NCCC-134
APPLIED COMMODITY PRICE ANALYSIS, FORECASTING AND MARKET RISK MANAGEMENT

The Value of Public Information in Storable Commodity Markets: Application to the Soybean Market

by

Christophe Gouel

Suggested citation format:

Gouel, C. 2018. “The Value of Public Information in Storable Commodity Markets: Application to the Soybean Market.” Proceedings of the NCCC-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management. Minneapolis, MN. [http://www.farmdoc.illinois.edu/nccc134].

The Value of Public Information in Storable Commodity Markets: Application to the Soybean Market*

Christophe Gouel†

Paper presented at the NCCC-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management
Minneapolis, Minnesota, April 16–17, 2018

*The author is grateful to Tamara Ben Ari, Benjamin Carton, Joe Glauber, Marc Nerlove, Will Martin, and Steve Salant for helpful comments. Special thanks to Scott Irwin for his insightful comments and for providing data on private forecasts. This work was supported by the INRA-Cirad metaprogramme GloFoodS and has been undertaken as a part of the CGIAR Research Programs on Policies, Institutions, and Markets (PIM), which is led by the International Food Policy Research Institute (IFPRI) and funded by the CGIAR Fund Donors. Christopher Marciniak and Maïmouna Yokessa provided excellent research assistance.

†Économie Publique, AgroParisTech, INRA, Université Paris-Saclay, 78850, Thiverval-Grignon, France and IFPRI, 1201 Eye St, NW, Washington, DC 20005, USA. Email: christophe.gouel@inra.fr

The Value of Public Information in Storable Commodity Markets: Application to the Soybean Market

Abstract: This study provides a framework to estimate the potential effects and benefits of the provision of market information in storable commodity markets. This framework is applied to the case of production forecasts for the soybean market. A rational expectations storage model of the global soybean market accounting for both inter-annual and intra-annual market dynamics is built. Shocks that occur between planting and harvesting affect the size of the potential harvest. Estimates of the size of these shocks are reported publicly, and affect the market equilibrium through adjustments to stock levels. By varying counterfactually the observability of seasonal shocks, we can estimate the efficiency gains related to the availability of advance information. They are equivalent to 2% of storage costs; the reduction of stock levels being the main channel explaining the welfare gains. The presence of advance information has a limited effect on inter-annual price volatility but redistributes price volatility during the season, increasing it just before harvest when almost all news has been received and stocks are tight, and decreasing it after harvest. The effect of news shocks is stronger on higher-order moments of the distribution with a strong decrease in skewness and kurtosis related to the lower frequency of price spikes.

Keywords: Commodity price dynamics, commodity storage, news shocks.

1 Introduction

Accurate and timely information is key to the good functioning of financial markets including commodity markets. What is specific to commodity markets is the extensive public involvement in the provision of market information. For example, national and international agencies regularly report stock levels and, in the case of agricultural commodities, forecast the coming harvests. Improving the transparency of the global food market was the rationale for the creation of the Agricultural Market Information System (AMIS) in 2011 by the G20 Agriculture Ministers as a response to the global food price hikes that occurred in 2007/08 and 2010. The important role of public information has been acknowledged in the literature (C-FARE, 2016) but evaluations of its counterfactual effects and benefits are rare despite the costs to public agencies of collecting data through surveys which is very high.

The objective of this article is to develop a framework to assess the economic role of public information in storable commodity markets. Public information can affect market allocation and welfare in many ways (see C-FARE, 2016, for a recent survey): by allowing efficient market allocation, by leveling the playing field for market participants with unequal access to private information, and by providing a basis for research on market behavior. Here, we focus on the allocation of resources over time and how it is affected by the timing of information. We build a rational expectations storage model that accounts for both inter-annual and intra-annual market dynamics, and in which shocks between planting and harvesting affect the size of the potential harvest. We apply the model to the case of soybean because its production is very concentrated—more than 80% of world production comes from Argentina, Brazil, and the United States (U.S.)—and its market is well integrated internationally (e.g., Merener, 2015, shows that soybean futures react similarly to weather events in Argentina, Brazil, and the U.S.), two features that simplify the modeling. To make the model computationally tractable, we define a simplified crop calendar corresponding to the most representative planting and harvesting dates and use monthly changes in U.S. Department of Agriculture (USDA) production forecasts to calibrate the size of the news shocks. Varying the size of the observable seasonal shocks counterfactually allows an assessment of the benefits of providing public information about potential shocks.

Because of its focus on the allocation of resources over time, the present article can be seen as updating the approaches proposed in Hayami and Peterson (1972) and Bradford and Kelejian (1977, 1978). These works developed analytical frameworks based on intra-annual storage models to assess the welfare costs of measurement errors in public information. Our approach differs in that we include the problem of seasonal news shocks in a modern rational expectations storage model such that seasonal information matters for both seasonal market dynamics and inter-annual market dynamics, because advance information about harvests allows smoother inter-annual stock adjustments. Also, the emphasis in our paper is slightly different. The above-mentioned papers were interested in assessing the welfare benefits of reducing forecast errors. Because monthly forecast errors are unobservable, we prefer to assume that production forecasts are without error, although we acknowledge that a part of the production shocks may not be observed. We use USDA data on last 40 years, which allows us to quantify the size of the informational shock at each period. However, adopting a modern rational expectations storage model framework has some drawbacks. This type of model is not analytically tractable, so we lose some of the insights provided by the analytical simplicity of Hayami and Peterson's framework. However, although it relies on numerical simulations, we have tried to keep our modeling framework as simple as possible to allow clear interpretation of all the results. For simplicity, we ignore one potential source of welfare gain associated to public information and considered in Hayami and Peterson (1972): adjustments of production to news shocks. We assume that supply is elastic at planting but that after that time it is not possible for production to react to shocks (see Bontems and Thomas, 2000, and Lechthaler and Vinogradova, 2017, for work on this issue).

Since the studies by Hayami and Peterson (1972) and Bradford and Kelejian (1977, 1978), most work on this topic adopted another approach inspired by market efficiency studies in finance which consists of assessing whether and how markets react to announcements (e.g., Garcia et al., 1997). This more recent literature investigates the value of USDA forecasts through their capacity to affect the market. USDA market outlooks are not the only source of market information. Large trading firms also gather information on market fundamentals. Some private companies specialize in selling market outlooks in advance of USDA releases. The existence of these other, private, sources of information raises questions about the economic function of public information. One way to assess its role is to quantify market reactions to the release of USDA reports. Adjemian (2012b) and Adjemian et al. (forthcoming) show that agricultural markets do react to the information contained in USDA reports, and that their reactions are consistent with storage theory.[1] However, although this literature quantifies how markets react to the release of public information, it is silent on the public benefits of this information, the object of the present study. Ensuring the tractability of our rational expectations model prevents us from accounting for the existence of multiple sources of

[1] Karali and Thurman (2009) show that the lumber futures market also reacts to announcements in a way that is consistent with the storage model.

information. We assume that all agents have the same set of information. So, we do not test whether USDA reports move the market through the provision of additional information; rather we assume that USDA reports reveal common information.

We use USDA production forecasts because their monthly releases are watched closely by market observers, and are considered a benchmark. USDA started producing monthly crop reports in 1863 during the Civil War when timely information on crop conditions was crucial for supplying the army (Adjemian, 2012a). Since then, with very few exceptions, USDA has released reports on crop production on around the 10$^{\text{th}}$ of every month. The 1972 "Great Grain Robbery" when the Soviet Union purchased large amounts of subsidized U.S. grains before world food prices spiked led to the establishment of an interagency process to produce balance sheets consolidating information on supply and demand (see figure A1 in the appendix for an example of the monthly balance sheets). Our study uses only production forecasts because after planting, changes to production forecasts can reasonably be considered exogenous and so can be seen as news shocks, while other balance sheet dimensions (domestic use, exports, ending stocks) mix endogenous market responses and news shocks. In 1980, the most important foreign markets for each commodity began to be included in the balance sheets, which provides us with changes in production forecasts for Argentina and Brazil. Other institutions than USDA—private firms and international organizations such as the International Grains Council, and since 2011, AMIS—also produce crop forecasts. So, assuming a counterfactual without any advance information would overestimate the effects of USDA reports. We use a series of private forecasts about the U.S. harvest to decompose the news shocks between components that are observed by all forecasters and components that are observed only by USDA. This decomposition allows us to differentiate the total effect of the existence of news shocks from the increased precision allowed by the USDA forecasts.

Our starting point is a rational expectations storage model extended to account simultaneously for inter-annual and intra-annual market dynamics. Adopting a rational expectations framework is a natural starting point as it implies that agents optimally process the information provided by news shocks.[2] Starting with early work by Lowry et al. (1987), Williams and Wright (1991, Ch. 8), and Chambers and Bailey (1996), a few papers have built rational expectations models with intra-annual dynamics. Only three (Ng and Ruge-Murcia, 2000; Osborne, 2004; Peterson and Tomek, 2005) include news shocks, where the word "news" takes the meaning it has been given in the macroeconomic literature (Beaudry and Portier, 2014) of observation of exogenous shocks with effects on fundamentals in the future. Ng and Ruge-Murcia (2000) show that news shocks can be a useful extension to the model to increase serial correlation in prices which has been a long-standing puzzle since the estimations by Deaton and Laroque (1992, 1996).[3] Peterson and Tomek (2005) introduce news shocks in a model of the U.S. corn market to improve its capacity to explain observed seasonal market behavior. Osborne (2004), which is the closest to our work, studies how precipitation observed in advance of harvest affects the market equilibrium in a storage model for the Ethiopian grain market. Compared to Osborne (2004), our model is richer, has elastic supply and more detailed news shocks structure. In particular, our monthly model (compared to Osborne's quarterly model) allows us to show that news shocks affect price volatility in subtle ways: they do

[2] The objectives of the recent work by Abbott et al. (2016) are similar to ours but without rational expectations, which makes it unclear how agents process the new information.

[3] Although some recent work (Cafiero et al., 2011; Gouel and Legrand, 2017) has gone a long way towards solving this puzzle.

not change much the average inter-annual price volatility but they redistribute intra-annual price volatility, increasing it just before harvest when the harvest is almost known and stocks are at their minimum, and decreasing it significantly otherwise. In addition, we do not have to build the news shocks from meteorological information; our reliance on USDA production forecasts means we have direct observations of news shocks expressed in quantities which allows better identification of the information available to agents. Lastly, a related problem of information frictions, not understood as news in the previous sense, was addressed recently by Steinwender (2018) in a rational expectations storage model. She shows how the introduction of the transatlantic telegraph reduced information frictions related to the international cotton trade between Great Britain and the U.S. Transatlantic shipping takes time and before the transatlantic telegraph was introduced, prices in the export and import markets were based on lagged information from the other market. Steinwender estimates that the instantaneous flow of information allowed by the telegraph generated efficiency gains equivalent to 9% of the export value. In our case, welfare gains arise mostly from better intertemporal allocation of resources, and amount to 2% of the value of the storage cost.

Finally, our work is related to macroeconomic studies which introduce news shocks into dynamic stochastic general equilibrium models, and see them as providing a positive theory of business cycle fluctuations (Beaudry and Portier, 2014). This is one of the motivations for introducing news into the storage model (Ng and Ruge-Murcia, 2000; Peterson and Tomek, 2005) but is not the only angle in this article which is concerned also with the welfare benefits associated to such information. Also, it is worth noting that there are at least two key differences in analyses of news shocks in macroeconomic models and the storage model. First, in agriculture while the underlying shocks are likely to be exogenous, observability of these shocks as news shocks is partly a function of public intervention. Second, most news shocks in agriculture have a clear physical interpretation (e.g., size of planted area, expectation of poor yields), and many are observable at some cost, while in macroeconomic models news shocks are rarely observed by the econometrician and are recovered either from financial variables or model-based structural decomposition.[4]

The rest of this paper is organized as follows. Section 2 develops our seasonal storage model. Section 3 describes how the model is taken to the data. It shows how USDA crop forecasts can be used to recover the size of the news shocks, and describes our approaches to recover storage cost and supply elasticity. Supply elasticity is estimated using a method inspired by Roberts and Schlenker (2013) but refined to account for the flows of information coming from asynchronous harvests in the two hemispheres. Section 4 presents the results of our counterfactual simulations. Section 5 offers some concluding remarks.

2 A seasonal storage model

We consider the soybean world market represented by a seasonal rational expectations storage model with 12 seasons per year. Years are indexed by t and seasons by $i \in \{1, \dots, I\}$ with the convention that $\{I + 1, t\} = \{1, t + 1\}$ and $\{1 - 1, t\} = \{I, t - 1\}$. There are two producing regions, indexed $r \in \{\mathrm{US}, \mathrm{LAC}\}$, corresponding to the U.S., and Argentina and Brazil combined. The two regions

[4]The work of Arezki et al. (2017) which analyzes the macroeconomic effect of giant oil discoveries is one of the rare exceptions where macroeconomic news shocks are observed directly.

differ in their planting and harvesting dates, and the size of their potential production. We adopt the convention that the year is the U.S. marketing year and starts with the U.S. harvest. We assume that trade is costless so that price, $P_{i,t}$, is the same in all regions and the localization of final demand or storers does not matter for the equilibrium.

Soybean generally takes around five months to grow from sowing to harvest. However, because of the variety of local climates, sowing to harvesting in the U.S. extends from mid-April to late December. The pattern is similar in Argentina and Brazil. To simplify the model, we assume that the U.S. production takes place over the five most active months from May/June to October/November, and the Argentina-Brazil production takes place between October/November and March/April. This simplified crop calendar means that the crop cycles do not overlap between regions: when planting occurs in one country harvest has been completed in the other. Since producers plant after the revelation of the size of the other hemisphere harvest, their price incentives are affected by this information through the effect of stock level adjustments on expected price, and their production plan are adjusted accordingly.

2.1 Consumers

Final demand is given by a demand function which is a downward sloping deterministic function of current price and is identical for every season: $D\left(P_{i,t}\right)$.

2.2 Storers

There is a single representative speculative storer that is risk neutral and acts competitively. Its activity is to transfer a commodity from one period to the next. Storing the non-negative quantity $S_{i,t}$ from period $\{i,t\}$ to period $\{i+1,t\}$ entails a purchasing cost, $P_{i,t}S_{i,t}$, and a storage cost, $k\bar{P}S_{i,t}$, with k the unit physical cost of storage expressed in proportion to the steady-state annual price $\bar{P}$. The benefits in period $\{i+1,t\}$ are the proceeds from the sale of previous stocks: $P_{i+1,t}S_{i,t}$. The storer follows a storage rule that maximizes its expected profit which, accounting for the non-negativity constraint on stocks, leads to the following non-arbitrage condition

$$\beta \, \mathrm{E}_{i,t}\, P_{i+1,t} - P_{i,t} - k\bar{P} \leq 0, = 0 \text{ if } S_{i,t} > 0, \tag{1}$$

where $\mathrm{E}_{i,t}$ is the expectation operator conditional on period $\{i,t\}$ information and β is the monthly discount factor and is assumed to be fixed.

Because of the convention that $\{I+1,t\} = \{1,t+1\}$, equation (1) applies equally to storers' behavior between seasons and between years.

2.3 Producers

Production in each region is undertaken by a representative competitive producer with decreasing returns to scale that takes the planting decision before knowing the selling price and the yield. The producer in region r plants in season i^r_π in year t in the expectation of harvesting the quantity $Q^r_{i^r_\pi,t}$ five months later. However, the production is affected by a multiplicative random shock $\epsilon^r_{i^r_\pi+5,t}$ which follows a distribution with unitary mean that is described later. We assume that after planting the producer cannot adjust its production level. The producer chooses its production level by solving the following maximization of expected profit:

$$\max_{Q^r_{i^r_\pi,t}} \beta^5 \,\mathrm{E}_{i^r_\pi,t} \left(P_{i^r_\pi+5,t} \epsilon^r_{i^r_\pi+5,t} Q^r_{i^r_\pi,t} \right) - \Psi^r \left(Q^r_{i^r_\pi,t} \right), \tag{2}$$

where $\Psi^r \left(Q^r_{i^r_\pi,t} \right)$ is the cost of planning the production $Q^r_{i^r_\pi,t}$ and $\epsilon^r_{i^r_\pi+5,t} Q^r_{i^r_\pi,t}$ is the realized production level. Profit maximization gives the following intertemporal equation

$$\beta^5 \,\mathrm{E}_{i^r_\pi,t} \left(P_{i^r_\pi+5,t} \epsilon^r_{i^r_\pi+5,t} \right) = \Psi^{r\prime} \left(Q^r_{i^r_\pi,t} \right), \tag{3}$$

which equalizes the marginal cost of production and the expected discounted marginal benefit of one unit of planned production.

2.4 News shocks

In contrast to standard annual storage models, we do not assume that the production shock is concentrated in one period. During the growing season, potential production is affected by a series of shocks. We assume that these production shocks are partly observable after their realization such that they provide advance information on the size of the coming harvest. They are news shocks in the sense that they do not directly affect current quantities but they do affect expectations about realizations of future quantities. Nonetheless, they affect the market equilibrium through immediate adjustments to stock levels. A higher than average news shock will increase expected output, decreasing in turn expected price at harvest and incentives to store. To maintain the non-arbitrage condition (1), storers will sell more to consumers driving the current price lower. The converse applies with a lower than average news shocks.

The literature on news shocks adopts two modeling strategies (Beaudry and Portier, 2014). It assumes either a noisy signal composed of the true shock plus a noise, or a shock composed of two elements, only one of which is observable. In the small theoretical literature on news shocks in commodity markets, Hayami and Peterson (1972) and Bradford and Kelejian (1977, 1978) adopt the noisy signal approach, and Ng and Ruge-Murcia (2000), Osborne (2004), and Peterson and Tomek (2005) adopt the other approach. In the noisy-signal approach, it is possible to generate richer dynamics since agents may have to react to false information in the case of a large noise shock. The downside to this approach is that it requires a larger state space, because it is necessary to follow separately the true shock and the associated noise which in the case of our storage model, would lead to a model that is more difficult to calibrate because of lack of information on the size of the potential noise. Therefore, we adopt the other approach.

The total production shock $\epsilon^r_{i^r_\pi+5,t}$ is assumed to be the product of a succession of five lognormallydistributed shocks occurring after planting:

$$\epsilon^r_{i^r_\pi+5,t} = \exp\left(\sum_{i=1}^{5} \eta^r_{i^r_\pi+i,t}\right), \tag{4}$$

where the $\eta^r_{i,t}$ are i.i.d. and follow normal distributions with mean μ^r_i and standard deviations σ^r_i such that $\mu^r_i + \left(\sigma^r_i\right)^2/2 = 0$ which ensures that seasonal shocks have unitary mean. Under these assumptions, $\epsilon^r_{i^r_\pi+5,t}$ follow a lognormal distribution with parameters $\sum_{i=1}^{5} \mu^r_{i^r_\pi+i}$ and $\sqrt{\sum_{i=1}^{5} \left(\sigma^r_{i^r_\pi+i}\right)^2}$. Given that $\mu^r_i + \left(\sigma^r_i\right)^2/2 = 0$, this implies that the expected mean of $\epsilon^r_{i^r_\pi+5,t}$ is 1.

There is no need for an explicit representation of shocks that are unobservable. In assuming that production is perfectly observed after harvest, the implication is that the unobservable part of the seasonal shocks is shifted to the harvest season, so in $\eta^r_{i^r_\pi+5,t}$, when it becomes observable.

Let $\hat{Q}^r_{i,t}$ denote region-r expected production in period $\{i, t\}$, for $i^r_\pi + 1 \leq i \leq i^r_\pi + 5$, as the product of planned production and realized production shocks:

$$\hat{Q}^r_{i,t} = \mathrm{E}_{i,t}\left[Q^r_{i^r_\pi,t} \exp\left(\sum_{j=i^r_\pi+1}^{i^r_\pi+5} \eta^r_{j,t}\right)\right] = Q^r_{i^r_\pi,t} \exp\left(\sum_{j=i^r_\pi+1}^{i} \eta^r_{j,t}\right). \tag{5}$$

2.5 Market equilibrium

Markets clear by equalizing the availability, noted $A_{i,t}$, which is the sum of recently produced and stored commodities, to final and storage demand:

$$A_{i,t} = D\left(P_{i,t}\right) + S_{i,t}. \tag{6}$$

2.6 Recursive equilibrium

The variables defining the state of the system change with the season. The availability is always part of the state variables, and is defined by available stocks plus current production if the season is a harvest season:

$$A_{i,t} = \begin{cases} S_{i-1,t} + \epsilon^r_{i,t} Q^r_{i-5,t} & \text{if } i = i^r_\pi + 5, \\ S_{i-1,t} & \text{if } i \neq i^r_\pi + 5. \end{cases} \tag{7}$$

In seasons after planting but before harvesting, defined by the set $\mathcal{I}_\pi = \{i^r_\pi + j \,|\, j = \{1, \ldots, 4\}$ and $r = \{\mathrm{US}, \mathrm{LAC}\}\}$, the expected production, $\hat{Q}^r_{i,t}$, is also a state variable since it will affect resource allocation via storage. Past storage and expected production play similar roles in characterizing the market supply but can be summed only at harvest time. Until then, expected production is still uncertain and so is not equivalent to stocks. In addition, even if expected production were certain, a future production cannot alleviate a current market scarcity, so the non-negativity of stocks would create a difference between these two state variables.

We define the set of state variables at season i and year t by $s_{i,t}$. $s_{i,t}$ includes availability, and if relevant, expected production: $s_{i,t} = \left\{A_{i,t}, \hat{Q}^r_{i,t}\right\}$ if $i \in \mathcal{I}_\pi$ and $s_{i,t} = \left\{A_{i,t}\right\}$ if $i \notin \mathcal{I}_\pi$.

From the above we can define the recursive equilibrium of the problem:

Definition. A recursive equilibrium is a set of functions $\mathcal{S}_i\left(s_{i,t}\right)$, $\mathcal{Q}^r_{i^r_\pi}\left(s_{i^r_\pi,t}\right)$, and $\mathcal{P}_i\left(s_{i,t}\right)$, defining storage, production, and price over the state variables and the transition equations (5) and (7) such that (i) storer solves (1), (ii) producer solves (2), and (iii) the market clears.

2.7 Welfare

To assess the value of public information, we define welfare as follows. A standard measure of instantaneous welfare, $w_{i,t}$, is provided by the sum of consumers' surplus, producers' surplus, and storers' surplus. However, since planting and harvesting occur at different season, the producers' surplus has to be split between revenues and production costs which have to be allocated to their respective season:[5]

$$w_{i,t} = \int_{P_{i,t}}^{P_{\max}} D\left(p\right) \mathrm{d}\, p + P_{i,t} S_{i,t-1} - \left(k\bar{P} + P_{i,t}\right) S_{i,t} + \begin{cases} P_{i,t}\epsilon^r_{i,t} Q^r_{i-5,t} & \text{if } i = i^r_\pi + 5, \\ -\Psi^r\left(Q^r_{i,t}\right) & \text{if } i = i^r_\pi, \\ 0 & \text{if } i \notin \left\{i^r_\pi, i^r_\pi + 5\right\}, \end{cases} \tag{8}$$

where $P_{\max}$ is the maximum price which since it is independent of all choices and variables, does not affect welfare comparison. Using equation (7), this expression can be simplified to

$$w_{i,t} = \underbrace{\int_{P_{i,t}}^{P_{\max}} D\left(p\right) \mathrm{d}\, p + P_{i,t}\left(A_{i,t} - S_{i,t}\right)}_{\text{Consumers' efficiency gains}} - k\bar{P} S_{i,t} - \mathbf{1}_{i=i^r_\pi} \Psi^r\left(Q^r_{i,t}\right), \tag{9}$$

where $\mathbf{1}_{(\cdot)}$ is the indicator function. This equation presents three easily interpretable efficiency terms. Combining consumers' surplus with the value of demand removes from the consumers' welfare any

Then there are the storage costs, and finally the production costs. This decomposition will help understanding of the source of welfare changes.

Using this expression of instantaneous welfare, we can calculate the intertemporal welfare, normalized to a monthly value by

$$W_{i,t} = \left(1 - \beta\right) w_{i,t} + \beta\, \mathrm{E}_{i,t}\, W_{i+1,t}. \tag{10}$$

The welfare is a function of the season and of the state variables. To remove the dependence on

[5]When producing soybean, production costs are not all incurred at planting; harvesting is also a significant source of costs. As long as production is inelastic after planting, the exact timing of costs does not matter for the results and, for simplicity, we assume they are all incurred at planting when the production decision is taken.

the state variables, we calculate welfare as an average over the asymptotic distribution of the state variables. Once averaged, welfare varies little with season because of the low discounting we adopt; therefore, without any consequences for the results, we choose the first season to calculate the welfare results.

3 Taking the model to the data

We need to assume functional forms to take the model to the data. We assume functional forms to be isoelastic, with demand given by

$$D(P) = \frac{\bar{D}}{12}\left(\frac{P}{\bar{P}}\right)^{\alpha^D}, \tag{11}$$

where $\bar{D}$ is the steady-state annual demand and $\alpha^D < 0$ is the price elasticity of demand. Similarly, production costs are given by

$$\Psi^r(Q^r) = \frac{\beta^5 \bar{P}}{\left(\theta^r \bar{D}\right)^{1/\alpha^Q}} \frac{(Q^r)^{1+1/\alpha^Q}}{1+1/\alpha^Q}, \tag{12}$$

where $0 \leq \theta^r \leq 1$ is the steady-state share of region r in world production, and $\alpha^Q > 0$ is the supply elasticity, assumed to be the same in all regions. With this expression of production costs, equation (3) simplifies to

$$Q^r_{i^r_\pi,t} = \theta^r \bar{D}\left[\mathrm{E}_{i^r_\pi,t}\left(\frac{P_{i^r_\pi+5,t}}{\bar{P}}\epsilon^r_{i^r_\pi+5,t}\right)\right]^{\alpha^Q}. \tag{13}$$

$\bar{P}$ and $\bar{D}$ are defined as steady-state values of price and demand in the annual model corresponding to the seasonal model with all decisions collapsed into one period, and production and stockpiling decisions taken inter-annually (see section A in the appendix). It is easier to work with these two annual steady-state values than with the 24 seasonal values of the monthly model.

Using these isoelastic functional forms, in all the equations the variables in levels can be substituted by the corresponding ratio to their annual steady-state levels with the exception of $S_{i,t}/\bar{D}$ and $A_{i,t}/\bar{D}$, which can be interpreted as the stock and availability to annual steady-state use ratios. Similarly, welfare can be expressed unit-free by normalizing $W_{i,t}$ by the monthly steady-state consumption value, $\bar{P}\bar{D}/12$. Since $\bar{P}$ and $\bar{D}$ serve only to define the level of prices and quantities without affecting any of the other results, we normalize their values to 1 and 12. So the monthly steady-state values of price and demand will be close to 1. Table 1 gives the calibration values of all the parameters.

We fix the discount factor by assuming an annual real interest rate of 2% which is close to the world average preceding the 2008 crisis and the subsequent very accommodating monetary policies (IMF, 2014, Ch. 3). The share of production from the U.S., θ^{US}, is 42%, its average 2007–2016 value (considering that Argentina, Brazil, and the U.S. are the only producers).

Table 1: Parameterization

Parameter	Economic interpretation	Value	
$\bar{P}$	Steady-state annual price	1	
$\bar{D}$	Steady-state annual demand	12	
β	Monthly discount factor	0.9984	
k	Monthly unit storage cost (% of $\bar{P}$)	0.56	
θ^{US}	Share of US production in total production (%)	42	
α^D	Demand elasticity	−0.2	
α^Q	Supply elasticity	0.4	
		US	LAC
$\sigma^r_{i^r_\pi+1}$	Standard deviation of the 1st news shocks (%)	3.65	2.96
$\sigma^r_{i^r_\pi+2}$	Standard deviation of the 2nd news shocks (%)	4.31	3.21
$\sigma^r_{i^r_\pi+3}$	Standard deviation of the 3rd news shocks (%)	3.63	2.29
$\sigma^r_{i^r_\pi+4}$	Standard deviation of the 4th news shocks (%)	3.33	2.56
$\sigma^r_{i^r_\pi+5}$	Standard deviation of the 5th news shocks (%)	1.73	2.43
$\sqrt{\sum_{i=1}^{5}\left(\sigma^r_{i^r_\pi+i}\right)^2}$	Standard deviation of the total production shock (%)	7.69	6.07

3.1 Behavioral parameters

This storage model is too complex and takes too long to solve to be amenable to a fully structural estimation with cross-equation restrictions. Instead, we estimate supply elasticity with an approach that follows Roberts and Schlenker (2013) and recover storage cost and demand elasticity from other sources.

3.1.1 Storage cost

To choose the value of storage cost k, we can note that the Chicago Board of Trade's soybean futures contract defines a maximum storage rate at delivery locations of 0.165 cent per bushel per day, about 5 cents per month. This rate has been rarely changed since the 1980s, and only by small amounts. Because of arbitrage, this implies that the interest-adjusted spread between two contracts cannot exceed this rate. It is a maximum rate, so if actual storage costs at delivery locations or close by were significantly lower, the spread would rarely be close to this upper bound. We can assess the relevance of this value by plotting the interest-adjusted spread. We calculate the spread between the nearby contract at delivery and the next-to-expire contract for all available soybean futures contracts at the Chicago Board of Trade between 1959 and 2018. We take the settlement price on the first delivery day of the nearby contract, and for the same day, the settlement price of the next-to-expire futures contract. The next-to-expire futures is adjusted for the nominal interest rate using the secondary market rate of the 3-month treasury bills. The calculated spread is not deflated by the consumer price index because of the near-constancy of the storage rate since the 1980s (we calculate later the spread in percentage which will correct for changes in the price index).

All the financial data are taken from Quandl.

Figure 1 represents the spread for each futures contract using a violin scatter plot. The x-axis starts with the November contract, the first contract after the harvest (and usually considered to be representative of the harvest-time price). The distribution of spreads across the contracts is consistent with Working's (1949) storage supply curve theory. The storage supply curve implies a richer theory of storage than relied on in this paper (see Joseph et al., 2016, for empirical evidence on the current relevance of Working's storage supply curve). Here, we consider only the speculative motive for storage: if the spread between nearby and distant futures is not sufficient to cover the storage and opportunity costs, stocks are not carried out. There are other motives for stockpiling (e.g., transaction and precaution motives), so discretionary stocks are never zero, and stocks may be carried out at apparent losses. The benefits, other than speculative, that storers derive from holding the physical commodity are dubbed the "convenience yield" (Kaldor, 1939). Since after the harvest soybean is widely available, the convenience yield is likely to be small for the November contract: spreads are rarely negative and are likely to be at full carry. The more further in time one moves from November, the more dispersed the spreads are, with many more occurrences of negative spreads that are rare after the harvest. The points are colored by year to account for possible changes in storage costs across time.

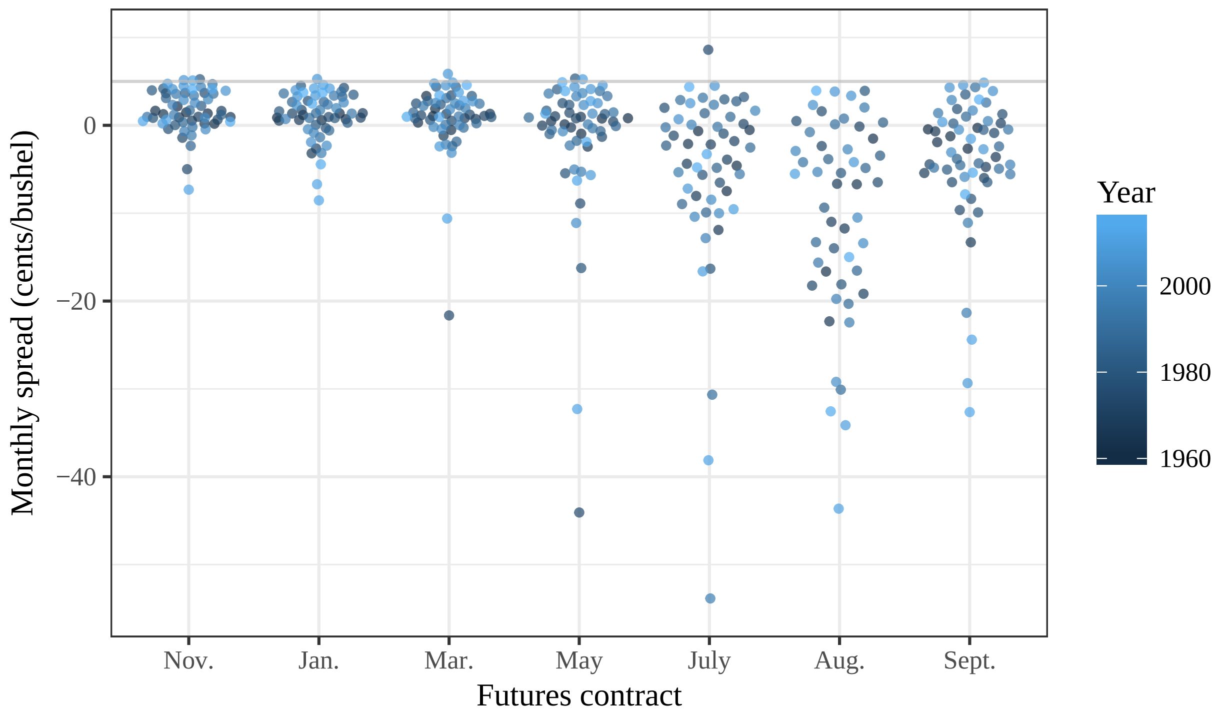

Figure 1: Interest-adjusted monthly price spread between a contract on its first delivery date and the following contract. Truncated y-axis: 9 spreads are below -55 cents per bushel.

The storage rate of 5 cents per bushel per month (indicated by a gray horizontal line) appears to effectively play the role of an upper bound with many spread values clustered just below. So, we retain this value as representing long-run storage costs. In the short-run, if stock levels are high, storage costs can exceed this value, but this cannot be observed in figure 1. If storage costs exceed the maximum rate at delivery locations, this can result in convergence failure between the price of the expiring futures contract and the spot market, with the spread between the spot and the next-to-expire contract exceeding the storage rate at delivery (Garcia et al., 2015). Such failures occurred repeatedly between 2006 and 2010 in grain markets, indicating that storage costs can exceed 5 cents per bushel under some conditions. This value of storage cost is also consistent with a study by World Bank and FAO (2012, Figure 2-4) that reports monthly storage cost for wheat to be \$2.02/ton (5.5 cents per bushel) in the U.S. in 2009. In the model, k is dimensionless and expressed as a percentage of the annual steady-state price. So, we have to normalize the storage cost by a reference price. We use the average delivery price over the last two decades, \$8.98 per bushel, which leads to $k = 0.0056$. On an annual basis, this number implies an annual storage cost of 6.72% of the steady-state price. This value is obviously a function of the choices made about the storage cost and the reference price. To account for possible trends in prices or storage costs that could have affected our choice, we represent in figure 2 the monthly spread as a percentage of the delivery price. Expressed in percentage, the maximum spreads do not seem to be function of the time period. The chosen value of 0.56% is within the cluster of high values but below the maximum that can reach 1.12%.

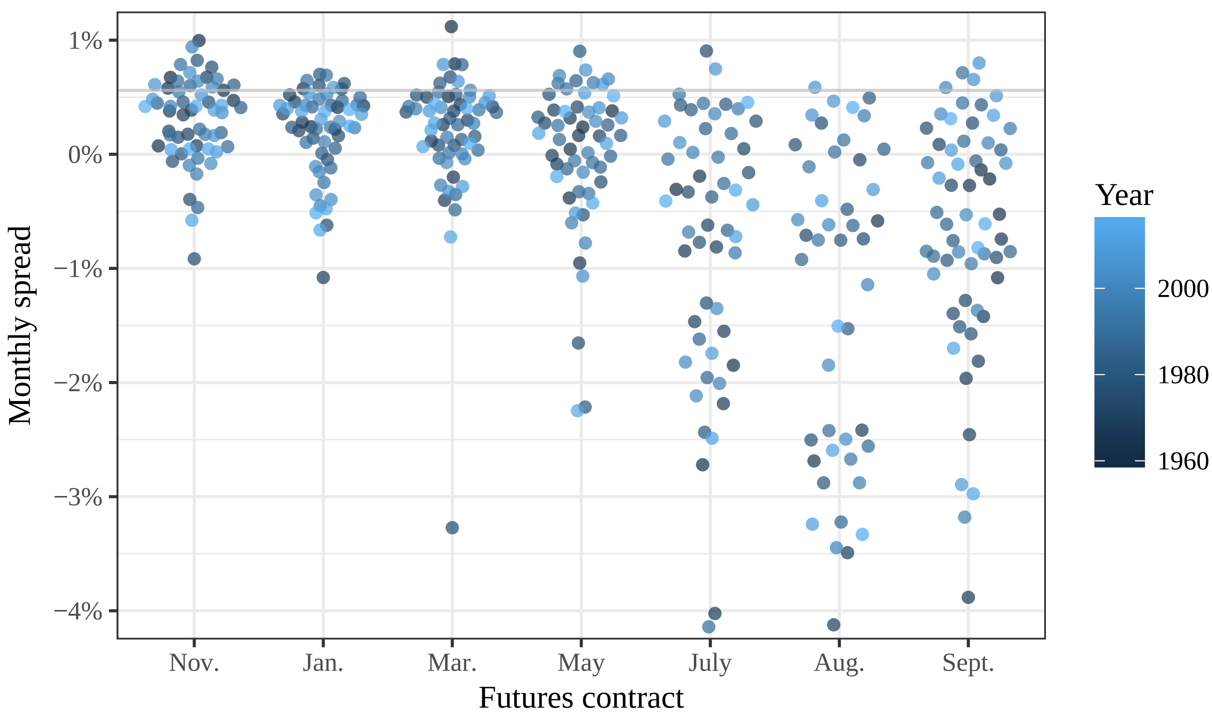

Figure 2: Interest-adjusted monthly price spread in percentage between a contract on its first delivery date and the following contract. Truncated y-axis: 17 spreads are below −4.2%.

3.1.2 Supply elasticity

To estimate supply elasticity, we adopt Roberts and Schlenker's (2013) instrumental variable approach which we modify to fit our framework. Roberts and Schlenker's approach provides two key insights. First, they argue that using the futures price as an explanatory variable is not enough to purge the model of endogeneity because futures prices can reflect anticipated shocks unobserved to the econometrician. Second, since speculative storage links prices between periods, futures prices are affected by past shocks, with the result that past shocks can be used as instruments.

We make the following adjustments to Roberts and Schlenker's empirical model. Since we are interested in the supply response of the U.S. and Argentina-Brazil, and because of their crop calendar the two regions face different price incentives, unlike Roberts and Schlenker we do not estimate the model on a world aggregate but create a panel of the two regions. Our dependent variable is the log quantity supplied, denoted $q_{rt} = \log\left(Q^r_{i^r_\pi+5,t}\epsilon^r_{i^r_\pi+5,t}\right)$. Our explanatory variables are the log of the harvest-time price expected at planting time, $p^q_{rt} = \log\left(\mathrm{E}_{i^r_\pi,t} P_{i^r_\pi+5,t}\right)$, corresponding empirically to the futures price, and a region-specific random shock, $\omega_{rt} = \epsilon^r_{i^r_\pi+5,t}$. The random shock is included as an explanatory variable to purge the dependent variable of its stochastic component and to reduce the endogeneity problem since realized shocks are likely to be a good proxy for anticipated supply shocks (Hendricks et al., 2015).

For the first-stage equation, we follow Roberts and Schlenker (2013) and use lagged yield shocks as instrument but to be consistent with our theoretical model we distinguish yield shocks by hemisphere.[6] We use ω_{ht} to denote the yield shock in the hemisphere h. The most recent shock to instrument the futures price faced by U.S. farmers is the Southern hemisphere harvest which is contemporaneous with sowing in the U.S. Similarly, the Northern hemisphere shock is used to instrument the futures price faced by farmers in Argentina-Brazil. In our case, lagged supply shocks on their own tend to be a weak instrument. Indeed, theoretically, because of storage, supply shocks play a time-varying role in determining prices: they have a limited effect on prices when stocks are abundant, and conversely, have an important effect when stocks are low. We also use an additional instrument, not employed by Roberts and Schlenker (2013). To account for the fact that lagged yield shocks have differential effects depending on stock levels, we add lagged harvest-time prices, $p^d_{rt-1} = \log\left(P_{i^r_\pi+5,t-1}\right)$, and their interaction with the hemispheric yield shocks as instrument. Harvest-time prices can be seen as proxying for market availability, and their interaction with yield shocks accounts for the possibly non-constant effect of yield shocks.

The resulting empirical model is

$$q_{rt} = \iota + \alpha^Q p^q_{rt} + \gamma\omega_{rt} + f^q_r(t) + \delta_r + u_{rt}, \tag{14}$$

$$p^q_{rt} = \zeta + \kappa\omega_{rt} + \lambda\omega_{ht-1} + \nu p^d_{rt-1} + \xi\omega_{ht-1}p^d_{rt-1} + f^p_r(t) + \tau_r + v_{rt} \text{ for } r \notin h. \tag{15}$$

Each equation includes a region-specific time trend $f^j_r(t)$, modeled by restricted cubic spline ($j \in \{q, p\}$),[7] and region fixed effects, δ_r and τ_r. This empirical model is consistent with our theoretical model with the exception of the anticipated shocks at planting time (unobserved by the econometrician) which justify use of an instrumental variable approach, and which for simplicity, we neglect in our theoretical model.

[6]See also Winne and Peersman (2016) for a recent study using harvest timings around the world to build the exogenous variables.

[7]Following the heuristics proposed by Harrell (2001), the knots for the cubic spline with three knots are located in 1967, 1988, and 2009. The spline with four knots uses 1965, 1980, 1996, and 2011, and the spline with five knots uses 1965, 1976, 1988, 2000, and 2011.

Soybean production and yields are from FAOSTAT (FAO, 2017). Countries are classified into Northern and Southern hemisphere using Sacks et al.'s (2010) crop calendar: countries with mean planting date between February 19th and July 19th are classed as Northern hemisphere. In the case of countries not covered by Sacks et al. (2010) which are mostly minor producing countries, we use a simple heuristic and assign to Northern hemisphere all countries whose capital city is located north of Mexico City's latitude. Futures prices are from the Chicago Board of Trade with a delivery month of November for the U.S., and May for Argentina-Brazil. p^d_{rt} is constructed as the log of the average futures price during the month of delivery, and p^q_{rt} is constructed as the log of the average futures price in April for the U.S. and in October of the previous year for Argentina-Brazil. All prices are deflated by the consumer price index from the Bureau of Labor Statistics. Yield shocks are constructed using the method described in Roberts and Schlenker, so ω_{ht} is the average across countries of hemisphere h of the log yield deviations from a three-knot restricted cubic spline trend. The hemispheric shocks are scaled to represent deviations from the world production trend, so that their sum is identical to the yield shock variable (ω_t) used in Roberts and Schlenker. For example, $\omega_{ht} = -0.02$ indicates a -2% shock to world production related to the deviation of the average yield in hemisphere h from its trend. Similarly, ω_{rt} is the region r log yield deviation from trend, calculated using jackknifed residuals. FAOSTAT data are available from 1961 to 2014 but our lagged yield shocks mean that our sample starts in 1962.

Estimation results are presented in table 2 and include 2SLS and OLS estimates with different specifications of the time trend. The F-statistics indicates that the first stage is strong. The overidentification test does not reject the validity of the instruments. The first-stage coefficients follow the intuitions from theory. The coefficient of p^d_{rt-1}, which can be interpreted directly because ω_{ht} has zero mean, indicates that lagged delivery prices have a positive influence on harvest-time prices which is consistent with storage theory. As expected, the marginal effect of lagged yield shock, ω_{ht-1}, depends on the market condition: evaluated at the mean delivery price it is -1.68, evaluated at the minimum sample price it is -0.01, and evaluated at the maximum sample price it is -3.92. So, the same supply shock has very different price effect depending on the prevailing market conditions. The estimated supply elasticities are higher than Roberts and Schlenker's estimate for a caloric aggregate but similar to their estimate in a four-crop system for soybean alone (Table A7). They are similar also to Haile et al.'s (2016) estimates. The estimates are sensitive to the flexibility of the time controls, with a much higher supply elasticity for a three-knot spline (in 2SLS and OLS).[8] This sensitivity which is not present in Roberts and Schlenker (2013) is explained by the fact that our dependent variable is regional not world production. World production has a linear trend over the period and so is straightforward to detrend but soybean production in Latin American countries took off only in the 1970s after the 1973 U.S. embargo on export of soybean and, so requires more flexibility in the trend. Supply elasticities estimated by OLS are much lower than estimated by 2SLS, and the Hausman test confirms the endogeneity of futures prices. These differences between 2SLS and OLS suggest also that Hendricks et al.'s (2015) conclusion that it is not necessary to instrument futures price when current yield shock is included in the explanatory variables may not always hold. Consequently, our preferred estimate uses the 2SLS estimator and the more flexible time trend, and for the simulations we retain a supply elasticity of 0.4.

[8]The results for the six-knot spline estimation are very close to those for the five-knot spline so are not reported here.

Table 2: Estimates of supply elasticity

	2SLS			OLS		
	(1a)	(1b)	(1c)	(2a)	(2b)	(2c)
Supply elasticity α^Q	0.60^{***}	0.25^{**}	0.38^{***}	0.51^{***}	0.09	0.15^{**}
	(0.08)	(0.10)	(0.10)	(0.07)	(0.07)	(0.07)
Shock ω_{rt}	1.60^{***}	1.39^{***}	1.21^{***}	1.62^{***}	1.41^{***}	1.24^{***}
	(0.22)	(0.15)	(0.15)	(0.21)	(0.15)	(0.14)
First stage ω_{rt}	0.16	0.10	0.08			
	(0.16)	(0.14)	(0.15)			
First stage ω_{ht-1}	10.48^{*}	9.68^{*}	10.08^{*}			
	(6.00)	(5.45)	(5.50)			
First stage p^d_{rt-1}	0.79^{***}	0.58^{***}	0.57^{***}			
	(0.04)	(0.06)	(0.06)			
First stage $\omega_{ht-1}p^d_{rt-1}$	-2.00^{**}	-1.84^{**}	-1.91^{**}			
	(0.97)	(0.88)	(0.89)			
First-stage F-statistics	116.03	38.77	35.67			
p-value for Hausman test	0.02	0.01	0.00			
p-value for overid. test	0.55	0.77	0.79			
Observations	106	106	106	106	106	106
Spline knots	3	4	5	3	4	5

Note: Columns 1a–1c use two-stage least squares; columns 2a–2c use ordinary least squares. Columns a, b, and c respectively include restricted cubic splines in time with 3, 4, and 5 knots. The coefficients in the first two rows are the results for log supply; the coefficients from the third to the sixth row give the first-stage results of the log price. Coefficients of time trends and regions are suppressed. ***, **, and * indicate significance at the 99%, 95%, and 90% levels, respectively.

3.1.3 Demand elasticity

We have not been able to estimate soybean demand elasticity using Roberts and Schlenker's method. They obtain a statistically significant demand elasticity estimating a four-crop system, only by imposing symmetry and by using USDA rather than FAOSTAT data (Table 8). They obtain an elasticity of −0.236. We are unaware of any other recent source of credibly estimated demand elasticity for soybean, so we follow Roberts and Schlenker and use a demand elasticity of −0.2.

3.2 Seasonal shocks

3.2.1 Public forecasts

To calibrate the size of the seasonal shocks, we use the monthly changes in production projections from the World Agricultural Supply and Demand Estimates (WASDE) reports (USDA, 2017b). The WASDE reports are balance sheets consolidating the information available in different USDA services. For the forecasts regarding U.S. production, the original information comes from the National Agricultural Statistics Service (NASS). The forecasting cycle for soybean works as follows (Good and Irwin, 2006). It begins with the Prospective Plantings report that is released at the end of March and that contains the farmers' planting intentions. Planted acreages are then updated at the end of June in the Acreage report. In the following months, estimates of acreages will be adjusted only slightly, and the focus shifts to yields forecasts. Around the 10th of every month, NASS releases a Crop Production report, the main elements of which can be found in the WASDE report released at the same moment and used in this paper. The August report contains the first non-trend yield forecast, and these forecasts are updated in the following months. Before the August report, the U.S. production forecast is the product of the acreage estimates by the trend yields. The production forecasts for foreign countries have a less rigid schedule and are based on information from USDA Foreign Agricultural Service attaché reports, foreign official sources, and satellite imagery, among other sources.

To make the link between the model and the data, we note that using equation (5), we obtain $\exp\left(\eta^r_{i,t}\right) = \hat{Q}^r_{i,t}/\hat{Q}^r_{i-1,t}$: the seasonal shocks can be identified by calculating the ratio between two consecutive projections. Given the assumption of shocks with unitary mean, we have $\operatorname{var}\left(\hat{Q}^r_{i,t}/\hat{Q}^r_{i-1,t}\right) = \exp\left(\left(\sigma^r_i\right)^2\right) - 1 \approx \left(\sigma^r_i\right)^2$. One benefit of assuming multiplicative shocks is that we do not need to detrend the data to compare shocks across years regardless of the trending production level. Based on these seasonal shocks, table 3 reports the mean and standard deviation of the monthly adjustments to production projections. We cannot reject the null hypothesis that the mean is equal to 1, as is assumed in the model, for any of the months considered. Therefore, the standard deviation can be interpreted as a coefficient of variation and is multiplied by 100 to ease interpretation.

For the U.S. market, projections for the new harvest begin in the WASDE reports with planting in the reports for May (in earlier reports, projections for the new crop year started in June or July) based on the combination of information on planting intentions and trend yields. There are no changes in projections between May and June, as acreage and yield estimates are not updated. The normal planting and harvesting periods for U.S. soybean are as follows (USDA, 2010). Planting of soybean begins in late April/early May and ends in early July with second half of May being the most active period. U.S. soybean harvest begins in September and is mostly completed by end November. This is consistent with the WASDE reports. Most production changes occur between June and November, with the largest changes observed in the August report which is the first to include survey-based yield forecasts and not just trend yields. Between June and July, yield estimates do not change and all changes in production are explained by changes in planted and harvested acreages. After July, most of the adjustments come from changes in yields, acreage estimates being fairly stable. No adjustments are made to production projections between November and December; some small adjustments are made between December and March based on data revisions. Further revisions can occur several months later as additional data (e.g., crushing or exports) become available.

Based on the usual planting and harvesting dates and the WASDE reports, we assume that the first season in the model, $i = 1$, when U.S. soybean is harvested corresponds to the period October 15 to November 15, denoted October/November. Therefore, in the model we represent the U.S. soybean crop cycle as starting with the planting decision in May/June and ending with harvesting

Table 3: Mean and standard deviation of monthly adjustments to production projections

	U.S. (1973–2016)[a]			Argentina and Brazil (1984–2016)[a]					
					Argentina		Brazil		LAC[b]
Month	# Obs.	Mean	SD (%)	# Obs.[d]	Mean	SD (%)	Mean	SD (%)	SD (%)
June	22	1	0	8	1	0	1	0	0
July	30	0.992	3.654	13	0.999	0.509	1	0	0.199
Aug.	42	0.991	4.309	23	1.002	0.753	1.003	1.189	0.708
Sept.	43	0.996	3.625	23	1.007	1.214	1.001	1.523	1.015
Oct.	43	1.004	3.325	31	1.007	2.011	1.006	2.086	1.607
Nov.	43	1.006	1.730	31	1.006	1.705	1.004	1.610	1.347
Dec.	39	1	0	32	1.005	1.426	0.999	0.839	0.660
Jan.	39	1.001	1.382	33	1.001	1.624	1.000	3.234	2.313
Feb.	39	0.999	0.405	33	1.000	4.836	1.002	2.322	2.508
Mar.	39	1.000	0.009	33	1.003	2.194	1.004	2.493	1.784
Apr.	44	1	0	33	1.001	2.794	0.997	2.285	1.998
May	39	1	0	32	0.988	2.987	0.999	2.246	1.899
June	39	1	0	32	1.001	2.903	1.002	1.365	1.139
July	43	1.000	0.267	32	0.995	2.258	1.000	0.975	0.906
Aug.	43	1	0	32	0.999	1.145	1.002	0.541	0.544
Sept.	43	1	0	23	1.001	0.887	1.000	0.378	0.396
Oct.	42	1.002	1.456	22	1.001	0.461	1.003	0.722	0.555
Nov.	42	1.000	0.009	22	1.002	1.242	1.001	0.407	0.515
Between-year volatility[c]			14.791			21.656		12.844	11.674

Source: Authors' calculation based on WASDE (USDA, 2017b) and PSD (USDA, 2017a). If there is more than one production forecast per month, as in the 1970s, we select the release date that is closest to the 10th of the month.
Note: [a] Marketing years used to calculate the statistics. [b] The last column contains the standard deviation of monthly adjustments to production projections of Argentina plus Brazil. [c] Calculated as the standard deviation of the first-order difference of the logarithm of the final production as reported in PSD for the same marketing years as the monthly adjustments. [d] There are fewer observations outside of the months of interest for the Latin American countries, because some could not be collected automatically from the original pdf files. Since they would not affect any of the results, they were not collected manually.

in October/November. There are five months following planting each with a seasonal shock. The standard deviations of the seasonal shocks, σ_i^{US}, are simply the standard deviations of the adjustments to production between July and November (table 3). Using these five standard deviations for seasonal shocks implies a standard deviation for the equivalent aggregate shock of 7.7%. This number can be compared to a between-year volatility in production of 14.8% (standard deviation of the first-order difference of the logarithm of final U.S. production values as reported in USDA Production, Supply and Distribution database): the combined seasonal shocks appear to account for a sizable share of this volatility. The remaining production volatility is attributable to shocks occurring before planting, the endogenous reactions of farmers to market incentives, unobservable seasonal shocks that are revealed in the data revisions made after harvest, and other types of shocks such as input price shocks.

The inclusion of the standard deviation of production for July constitutes a small deviation from the theoretical framework presented in section 2. The changes in the July production forecast for the U.S. come only from changes in acreages not yields. The changes in acreages between the May and July WASDE reports are caused by both endogenous farmers reactions to changes in the market and exogenous shocks (e.g., wet weather preventing planting). By using the July report in the calibration of the shocks, we are assuming that the changes are fully explained by exogenous shocks. This assumption avoids the need to decompose the shock between its components, which has never been done in the literature. Karali et al. (2018) show that the June acreage reports underlying the July WASDE reports generate important price reaction, which could be interpreted as indicating a large, stochastic, and unobserved-by-the-market component, supporting our choice.

Our modeling framework focuses on news shocks, and so, overlooks other types of informational shocks. For example, it neglects the revisions made to production estimates a few months after harvest. Since soybean is mostly crushed or exported, its use can be accounted for with more precision than in the case of other crops with more diverse uses (such as maize or wheat). The data on use provide information on the size of the recent harvest, and are used to update the production estimates. Also, it takes a couple of months to collect all the information on final yields. These two pieces of data explain the non-zero numbers in table 3 for January, February, and later months after the harvest. These data revisions involve a different type of informational problem from the one studied in the present article. They imply that the state of the system is observed only with measurement errors, and the new observables at each period allow the agents to update their estimations of the state. This is a problem that is related more to filtering—how agents extract information about the state from the observables, and how it affects market outcomes (studied recently by Leduc et al., 2016, in the context of the oil market and by Galvão, 2017, in the context of revising the macroeconomic data)—than to news shocks. Therefore, given our aim to assess the economic value of news shocks, neglecting these post-harvest shocks should not affect our results.

For Argentina-Brazil, identification of the seasonal shock structure is more challenging. Taken together, these countries present much greater agro-climatic diversity than the U.S., so the planting and harvesting of the main soybean crop is spread over more time. In addition, because of the tropical climate in these countries, double-cropping systems are possible, spreading soybean production even further over time. The WASDE reports began separate coverage of Argentina and Brazil only in January 1985; 1982 was the first marketing year for which some data were published, and 1985 was the first marketing year with complete data. Therefore, we have fewer observations than for the U.S. The production projections for these countries present different patterns from those for the U.S. They start with the U.S. projections, in May in the most recent reports, so several months before planting in the Southern hemisphere. The initial projections are estimates based on likely production level trends by USDA oilseeds experts. These early projections can be adjusted before the planting seasons, based on additional information. The variations between May and October are different from pure news shocks since they are the difference between the USDA trend estimates and the first information on planting conditions, or planting intentions. Another feature that differs from U.S. projections is that the projections for Argentina and Brazil do not seem to converge smoothly to some final estimates, even in the case of August, some months after the end of the main harvest. Crop adjustments are much larger in Argentina than in Brazil or the U.S., but because it is a smaller producer than Brazil and their production shocks are not correlated, the volatility of the Argentina-Brazil region is often smaller than the volatility of the countries considered separately.

In order to obtain a series of five seasonal shocks for this region, we have to make some assumptions. Based on the AMIS crop calendar,[9] most planting in Brazil happens in October and November, and the harvest takes place in March and April. In Argentina, planting is during November and December, and harvesting takes place in April and May. For this region, the model assumes that planting takes place in October/November and harvesting takes place in March/April. We consider that the main news shocks are the five shocks with the largest standard deviations between January and May. However, were we to use the values for these months in the last column of table 3, this would result in a standard deviation of the total production shock of 4.74%, well below the 11.67% that is observed inter-annually. Consequently, we rescale seasonal shocks so that the total production shock represents 52% of the inter-annual volatility, which is observed for the U.S. The resulting standard deviations are presented in table 1. They show a more regularly spread variability than in the case of the U.S.

Figure 3 plots the densities of the aggregate shocks conditional on information available at the month of forecast and assuming previous shocks to be at their mean. It illustrates how the assumed standard deviations of the successive monthly shocks result in conditional densities that narrow the more one moves in the production season. In the planting months, October/November for Argentina-Brazil and May/June for the U.S., the densities are widely spread and the possibilities of crop failures or exceptionally good crop could not be excluded. On the contrary, one month before the harvest, respectively February/March and September/October, the distributions have narrowed a lot and the size of the harvest is almost known. In this last month, the precision is much higher for the U.S. than for Argentina-Brazil where the last month still displays a large standard deviation.

3.2.2 Private forecasts

In addition to USDA forecasts, there are also private analysts who produce their own harvest forecasts and release them to their clients a few days before the USDA reports. To avoid overestimating the gains from the provision of public information, it is important to be able to account for the information that would come from these other sources in the absence of USDA reports. But were the private forecasts really private, they would not affect our assessment of the benefits of public information. Actually, they do not stay private for long. After their initial private release based on which traders adjust their positions, private forecasts are often made public in specialized professional news outlets (e.g., Bloomberg and Reuters) through surveys of the traders or directly from the firms establishing the forecasts.

We use a database of private forecasts about the U.S. harvest extending from 1983 to 2017 for the July forecasts and from 1970 to 2017 for the August, September, October, and November forecasts.[10] There is no equivalent database for the Argentina and Brazil crop forecasts, but their production is also monitored by some market analysts. What we have called the July forecast is different from the forecast in the later months, because it is concerned only with forecasting the acreages and is released in the NASS Acreage report at the end of June before being incorporated in the July WASDE report. For 1983–2015, the July forecasts are an average of analyst forecasts collected by Bill Tierney. After 2015, it comes from Reuters. For 1970–2000, the August, September, October,

[9] http://www.amis-outlook.org/amis-about/calendars/en/, retrieved March 27, 2017.

[10] This database was kindly provided to us by Scott Irwin.

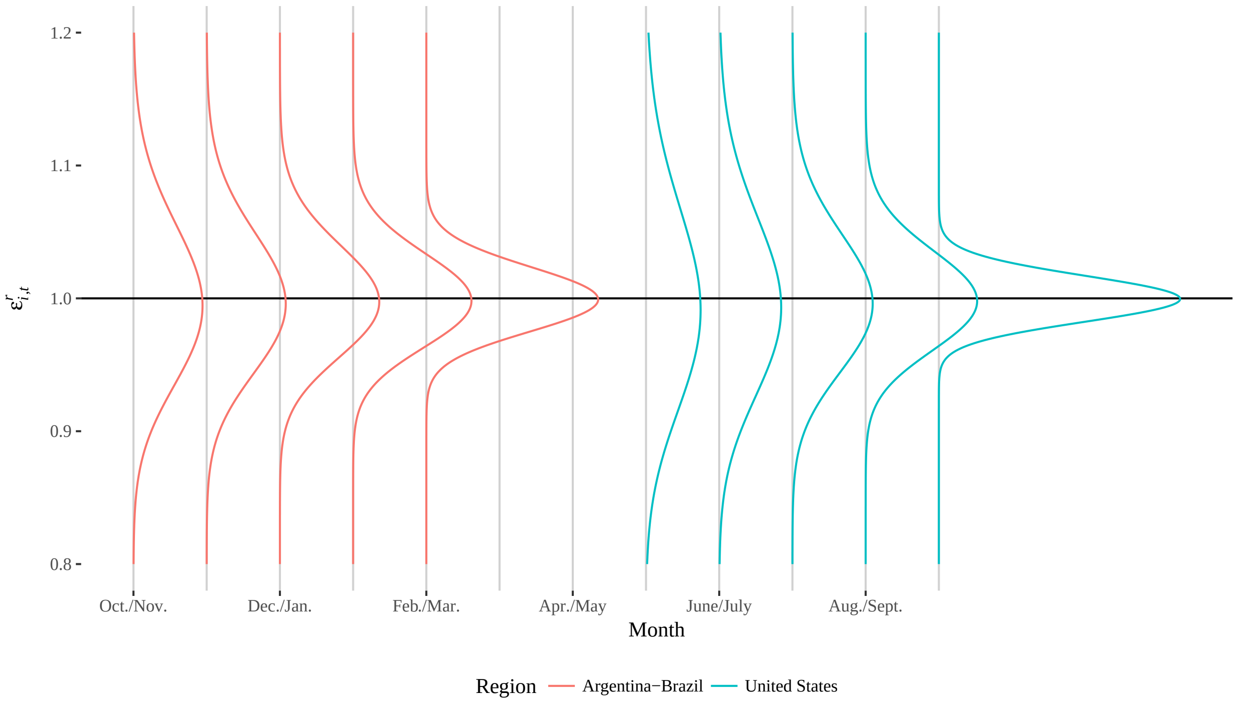

Figure 3: Density of the aggregate production shock conditional on information available at the month of forecast for previous shocks at their mean

and November forecasts are obtained as a simple average of the production forecasts by Conrad Leslie and Informa Economics (formerly Sparks Companies), selected because they were considered to be the most influential, they were widely reported and they are available over an extended period of time. For 2001–5, the forecasts are a simple average of Informa Economics estimate and the average analyst estimate reported by the Dow Jones Newswire survey. For 2006–12, the forecasts are the Dow Jones survey average. For 2013–5, the forecasts are the Bloomberg survey average. For 2016–7, the forecasts are from Reuters. See Egelkraut et al. (2003) and Good and Irwin (2006) for details about these private forecasts and their performance.

To identify the private shocks, we assume that the monthly shocks to production are the sum of two elements: a part observed by both private forecasting firms and USDA, denoted $\tilde{\eta}^r_{i,t}$, and a part observed by USDA only, denoted $\dot{\eta}^r_{i,t}$, so that the total shock is the sum of the two, $\eta^r_{i,t} = \tilde{\eta}^r_{i,t} + \dot{\eta}^r_{i,t}$. These two shocks follow normal distributions with means $\tilde{\mu}^r_i$ and $\dot{\mu}^r_i$, and standard deviations $\tilde{\sigma}^r_i$ and $\dot{\sigma}^r_i$, such that $\left(\sigma^r_i\right)^2 = \left(\tilde{\sigma}^r_i\right)^2 + \left(\dot{\sigma}^r_i\right)^2$, $\tilde{\mu}^r_i + \left(\tilde{\sigma}^r_i\right)^2/2 = 0$, and $\dot{\mu}^r_i + \left(\dot{\sigma}^r_i\right)^2/2 = 0$. We assume that private production forecasts, noted $\hat{\hat{Q}}$, build on previous public forecasts if available:

$$\hat{\hat{Q}}^r_{i,t} = Q^r_{i^r_\pi,t}\begin{cases}\exp\left(\sum_{j=i^r_\pi+1}^{i-1}\eta^r_{j,t} + \tilde{\eta}^r_{i,t}\right) & \text{if } i^r_\pi + 2 \le i \le i^r_\pi + 5,\\ \exp\left(\tilde{\eta}^r_{i,t}\right) & \text{if } i = i^r_\pi + 1.\end{cases} \tag{16}$$

Combining equations (5) and (16) gives $\exp\left(\dot{\eta}^r_{i,t}\right) = \hat{Q}^r_{i,t}/\hat{\hat{Q}}^r_{i,t}$: the size of the news shocks provided by USDA only can be obtained from the ratio of public and private projections. As before, we obtain that var $\left(\hat{Q}^r_{i,t}/\hat{\hat{Q}}^r_{i,t}\right) \approx \left(\dot{\sigma}^r_i\right)^2$.

This approach makes several simplifications. First, it assumes that private forecasts are immediately made public, which prevents assessing the benefits accruing to speculators from their privileged access to early information. Second, it assumes that there is one representative private forecast, while the reality is of a diversity of forecasts, which would give USDA forecasts a special role in decreasing the uncertainty related to the diversity of previsions. Third, and related, it considers that the only economic role of USDA forecasts is to improve the precision of information about fundamentals over a situation with only private forecasts. In the global games framework of Morris and Shin (2002), public information has another role: it allows agents with private information to coordinate while interacting strategically. Given the purpose of the paper, these simplifications allow us to assess the benefits of public information compared to private information without the need for a model detailing the timing of each shock and without accounting for heterogeneous sets of information.

The theoretical framework assumes that public forecasts are more precise than private forecasts. So, a first step before decomposing the news shocks in their private and public elements is to verify that it is the case. Table 4 displays the root mean square percentage error (RMSPE) for private and USDA forecasts using as final production value the January estimate. The forecasting error declines with time and is higher for private forecasts than for public forecasts.[11] This result confirms the model assumption. Using the ratio of public to private forecasts to identify the shock observed only by USDA, we cannot reject the null hypothesis that its mean is equal to 1. The standard deviation of this ratio gives $\dot{\sigma}^{\mathrm{US}}_i$ from which $\tilde{\sigma}^{\mathrm{US}}_i$ can be calculated by targeting the standard deviation of the total shock. Using this decomposition, the private forecasts predict between 61 and 77% of the total shocks variance.

Table 4: Informational content of private forecasts about U.S. harvest over 1970–2017

	RMSPE compared to January estimate			$\hat{Q}^{\mathrm{US}}_{i,t}/\hat{\hat{Q}}^{\mathrm{US}}_{i,t}$		
Month	Private (%)	USDA (%)	Difference (%)	Mean	SD (%)[a]	$\tilde{\sigma}^{\mathrm{US}}_i$ (%)[b]
July[c]	8.816	7.929	−0.887	1.000	1.897	3.123
August	6.312	6.207	−0.104	1.000	2.350	3.612
September	5.169	4.849	−0.320	1.004	1.738	3.181
October	3.280	2.677	−0.603	1.000	1.665	2.878
November	2.013	1.634	−0.379	1.002	1.081	1.351

Source: Authors' calculation based on forecasts provided by Scott Irwin.

Note: [a] Correspond to $\dot{\sigma}^{\mathrm{US}}_i$. [b] Calculated as $\sqrt{\left(\sigma^{\mathrm{US}}_i\right)^2 - \left(\dot{\sigma}^{\mathrm{US}}_i\right)^2}$. [c] Based on data over 1983–2017 and calculated using USDA yield forecasts multiplied by private or USDA acreage forecasts.

[11]The results would be similar with a measure of mean absolute percentage error except for August when the private forecasts would be slightly more precise than the public forecasts.

4 Quantitative analysis

4.1 Effects on market dynamics

We now turn to analysis of the effects of news shocks. We compare the model with news shocks described above, with a model without news shocks where the production shocks are degenerate and concentrated in the harvesting periods. Table 5 presents the descriptive statistics on the asymptotic distribution of prices and stocks for the models with and without news shocks.[12] The changes in mean prices are not represented in the table because they are very small (below 0.3%). News shocks affect the distribution of prices without changing their means.

Table 5: Descriptive statistics on the asymptotic distribution (all results in percentages)

		Prices					Stocks
		Standard deviation			Skewness	Kurtosis	Mean
Period	Shock	No news	News	Changes	Changes	Changes	Changes
Oct./Nov.	σ_1^{US}	17.25	14.64	−15.1	−25.7	−41.7	−1.4
Nov./Dec.	σ_2^{LAC}	17.28	15.25	−11.7	−34.4	−49.8	−1.7
Dec./Jan.	σ_3^{LAC}	17.30	16.15	−6.7	−42.7	−58.5	−2.5
Jan./Feb.	σ_4^{LAC}	17.33	16.74	−3.4	−45.8	−62.3	−4.2
Feb./Mar.	σ_5^{LAC}	17.42	17.67	1.4	−47.5	−65.4	−14.8
Mar./Apr.	σ_6^{LAC}	15.43	14.35	−7.0	−19.0	−30.7	−0.9
Apr./May		15.46	14.38	−7.0	−19.0	−30.7	−1.1
May/June		15.48	14.40	−7.0	−19.0	−30.7	−1.3
June/July	σ_9^{US}	15.51	14.91	−3.8	−26.8	−38.6	−1.7
July/Aug.	σ_{10}^{US}	15.53	15.87	2.2	−34.7	−48.3	−2.4
Aug./Sept.	σ_{11}^{US}	15.56	16.83	8.1	−36.2	−51.6	−4.1
Sept./Oct.	σ_{12}^{US}	15.70	18.00	14.7	−31.7	−48.1	−15.7
Inter-annual[a]		12.70	12.86	1.3	−24.1	−32.2	–
Intra-annual[b]		3.37	4.64	37.5	–	–	–

Source: Statistics calculated over 1,000,000 sample observations from the asymptotic distribution simulated with the model.

Note: [a] Annual price calculated as $\sum_{1=1}^{12} P_{i,t}/12$.

[b] Intra-annual standard deviation calculated as $\mathrm{E}\sqrt{\sum_{i=2}^{12}\left[\ln\left(P_{i,t}/P_{i-1,t}\right)-\ln\left(P_{12,t}/P_{1,t}\right)/11\right]^2/10}$.

Before analyzing the effects of introducing news shocks, let us first explain the pattern of seasonal price volatility in the absence of these shocks. Without news shocks, the standard deviation of seasonal prices does vary with the seasons but only a little (1.99 percentage point maximum), increasing by about 0.03 percentage points between seasons except after new harvests when there are small jumps. These variations are easily explained. Between a harvest and the month immediately before the next harvest, there is no new information, and stocks are always positive, so from

[12]See section B in the appendix for details of the numerical methods.

equation (1), the non-arbitrage condition of storage, successive prices are deterministic functions of the harvest-time price: $P_{i+1,t} = \left(P_{i,t} + k\bar{P}\right)/\beta$. If the only storage cost was the opportunity cost, the standard deviation and the mean price would increase between two periods by $1/\beta$ but the coefficient of variation would remain constant. The presence of additive storage costs breaks this multiplicative relationship, and the coefficient of variation changes slightly with the period.

The presence of news shocks changes seasonal price volatility a great deal. The differences between seasons increase with a maximum spread of 3.65 percentage points. In most seasons the volatility decreases significantly but not in February/March, July/August, August/September, or September/October; the volatility increases are the highest in the periods closest to the harvests (+14.7% in September/October preceding the U.S. harvest). As expected, between the seasons without news arrival, the same 0.03 percentage points deterministic increase in price volatility occurs as observed in the model without news shocks. The decrease in price volatility observed in the other seasons follows simple intuition: having advance information allows the market to adjust stock levels before the harvests which avoids abrupt changes at the harvest time. Thus, adjustments are smoothed over the growing seasons. However, the increased volatility immediately before the harvest is less intuitive. It is due to the fact that most of the aggregate productive shock has been observed before the harvest (during the 4-month growing season before the harvest, the news shocks aggregate to standard deviations of 7.49% for the U.S. and 5.56% for Argentina-Brazil). Therefore, although the harvest is one period later, the market is fairly confident about its size. On the other hand, the seasons immediately before harvests are the seasons when stocks are at their lowest levels, and so are seasons when the market is very sensitive to news because adjustment capacities are limited. The fact that the information on the harvest size has been moved to seasons when the market is less able to cope with shocks explains the increased volatility in these seasons.

The most dramatic effects of the introduction of news shocks can be observed in the higher-order moments: skewness and kurtosis. In all seasons, as expected in the presence of speculative storage, prices are skewed positively (not represented in table 5): prices are concentrated below the mean with occasional positive deviations which possibly are larger than the negative ones. News shocks reduce the skewness considerably (between 19% and 48%) by reducing the occurrence of high prices significantly. Similarly, kurtosis of the price distribution is reduced by at least a third (although the distribution remains leptokurtic) by the introduction of news shocks because of the decreased probability of extreme price events.

It is striking that this reduction in seasonal price volatility and occurrence of high prices is obtained despite a reduction in mean stock levels. Without news shocks, it is profitable to keep higher levels of stocks before the harvest because there is a non-negligible likelihood of a bad harvest, thus, keeping more stocks might be profitable. With news shocks, before the harvest the market has a pretty good idea of the coming harvest so the speculative motive is decreased. The reduction in stock levels in periods other than just before harvests seems less impressive because in these periods the mean stock levels are higher to smooth the bi-annual harvest over the whole year. However, in levels, not in proportions, the decrease is very similar for all periods. Stocks do not always reduce with news shocks. If early news shocks indicate a bad harvest, destocking is reduced compared to the situation without news which mitigates the price increase.

Table 5 also presents inter- and intra-annual price volatility statistics. Inter-annual price volatility is calculated by averaging seasonal prices over the U.S. marketing year. A similar approach is adopted for the World Bank Pink Sheets although the World Bank averages over the calendar year. The inter-annual price volatility is significantly lower than any of the seasonal price volatilities because averaging over seasons removes the intra-annual volatility, related to the non-overlapping crop cycles which allows producers in one hemisphere to react to observation of the other hemisphere's harvest. In the inter-annual model in appendix (section A) corresponding to the monthly model except that crop cycles are collapsed, the coefficient of variation of price is 16.77%, much higher than the 12.86% inter-annual price volatility of the monthly model without news because of lack of adjustments after one harvest. So, the inter-hemispheric supply response reduces price volatility by 3.91 percentage point. The stabilization benefits of the inter-hemispheric supply response are obtained even without news because we have assumed for simplicity that the crop cycles do not overlap. In reality, in some regions of Argentina-Brazil planting occurs while harvesting is not finished everywhere in the U.S. So, part of this reduction of price instability by the inter-hemispheric supply response could probably be related to the presence of news shocks. The size of the inter-hemispheric supply response is analyzed in Lybbert et al. (2014) but the present article is the first to quantify its effect on price volatility.

The level of inter-annual price volatility is only scarcely affected by the presence of news shocks despite the important changes to seasonal price volatility. This shows that one of the main effects of news shocks is to reorganize the price volatility across the seasons. However, among the higher-order moments of the distribution, the inter-annual and the seasonal results are similar: very strong reduction in skewness and kurtosis related to smaller occurrence of price spikes. An annual statistic calculated as an average would not make much sense in the context of stocks. Annual stocks usually are considered as the level at the beginning of the marketing year. Since in our model, the periods when there are non-zero probabilities of stockouts are February/March and September/October, the change in annual stock levels could be seen as the change in one of these seasons statistics: -14.8% or -15.7%.

Intra-annual price volatility is calculated as the mean over the asymptotic distribution of the bias-corrected root mean square of the difference between within-year month-to-month price changes and the yearly price changes: $\mathrm{E}\sqrt{\sum_{i=2}^{12}\left[\ln\left(P_{i,t}/P_{i-1,t}\right)-\ln\left(P_{12,t}/P_{1,t}\right)/11\right]^2/10}$. This measure of intra-annual price volatility increases significantly with news shocks. Indeed, in the absence of news shocks and neglecting storage costs, intra-annual volatility relates only related to the Argentina-Brazil harvest which is the only within-year shock in the model that would justify non-deterministic changes in prices. Outside harvest periods, without news shocks prices are deterministic functions of past prices, and are constant without any storage costs. With news shocks, this changes: with the exception of two periods without news shocks there are always some informational shocks, and prices adjust more regularly resulting in higher intra-annual price volatility.

With the introduction of news shocks, the first-order auto-correlation of the inter-annual price average increases by 8 percentage points (a result found also by Ng and Ruge-Murcia, 2000, and Osborne, 2004). This is interesting because this result is obtained despite lower stock levels and in the knowledge that higher stock levels usually are associated to higher auto-correlation levels. The intuition behind this result is that news shocks create links between different marketing years given that the conditions of the coming harvest are mostly known a few months in advance.

To summarize the consequences of news shocks for market dynamics: the presence of news shocks leads to a different repartition of seasonal price volatility with higher volatility just before harvests and lower volatility at other seasons but has little effect on inter-annual price volatility. However, by allowing stocks to adjust before the harvest, it also reduces stock levels, and the occurrence of price spikes.

4.2 Welfare effects

Table 6 decomposes the welfare gains in the three efficiency components identified in equation (9): consumer efficiency gains, storage costs, and production costs. Having news shocks compared to having only one production shock at the harvest increases welfare. Total gains amount to 0.38‰ of the value of annual steady-state consumption (column 1). To better understand the level of the welfare gains, we can compare them to the maximum gains attainable from having advance information about the coming harvests. For this we assume that the total production is fully known one period after planting, and that there are no other shocks following this. In this case, presented in column 2, total welfare gains equal 0.62‰ of the value of annual steady-state consumption. Thus, the actual news shocks are responsible for 61% of the potential gains. The small size of the gains, especially compared to the important changes in price dynamics shown in table 5, is explained by the simplicity of the model which includes few channels through which price volatility could affect welfare. Consumers are not risk averse, and so are mostly indifferent to price volatility (Gouel, 2014). Similarly, in the model there are very few reasons why production costs should be affected by news shocks since production decisions are taken before observing news shocks, and mean prices change very little. The only significant source of welfare gains is related to storage costs. Advance information means that there is no need for high levels of stocks. The reduction in storage costs explains most of the welfare gains but given that storage costs are small compared to the value of consumption, the eventual welfare gains are also low. That most of the welfare gains are related to the reduced inter-annual stock levels confirms the choice to use a storage model accounting for both inter- and intra-annual market dynamics rather than the approaches in Hayami and Peterson (1972) and Bradford and Kelejian (1977, 1978) which focus only on intra-annual storage and miss most of the effects of news shocks.

Table 6: Welfare changes (‰ of the value of annual steady-state consumption)

	All information[a]				Public information[b]
Welfare element	Actual news shocks (1)	All news in the first shock (2)	US news shocks only (3)	LAC news shocks only (4)	US news shocks only (5)
Consumer's efficiency gains	0.07	0.12	0.03	0.02	0.01
Storage costs	0.35	0.55	0.18	0.16	0.05
Production costs	−0.04	−0.05	−0.03	−0.02	−0.00
Total gains	0.38	0.62	0.18	0.17	0.06

Note: [a] Comparison between the news shocks structure described in the column header and a situation without news shocks. [b] For the U.S., comparison between a situation with actual news shocks and a situation with only the shocks that would be observed by private analysts, $\tilde{\eta}_{i,t}^{\mathrm{US}}$; for Argentina-Brazil no news shocks.

Given that most welfare gains are related to reductions in storage costs, it is reasonable to express them as a proportion of storage costs without news shocks. In this case, the total welfare gains from news amount to 2.31% of storage costs. However, we should not expect all storage decisions to be affected by the presence of news. In a storage model with intra-annual dynamics, most storage costs are incurred for smoothing consumption between harvests. These irreducible storage costs can be evaluated from the steady-state values of stock levels which are determined by assuming an absence of risk. If we express welfare gains in proportion to "risky" storage costs (storage costs minus their steady-state value), these welfare gains represent 17.37% of these costs. Thus, a sizable portion of the storage costs generated by uncertainty are reduced by the provision of news shocks. These gains are consistent with the reduction in ending stocks observed in table 5.

Since we had to make more assumptions to identify the size of the news shocks for Argentina-Brazil than for the U.S., it is interesting to calculate the value of public information separately for each region. This is done in the columns 3 and 4 of table 6 by comparing a model where only one region has news shocks with the counterfactual without news shocks for any region. The welfare gains are very similar across regions, slightly smaller than half of the gains with both regions. The U.S. share of production of 42% should have commanded lower welfare gains than for Argentina-Brazil, but this is counterbalanced by the higher precision of the news shocks on the U.S. harvest.

The previous welfare results assume that USDA is the only source of seasonal forecasts and use as counterfactual a situation without advance information at all, which overestimates the benefits since some information would be collected privately even in the absence of USDA. We account now for the presence of private forecasts. This can only be done for the shocks related to the U.S. harvest for which we have information about private forecasts. It is difficult to choose the appropriate counterfactual, because private forecasts would be unlikely to stay the same in the absence of USDA forecasts. Private forecasters could increase their data collection and their accuracy, or on the contrary, reduce their accuracy because their forecasts partly build on public data that would be too costly to collect privately. For example after July, Conrad Leslie and Sparks Companies used USDA acreage estimates for their production forecasts (Egelkraut et al., 2003), focusing their efforts on yield prediction.

In section 3.2.2, we have decomposed U.S. monthly shocks into a part observed by private analysts, $\tilde{\eta}^{r}_{i,t}$, and a part observed by USDA only, $\dot{\eta}^{r}_{i,t}$. For the counterfactual with only private information, we assume that $\dot{\eta}^{r}_{i,t}$ is not observed, which moves part of the shocks to harvest time when production is observable by all and we use the last column of table 4 to calibrate $\tilde{\sigma}^{\mathrm{US}}_{i}$. Note that to maintain constant the standard deviation of the aggregate shock, the observability of production at harvest implies $\tilde{\sigma}^{\mathrm{US}}_{1} = \sqrt{0.0769^2 - \sum_{i=9}^{12} \left(\tilde{\sigma}^{\mathrm{US}}_{i}\right)^2}$, which means that we cannot make use of the last row of table 4 related to the shocks at harvest.

Column 5 of table 6 presents the results in which a situation with the actual U.S. news shocks is compared to situations with only private shocks. For Argentina-Brazil, a situation without news shocks is assumed. These results allow us to estimate the contribution of USDA information accounting for the information provided by private analysts. The results can be compared with column 3 to estimate the share of the welfare gains related to private and public forecasts in total gains. The information uniquely provided by USDA accounts for one-third of the total welfare gains, consistent with the observation from table 4 that the private forecasts predict around two-thirds of the total shocks variance.

To illustrate the welfare gains, we use a price of $350 per ton of soybeans, and world production of 320 million tons based on recent data. World production then is valued at $112 billion, and the total benefits from news shocks are valued at $42.6 million. If we focus just on the benefits of the USDA reports on the U.S. harvest, accounting for the presence of private forecasts the welfare benefits are $6.7 million. For a comparison, the combined costs of the three main services involved in the elaboration of the WASDE reports—the National Agricultural Statistics Service, the Foreign Agricultural Service, and the World Agricultural Outlook Board—were $211 million in 2013 (C-FARE, 2016, Table 1) for missions which include other commodities than soybeans, demand sides of the balance sheets, which are not considered here, and several reports other than the monthly WASDE reports.

4.3 Improving the quality of advance information

In this section, we use the model to assess the effects of small improvements in the informational content of news shocks. These counterfactuals allow us to quantify what would be the benefits of improving the current crop forecasts (if it is possible). For example, by making available much more data about earth observation at higher frequencies, the CubeSat revolution is providing a basis for new approaches to crop production forecast, and in the recent years, new firms are relying on these big data to propose what they claim to be improved crop production forecasts. We consider four counterfactuals in which the last shock to the U.S. harvest has its standard deviation reduced by 1 percentage point to 0.73% by assuming that this part of the shock can be observed in one of the four earlier growing seasons. To maintain constant the standard deviation of the aggregate shock, $\epsilon_{1,t}^{\mathrm{US}}$, the standard deviation of an earlier shock is increased according to $\sigma_i^{\mathrm{US}} = \sqrt{0.0769^2 - \sum_{j\in\{1,9..12\},j\neq i}\left(\sigma_j^{\mathrm{US}}\right)^2}$ for $i \in \{9\,..\,12\}$. The distribution means are adjusted to maintain unitary means for the production shocks.

Table 7 presents these counterfactual results. As can be expected, welfare gains are higher when the shock is moved to an earlier month. The maximum gains are achieved when moving the shock to η_9^{US} (column 3). Compared to the gains coming from news shocks on the US harvest (column 3 of table 6), the welfare gains amounts to an 4% increase. Regarding price volatility, table 7 presents the pattern described previously. Moving the information earlier in time barely affects inter-annual volatility. The effect on intra-annual volatility is also small but shows a clear pattern of increased volatility the later the information arrives (it increases from column 3 to 6).

In these simulations, price volatility decreases in the first five months and increases after. The increases are limited (below 0.05 percentage points) before the period to which the shock is moved. The increase is more significant at the period to which the shock is moved and until the U.S. harvest. In column 3, price volatility increases by 0.15 percentage points in June/July when additional information is received. Price volatility increases similarly in the following months before the harvest. In column 6, the shock is just moved one month earlier. Price volatility increases by 0.02 percentage points before September/October and by 0.29 percentage points in September/October. These simulations show us that news shocks move the volatility earlier in times, when the information is received, and decrease it at harvest and later months because of a more efficient allocation of

Table 7: Standard deviation and welfare gains with various structures of news shocks

Period	Shock	No news (1)	News (2)	Reduction of σ_1^{US} and increase of σ_9^{US} (3)	σ_{10}^{US} (4)	σ_{11}^{US} (5)	σ_{12}^{US} (6)
		Standard deviation (%)					
Month							
Oct./Nov.	σ_1^{US}	17.25	14.64	14.34	14.37	14.41	14.49
Nov./Dec.	σ_2^{LAC}	17.28	15.25	14.97	15.00	15.03	15.11
Dec./Jan.	σ_3^{LAC}	17.30	16.15	15.89	15.91	15.94	16.01
Jan./Feb.	σ_4^{LAC}	17.33	16.74	16.49	16.51	16.55	16.61
Feb./Mar.	σ_5^{LAC}	17.42	17.67	17.44	17.46	17.49	17.55
Mar./Apr.	σ_6^{LAC}	15.43	14.35	14.40	14.40	14.39	14.38
Apr./May		15.46	14.38	14.43	14.42	14.41	14.40
May/June		15.48	14.40	14.45	14.45	14.44	14.42
June/July	σ_9^{US}	15.51	14.91	15.05	14.95	14.94	14.93
July/Aug.	σ_{10}^{US}	15.53	15.87	16.02	16.04	15.90	15.89
Aug./Sept.	σ_{11}^{US}	15.56	16.83	16.99	17.01	17.03	16.84
Sept./Oct.	σ_{12}^{US}	15.70	18.00	18.19	18.21	18.24	18.29
Inter-annual[a]		12.70	12.86	12.84	12.84	12.84	12.84
Intra-annual[b]		3.37	4.64	4.66	4.66	4.67	4.68
			Total welfare gains (‰)				
			0.383	0.398	0.396	0.393	0.389

Source: Statistics calculated over 1,000,000 sample observations from the asymptotic distribution simulated with the model.
Note: [a] Annual price calculated as $\sum_{1=1}^{12} P_{i,t}/12$.
[b] Intra-annual standard deviation calculated as $\mathrm{E}\sqrt{\sum_{i=2}^{12}\left[\ln\left(P_{i,t}/P_{i-1,t}\right)-\ln\left(P_{12,t}/P_{1,t}\right)/11\right]^2/10}$.

stocks. In consequence, news shocks about the U.S. harvest increase price volatility during the U.S. growing season and decrease volatility during the Argentina-Brazil growing season. Conversely, news shocks about the Argentina-Brazil harvest increase price volatility during the Argentina-Brazil growing season and decrease volatility during the U.S. growing season. When both sources of news shocks are combined, price volatility decreases everywhere except just before the harvests. These results help better explain that in table 5 the presence of news increases volatility more before the U.S. harvest than before the Argentina-Brazil harvest. The news about the Argentina-Brazil harvest reduces price volatility during the U.S. growing season, but since the news about the U.S. crops are quite precise the volatility tends to increase before the U.S. harvest and very significantly one month before. On the other hand, the precision of the news about the Argentina-Brazil harvest is much lower and the reduction of price volatility caused by the news about the U.S. harvest tends to dominate.

4.4 Sensitivity analysis

This section analyzes the sensitivity of the main results to the 3 key behavioral parameters: demand elasticity, supply elasticity, and storage cost. For each parameter, we consider two scenarios: a halving and a doubling of the benchmark parameter, which should provide reasonable lower and upper bounds of the likely values. Table 8 presents the results of the sensitivity analysis with for parsimony a focus on the most important results. To help the comparison, column (1) reproduces the results under the default parameters.

Table 8: Sensitivity analysis

	Default	Demand elasticity		Supply elasticity		Storage cost	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
α^D: Demand elasticity	−0.2	−0.1	−0.4	−0.2	−0.2	−0.2	−0.2
α^Q: Supply elasticity	0.4	0.4	0.4	0.2	0.8	0.4	0.4
k: Storage cost (% of $\bar{P}$)	0.56	0.56	0.56	0.56	0.56	0.28	1.12
Price SD in Sept./Oct. under news (%)	18.00	28.16	11.29	19.31	16.75	14.89	22.10
Change in price SD in Oct./Nov. (%)	−15.1	−22.0	−16.4	−16.2	−13.6	−14.6	−17.3
Change in price SD in Feb./Mar. (%)	1.4	−5.6	−2.7	−0.8	4.7	2.4	−2.2
Change in price SD in Sept./Oct. (%)	14.7	13.7	7.9	13.2	16.9	16.0	10.1
Change in inter-annual price SD (%)	1.27	−2.76	−2.76	−0.68	3.84	1.84	−2.08
Ratio of total welfare gains to storage costs (%)	2.31	1.96	2.30	2.66	1.91	2.28	2.27
Ratio of total welfare gains to risky storage costs (%)	17.37	10.97	29.59	19.85	14.42	13.53	24.23

In the first row after the parameters, we can first verify with the price standard deviation in September/October (the month with the highest volatility) under news that the parameters have the expected effect. The results are as expected. Demand elasticity has a strong influence on price volatility with an increase to 28% if $\alpha^D = -0.1$ and a decrease to 11% if $\alpha^D = -0.4$. Supply elasticity has a smaller effect since supply takes time to adjust. Lastly, a higher storage cost makes speculation more costly, so storers buy at lower prices and sell at higher prices leading to more price volatility.

Three changes in price volatility after the introduction of news are reported in table 8: October/November, February/March, and September/October, corresponding respectively to the U.S. harvest when the decrease in the volatility is the highest, the period before the Argentina-Brazil harvest when the volatility increases slightly in the benchmark, and the period before the U.S. harvest when the volatility increases the most. For the two extremes, the same pattern holds with a strong decrease in October/November and a strong increase in September/October. However, in February/March, the sign of the change depends on the parameters' value. Remember that there are two effects at play: the news shocks on the U.S. harvest decrease volatility during the Argentina-Brazil growing season and the news shocks on the Argentina-Brazil harvest increase it. Depending on the respective strength of the two effects, volatility may increase or decrease in February/March.

The results on inter-annual price volatility are inconsistent across the parameterizations: it decreases for some parameters and increases for others. The changes are always much lower than the changes in monthly price volatilities, supporting the conclusion that the main effect of news shocks is of redistributing volatility during the season with limited effect on inter-annual volatility.

Lastly, regarding the welfare gains, they are quite stable around 2% if expressed in ratio to storage costs. In ratio to risky storage costs, there are more variations because in this case both the numerator and the denominator may vary significantly. Since this expression of welfare tends to be related to the decrease in the level of the end-of-season stock, it shows that the presence of news shocks has a robust effect on decreasing end-of-season stocks.

5 Conclusion

In this article which uses the example of the global soybean market, we have shown how news shocks that provide advance information about future production of a storable commodity affect market dynamics, compared to a situation where production is known only when it is realized. The results were obtained by building a rational expectations storage model able to account for intra-annual and inter-annual dynamics, and running counter-factual simulations. The monthly crop production forecasts, from USDA and private firms, were used to calibrate realistic news shocks in the model. These shocks matter greatly for market dynamics because based on the information they provide, storers adjust their stocks before the harvest to be more consistent with the newly expected market tightness. These early stocks adjustments reduce the need to carry large inter-seasonal stocks and can reduce them significantly. In addition, if a poor harvest is forecast, storers reduce their destocking before the harvest which in turn reduces the occurrence of price spikes, although the overall price volatility tends to be little affected by the presence of news shocks. The reduction in stock levels allowed by news shocks explains most of the welfare gains from having advance information which amount to 2% of the value of storage costs.

This study has dealt with only a small part of the informational issues that occur in commodity markets, and thus, with only a small part of the potential benefits of public information in these markets. For example, there are problems related to reporting stocks. Stock levels are not directly observable by market participants but are reported regularly by government agencies and international organizations. These statistics although crucial for the market are recognized as not wholly reliable: the preciseness of their estimation is questionable since they are based either on surveys, and more frequently outside the U.S., as residuals in commodity balances. In our storage model framework, this means that agents do not observe the state variables with precision, and have to make their decisions based on incomplete information. The reporting of stock levels is another, and perhaps as economically important problem as advance information on the coming harvests. Another feature of our approach that might have led to underestimation of the benefits from public information is that we developed a simple model with few margins for adjustment to new information. An extension to our model that includes agents taking costly decisions during the growing season would increase the benefits: for example, planting crops that are close substitutes or buying livestock feed.

Beyond the question of the cost/benefit analysis of public information, this research may have other policy implications related to the opportunity for price stabilization policies. Following the 2007/08 food crisis, we observed renewed interest in food price stabilization policies (Gouel, 2014), in particular to protect consumers from price spikes. These policies frequently involve the combination of storage and trade policies which can be very costly fiscally (Gouel and Jean, 2015). The results in this paper show that provision of public information during the growing season is effective to reduce the occurrence of price spikes without the need for public interventions in the market. Since market information systems in developing countries are generally less developed than in the U.S., their development should be considered before considering any price stabilization policies.

The findings in this article on seasonal dynamics have consequences also for future research on the annual storage model. They illustrate some of challenges related to the estimation of this model which are acknowledged only superficially in the literature in the tradition of Deaton and Laroque (1992, 1996). Guerra et al. (2015) is the only work that touches on this issue by showing important differences between estimating an annual model using harvest-time prices and annual averages, which is consistent with our results: the price volatility of the annual average is very different from the volatility in every season. An even more difficult issue is that, calibrated with the same parameters, an annual storage model that implicitly neglects news shocks, and an intra-annual model with news shocks will present very different price dynamics. The second model will show with far fewer price spikes because of the stock adjustments made possible before harvest. Thus, estimation of an annual model based on prices generated by a more realistic seasonal model would likely not recover the original structural parameters. It remains to be seen which parameters are the most affected by this misspecification.

Appendix

A Inter-annual storage model

Here, we present the inter-annual storage model corresponding to the seasonal model used in our study. This section has two purposes. First, to lay down the equations of the inter-annual model used to generate some results in section 4. Second, to prove that the steady-state parameters, $\bar{D}$ and $\bar{P}$, used in the equations of the seasonal model have a counterpart in the annual model.

The inter-annual model is defined by four equilibrium equations:

$$S_t : \beta^{12} \mathrm{E}_t\, P_{t+1} - P_t - 12k\bar{P} \le 0, = 0 \text{ if } S_t > 0, \tag{A1}$$

$$Q_t^r : \beta^{12} \mathrm{E}_t \left(P_{t+1}\epsilon_{t+1}^r\right) = \Psi_I^{r\prime}\left(Q_t^r\right) \text{ for } r \in \{\text{US,LAC}\}, \tag{A2}$$

$$P_t : A_t = 12D\left(P_t\right) + S_t, \tag{A3}$$

and one transition equation:

$$A_t : A_t = S_{t-1} + \sum_{r\in\{\text{US,LAC}\}} \epsilon_t^r Q_{t-1}^r. \tag{A4}$$

A few adjustments are required to transform the model to a purely inter-annual model. Since the production decision is taken one year before the harvest, the crop is assumed to grow for one year, involving a discounting that is different on the left-hand side of equation (A2) compared to the left-hand side of equation (3). Because the discounting is increased, the marginal cost is increased similarly with

$$\Psi_I^r\left(Q^r\right) = \frac{\beta^{12}\bar{P}}{\left(\theta^r\bar{D}\right)^{1/\alpha^Q}} \frac{\left(Q^r\right)^{1+1/\alpha^Q}}{1+1/\alpha^Q}. \tag{A5}$$

It can be verified easily that this model's steady state is $A^{ss} = \bar{D}$, $S^{ss} = 0$, $Q^{r,ss} = \theta^r\bar{D}$, and $P^{ss} = \bar{P}$.

B Numerical methods[13]

The rational expectations storage model does not allow a closed-form solution; it must be approximated numerically. The numerical algorithm used here is based on a projection method with a collocation approach and is inspired by Fackler (2005) and Miranda and Glauber (1995). The results were obtained using MATLAB R2017b and solved using the solver for nonlinear rational expectations models RECS version 0.7 (Gouel, 2018). The numerical method is explained below but for a complete picture see the program code (RECS code is publicly available and the code for this paper is available upon request). The method presented is for a typical rational expectations model with informational subperiods, and not for the specific model equations. Following Fackler

[13] This section draws in part on the appendix in Gouel and Jean (2015).

(2005), rational expectations problems can be expressed using three groups of equations. For ease of exposition, below we exclude the year index, and index only by seasons. State variables s_i are updated through a transition equation:

$$s_i = g_{i-1}(s_{i-1}, x_{i-1}, e_i), \tag{A6}$$

where x_i are response variables and e_i are stochastic shocks. Response variables are defined by solving a system of complementarity equilibrium equations:

$$f_i(s_i, x_i, z_i) \leq 0, = 0 \text{ if } x_i > \underline{x}_i. \tag{A7}$$

Response variables can have lower bounds, $\underline{x}_i$. In cases where response variables have no bounds, equation (A7) simplifies to a traditional equation: $f_i(s_i, x_i, z_i) = 0$. z_i is a variable representing the expectations about the next period and is defined by

$$z_i = \mathrm{E}_i \left[h_i \left(s_i, x_i, e_{i+1}, s_{i+1}, x_{i+1} \right) \right]. \tag{A8}$$

There is one difference between the generic model defined by equations (A6)–(A8) and the model in the article: in the generic model, expectations are only function of next-period variables, while in the model producers takes the planting decision based on variables expected five periods before, $\mathrm{E}_{i^r_\pi,t}(P_{i^r_\pi+5,t}\epsilon^r_{i^r_\pi+5,t})$ in equation (3). We can use the law of iterated expectations to make the connection between the two expressions. For example:

$$\mathrm{E}_{i^r_\pi,t}(P_{i^r_\pi+5,t}\epsilon^r_{i^r_\pi+5,t}) = \mathrm{E}_{i^r_\pi,t}(\mathrm{E}_{i^r_\pi+1,t}(\mathrm{E}_{i^r_\pi+2,t}(\mathrm{E}_{i^r_\pi+3,t}(\mathrm{E}_{i^r_\pi+4,t}(P_{i^r_\pi+5,t}\epsilon^r_{i^r_\pi+5,t}))))). \tag{A9}$$

So, it suffices to add some auxiliary response variables to keep track of the expectations of variables that are more than one period ahead.

One way to solve this problem is to find a function that is a good approximation of the behavior of response variables. We consider a cubic spline approximation of the response variables,

$$x_i \approx \Phi_i \left(s_i, \theta_i \right), \tag{A10}$$

where θ_i are the parameters defining the spline approximation. To calculate this spline, we discretize the state space (using 150 nodes for availability and for expected production), and the spline has to hold exactly for all points of the grid.

The expectations operator in equation (A8) is approximated through 5-point Gaussian quadratures using functions from the CompEcon toolbox (Miranda and Fackler, 2002). The Gaussian quadrature defines a set of pairs $\left\{e^l_i, w^l_i\right\}$ in which e^l_i represents a possible realization of shocks, and w^l_i the associated probability. Using this discretization and equations (A6) and (A8)–(A10), we can express the equilibrium equation (A7) as

$$f_i \left(s_i, x_i, \sum_l w^l_i h_i \left(s_i, x_i, e^l_{i+1}, g_i \left(s_i, x_i, e^l_{i+1} \right), \Phi_{i+1} \left(g_i \left(s_i, x_i, e^l_{i+1} \right), \theta_{i+1} \right) \right) \right) \leq 0, = 0 \text{ if } x_i > \underline{x}_i. \tag{A11}$$

For a given spline approximation, θ_{i+1}, and a given s_i, equation (A11) is a function of x_i and can be solved using a mixed complementarity solver.

Once all the above elements are defined, we can proceed to the algorithm, which runs as follows:

Step 1. Initialization step. Choose an initial spline approximation, θ_1^n with $n = 1$, based on a first-guess (for example, the steady-state values).

Step 2. Time iteration step for subperiods. For $i = I, \ldots, 1$ do

Step 2.1. Equation solving step. For each point of the grid of state variables, s_i^j, solve equation (A11) for x_i^j using a mixed complementarity solver:

$$f_i\left(s_i^j, x_i^j, \sum_l w_i^l h_i\left(s_i^j, x_i^j, e_{i+1}^l, g_i\left(s_i^j, x_i^j, e_{i+1}^l\right), \Phi_{i+1}\left(g_i\left(s_i^j, x_i^j, e_{i+1}^l\right), \theta_{i+1}^n\right)\right)\right) \leq 0, = 0 \text{ if } x_i^j > \underline{x}_i. \tag{A12}$$

Step 2.2. Approximation step. Update the spline approximation using the new values of response variables, $x_i = \Phi_i\left(s_i, \theta_i^n\right)$.

Step 3. Terminal step. If $n = 1$ or $\|\theta^{n+1} - \theta^n\|_2 \geq 10^{-8}$ then increment n to $n + 1$ and go to Step 2.

Once the rational expectations equilibrium is identified, the spline approximation of the decision rules can be used to simulate the model.

To ensure precise solutions, the simulations are accomplished by solving the equilibrium equation (A11) at each iteration using the policy rules approximated by splines only to approximate the expectations. Regarding the welfare results, since welfare terms are expressed as recursive equations such as equation (10), they are calculated by value function iterations. The value function iterations generate a function that represents welfare as a function of the state variables. This function then is applied to the simulated observations. Welfare is calculated as the average of all welfare values in the first season.

Statistics on the asymptotic distribution are calculated over 1,000,000 observations from random outcomes of the stochastic variables, obtained by simulating 5,000 paths for 220 years, and after discarding for each path the first 20 years as burn-in period. The aggregate production shocks are the same for all scenarios. This is ensured first by drawing shocks for the model with news, and then multiplying them to obtain the aggregate shocks used in the scenarios without news or with all news concentrated in the first period after planting.

C Example of a WASDE report

August 2016

WASDE - 556 - 28

World Soybean Supply and Use 1/
(Million Metric Tons)

		Beginning Stocks	Production	Imports	Domestic Crush	Domestic Total	Exports	Ending Stocks
2014/15								
World 2/		61.96	319.78	123.74	263.26	300.90	126.13	78.46
United States		2.50	106.88	0.90	50.98	54.96	50.14	5.19
Total Foreign		59.46	212.90	122.84	212.29	245.94	75.99	73.27
Major Exporters 3/		41.43	170.05	0.32	84.30	91.56	68.70	51.54
Argentina		25.27	61.40	0.00	40.02	44.18	10.57	31.92
Brazil		16.02	97.20	0.31	40.44	43.41	50.61	19.50
Paraguay		0.13	8.15	0.01	3.65	3.69	4.49	0.12
Major Importers 4/		15.22	15.39	105.49	97.91	117.17	0.29	18.64
China		13.88	12.15	78.35	74.50	87.20	0.14	17.03
European Union		0.62	1.83	13.42	13.60	15.07	0.12	0.69
Japan		0.23	0.23	3.00	2.15	3.28	0.00	0.18
Mexico		0.12	0.35	3.82	4.18	4.21	0.00	0.07
2015/16 Est.								
World 2/		78.46	312.67	131.88	278.65	317.20	132.80	73.00
United States		5.19	106.93	0.68	51.71	54.69	51.17	6.95
Total Foreign		73.27	205.74	131.20	226.94	262.52	81.64	66.05
Major Exporters 3/		51.54	164.00	1.06	89.28	96.74	73.60	46.26
Argentina		31.92	56.50	0.60	44.25	48.60	10.50	29.92
Brazil		19.50	96.50	0.45	40.70	43.70	56.60	16.15
Paraguay		0.12	9.00	0.01	4.10	4.14	4.80	0.18
Major Importers 4/		18.64	15.25	110.93	106.29	126.99	0.36	17.46
China		17.03	11.60	83.00	81.80	95.50	0.15	15.98
European Union		0.69	2.26	13.60	14.00	15.58	0.15	0.82
Japan		0.18	0.24	3.25	2.40	3.55	0.00	0.12
Mexico		0.07	0.33	3.95	4.25	4.29	0.00	0.06
2016/17 Proj.								
World 2/	Jul	72.17	325.95	136.02	289.23	328.78	138.26	67.10
	Aug	73.00	330.41	136.62	289.89	329.82	138.97	71.24
United States	Jul	9.54	105.60	0.82	52.39	55.79	52.25	7.90
	Aug	6.95	110.50	0.82	52.80	56.22	53.07	8.97
Total Foreign	Jul	62.63	220.35	135.21	236.84	272.99	86.00	59.20
	Aug	66.05	219.91	135.81	237.09	273.59	85.90	62.27
Major Exporters 3/	Jul	42.61	172.00	0.36	89.25	96.93	77.79	40.26
	Aug	46.26	172.17	0.56	89.25	96.93	77.99	44.08
Argentina	Jul	27.02	57.00	0.05	44.30	48.75	10.65	24.67
	Aug	29.92	57.00	0.30	44.30	48.75	10.65	27.82
Brazil	Jul	15.45	103.00	0.30	40.50	43.60	59.70	15.45
	Aug	16.15	103.00	0.25	40.50	43.60	59.70	16.10
Paraguay	Jul	0.13	9.00	0.01	4.20	4.25	4.75	0.14
	Aug	0.18	9.17	0.01	4.20	4.25	4.95	0.16
Major Importers 4/	Jul	17.68	16.08	114.18	110.82	131.64	0.34	15.97
	Aug	17.46	16.07	114.58	111.32	132.59	0.34	15.19
China	Jul	16.23	12.20	87.00	87.00	100.80	0.15	14.48
	Aug	15.98	12.20	87.00	87.00	101.20	0.15	13.83
European Union	Jul	0.59	2.45	12.60	13.30	14.87	0.15	0.62
	Aug	0.82	2.44	13.00	13.80	15.39	0.15	0.72
Japan	Jul	0.24	0.24	3.10	2.20	3.33	0.00	0.26
	Aug	0.12	0.24	3.10	2.20	3.33	0.00	0.14
Mexico	Jul	0.06	0.37	4.00	4.28	4.33	0.00	0.11
	Aug	0.06	0.37	4.00	4.28	4.33	0.00	0.11

1/ Data based on local marketing years except Argentina and Brazil which are adjusted to an October-September year. 2/ World imports and exports may not balance due to differences in local marketing years and to time lags between reported exports and imports. Therefore, world supply may not equal world use. 3/ Argentina, Brazil, Paraguay, and Uruguay. 4/ China, European Union, Japan, Mexico, and Southeast Asia (includes Indonesia, Malaysia, Philippines, Vietnam, and Thailand).

Figure A1: Example of a WASDE report on soybean: Page 28 of the August 12th 2016 pdf report

References

Abbott, P., Boussios, D. and Lowenberg-DeBoer, J. (2016). *Valuing Public Information in Agricultural Commodity Markets: WASDE Corn Reports*. Conference paper, Agricultural and Applied Economics Association, 2016 Annual Meeting, July 31–August 2, Boston, Massachusetts.

Adjemian, M. K. (2012a). Emergence and impact of USDA's WASDE report. *Amber Waves*, June 05.

Adjemian, M. K. (2012b). Quantifying the WASDE announcement effect. *American Journal of Agricultural Economics*, 94(1), 238–256.

Adjemian, M. K., Johansson, R., McKenzie, A. and Thomsen, M. (forthcoming). Was the missing 2013 WASDE missed? *Applied Economic Perspectives and Policy*.

Arezki, R., Ramey, V. A. and Sheng, L. (2017). News shocks in open economies: Evidence from giant oil discoveries. *The Quarterly Journal of Economics*, 132(1), 103–155.

Beaudry, P. and Portier, F. (2014). News driven business cycles: Insight and challenges. *Journal of Economic Literature*, 52(4), 993–1074.

Bontems, P. and Thomas, A. (2000). Information value and risk premium in agricultural production: The case of split nitrogen application for corn. *American Journal of Agricultural Economics*, 82(1), 59–70.

Bradford, D. F. and Kelejian, H. H. (1977). The value of information for crop forecasting in a market system: Some theoretical issues. *The Review of Economic Studies*, 44(3), 519–531.

Bradford, D. F. and Kelejian, H. H. (1978). The value of information for crop forecasting with bayesian speculators: Theory and empirical results. *The Bell Journal of Economics*, 9(1), 123–144.

C-FARE (2016). *From Farm Income to Food Consumption: Valuing USDA Data Products*. Report, Council on Food, Agricultural & Resource Economics.

Cafiero, C., Bobenrieth, E. S. A., Bobenrieth, J. R. A. and Wright, B. D. (2011). The empirical relevance of the competitive storage model. *Journal of Econometrics*, 162(1), 44–54.

Chambers, M. J. and Bailey, R. E. (1996). A theory of commodity price fluctuations. *The Journal of Political Economy*, 104(5), 924–957.

Deaton, A. and Laroque, G. (1992). On the behaviour of commodity prices. *The Review of Economic Studies*, 59(1), 1–23.

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

Egelkraut, T. M., Garcia, P., Irwin, S. H. and Good, D. L. (2003). An evaluation of crop forecast accuracy for corn and soybeans: USDA and private information agencies. *Journal of Agricultural and Applied Economics*, 35(1), 79–95.

Fackler, P. L. (2005). A MATLAB solver for nonlinear rational expectations models. *Computational Economics*, 26(2), 173–181.

FAO (2017). FAOSTAT database. Retrieved May 25, 2017 from http://www.fao.org/faostat/en.

Galvão, A. B. (2017). Data revisions and DSGE models. *Journal of Econometrics*, 196(1), 215–232.

Garcia, P., Irwin, S. H., Leuthold, R. M. and Yang, L. (1997). The value of public information in commodity futures markets. *Journal of Economic Behavior & Organization*, 32(4), 559–570.

Garcia, P., Irwin, S. H. and Smith, A. (2015). Futures market failure? *American Journal of Agricultural Economics*, 97(1), 40–64.

Good, D. L. and Irwin, S. H. (2006). *Understanding USDA Corn and Soybean Production Forecasts: Methods, Performance and Market Impacts over 1970–2005*. Project Research Report 2006-01, AgMAS.

Gouel, C. (2014). Food price volatility and domestic stabilization policies in developing countries. In J.-P. Chavas, D. Hummels and B. D. Wright (eds.) *The Economics of Food Price Volatility*, National Bureau of Economic Research conference report, chapter 7, (pp. 261–306). Chicago, IL: University of Chicago Press.

Gouel, C. (2018). RECS: MATLAB solver for rational expectations models with complementarity equations – version 0.7. DOI: 10.5281/zenodo.1136095.

Gouel, C. and Jean, S. (2015). Optimal food price stabilization in a small open developing country. *World Bank Economic Review*, 29(1), 72–101.

Gouel, C. and Legrand, N. (2017). Estimating the competitive storage model with trending commodity prices. *Journal of Applied Econometrics*, 32(4), 744–763.

Guerra, E. A., Bobenrieth, E. S. A., Bobenrieth, J. R. A. and Cafiero, C. (2015). Empirical commodity storage model: The challenge of matching data and theory. *European Review of Agricultural Economics*, 42(4), 607–623.

Haile, M. G., Kalkuhl, M. and von Braun, J. (2016). Worldwide acreage and yield response to international price change and volatility: A dynamic panel data analysis for wheat, rice, corn, and soybeans. *American Journal of Agricultural Economics*, 98(1), 172–190.

Harrell, F. E. (2001). *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. Springer Series in Statistics, New York: Springer.

Hayami, Y. and Peterson, W. (1972). Social returns to public information services: Statistical reporting of U.S. farm commodities. *The American Economic Review*, 62(1/2), 119–130.

Hendricks, N. P., Janzen, J. P. and Smith, A. (2015). Futures prices in supply analysis: Are instrumental variables necessary? *American Journal of Agricultural Economics*, 97(1), 22–39.

IMF (2014). *World Economic Outlook: Recovery Strengthens, Remains Uneven*. Washington, DC: International Monetary Fund.

Joseph, K., Irwin, S. H. and Garcia, P. (2016). Commodity storage under backwardation: Does the Working curve still work? *Applied Economic Perspectives and Policy*, 38(1), 152–173.

Kaldor, N. (1939). Speculation and economic stability. *The Review of Economic Studies*, 7(1), 1–27.

Karali, B., Isengildina-Massa, O., Irwin, S. H., Adjemian, M. K. and Johansson, R. (2018). Are USDA reports still news to crop markets in the big data era?, mimeo, University of Illinois.

Karali, B. and Thurman, W. N. (2009). Announcement effects and the theory of storage: an empirical study of lumber futures. *Agricultural Economics*, 40(4), 421–436.

Lechthaler, F. and Vinogradova, A. (2017). The climate challenge for agriculture and the value of climate services: Application to coffee-farming in Peru. *European Economic Review*, 94, 45–70.

Leduc, S., Moran, K. and Vigfusson, R. J. (2016). *Learning in the Oil Futures Markets: Evidence and Macroeconomic Implications*. International Finance Discussion Paper 1179, FRB.

Lowry, M., Glauber, J. W., Miranda, M. J. and Helmberger, P. G. (1987). Pricing and storage of field crops: A quarterly model applied to soybeans. *American Journal of Agricultural Economics*, 69(4), 740–749.

Lybbert, T. J., Smith, A. and Sumner, D. A. (2014). Weather shocks and inter-hemispheric supply responses: Implications for climate change effects on global food markets. *Climate Change Economics*, 5(4), 1450010.

Merener, N. (2015). Globally distributed production and the pricing of CME commodity futures. *Journal of Futures Markets*, 35(1), 1–30.

Miranda, M. J. and Fackler, P. L. (2002). *Applied Computational Economics and Finance*. Cambridge, US: MIT Press.

Miranda, M. J. and Glauber, J. W. (1995). Solving stochastic models of competitive storage and trade by Chebychev collocation methods. *Agricultural and Resource Economics Review*, 24(1), 70–77.

Morris, S. and Shin, H. S. (2002). Social value of public information. *The American Economic Review*, 92(5), 1521–1534.

Ng, S. and Ruge-Murcia, F. J. (2000). Explaining the persistence of commodity prices. *Computational Economics*, 16(1–2), 149–171.

Osborne, T. (2004). Market news in commodity price theory: Application to the Ethiopian grain market. *The Review of Economic Studies*, 71(1), 133–164.

Peterson, H. H. and Tomek, W. G. (2005). How much of commodity price behavior can a rational expectations storage model explain? *Agricultural Economics*, 33(3), 289–303.

Roberts, M. J. and Schlenker, W. (2013). Identifying supply and demand elasticities of agricultural commodities: Implications for the US ethanol mandate. *The American Economic Review*, 103(6), 2265–2295.

Sacks, W. J., Deryng, D., Foley, J. A. and Ramankutty, N. (2010). Crop planting dates: an analysis of global patterns. *Global Ecology and Biogeography*, 19(5), 607–620.

Steinwender, C. (2018). Real effects of information frictions: When the States and the Kingdom became united. *The American Economic Review*, 108(3), 657–696.

USDA (2010). *Field Crops: Usual Planting and Harvest Dates*. Agricultural Handbook 628, United States Department of Agriculture, National Agricultural Statistics Service.

USDA (2017a). Production, Supply and Distribution (PSD) data. Released on July 12, 2017 and retrieved from https://apps.fas.usda.gov/psdonline/app/index.html.

USDA (2017b). World Agricultural Supply and Demand Estimates Report (WASDE). Retrieved from https://www.usda.gov/oce/commodity/wasde/.

Williams, J. C. and Wright, B. D. (1991). *Storage and Commodity Markets*. New York: Cambridge University Press.

Winne, J. D. and Peersman, G. (2016). Macroeconomic effects of disruptions in global food commodity markets: Evidence for the United States. *Brookings Papers on Economic Activity*, 47(2), 183–286.

Working, H. (1949). The theory of price of storage. *The American Economic Review*, 39(6), 1254–1262.

World Bank and FAO (2012). *The Grain Chain: Food Security and Managing Wheat Imports in Arab Countries*. Washington, DC.: World Bank.