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

Christophe Gouel

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

Christophe Gouel

Inter-annual storage model

Here, we present the inter-annual storage model corresponding to the seasonal model used in our study. This section aims 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:

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

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

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

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 $P^{ss} = \bar{P}$.

Seasonal storage model in deviation from steady state

Let's define variables in deviation from the deterministic annual steady state: $\check{P}_{i,t} = P_{i,t}/\bar{P}$, $\check{S}_{i,t} = S_{i,t}/\bar{D}$, $\check{Q}_{i^r,t} = Q_{i^r,t}/\left(\theta^r\bar{D}\right)$, $\check{A}_{i,t} = A_{i,t}/\bar{D}$, and $\check{\hat{Q}}_{i,t}^r = \hat{Q}_{i,t}^r/\left(\theta^r\bar{D}\right)$. $\check{S}_{i,t}$ and $\check{A}_{i,t}$ can be interpreted as the stock and availability to annual steady-state use ratios. Then, all the model's equations can be divided by a steady state value, which gives

$$\check{S}_{i,t} : \beta \operatorname{E}_{i,t} \check{P}_{i+1,t} - \check{P}_{i,t} - k \le 0, = 0 \text{ if } \check{S}_{i,t} > 0, \tag{A6}$$

$$\check{Q}^r_{i^r,t} : \check{Q}^r_{i^r,t} = \left[\operatorname{E}_{i^r,t} \left(\check{P}_{i^r+5,t} \epsilon^r_{i^r+5,t}\right)\right]^{\alpha^Q}, \tag{A7}$$

$$\check{P}_{i,t} : \check{A}_{i,t} = \frac{\left(\check{P}_{i,t}\right)^{\alpha^D}}{12} + \check{S}_{i,t}, \tag{A8}$$

$$\check{A}_{i,t} : \check{A}_{i,t} = \begin{cases} \check{S}_{i-1,t} + \epsilon^r_{i,t} \theta^r \check{Q}^r_{i-5,t} & \text{if } i = i^r + 5, \\ \check{S}_{i-1,t} & \text{if } i \ne i^r + 5, \end{cases} \tag{A9}$$

$$\check{\hat{Q}}^r_{i,t} : \check{\hat{Q}}^r_{i,t} = \check{Q}^r_{i^r,t} \exp\left(\sum_{j=i^r+1}^{i} \eta^r_{j,t}\right) \text{ for } i^r + 1 \le i \le i^r + 5. \tag{A10}$$

These equations are independent of the chosen values of $\bar{P}$ and $\bar{D}$. Thus, $\bar{P}$ and $\bar{D}$ only serve to scale the numerical results around given levels of quantities and prices.

Similarly, the variable part of instantaneous welfare can be expressed in deviation from the monthly steady-state consumption value, $\bar{P}\bar{D}/12$, as

$$\check{w}_{i,t} = \frac{12 w_{i,t}}{\bar{P}\bar{D}} = -\frac{\left(\check{P}_{i,t}\right)^{1+\alpha^D}}{1+\alpha^D} + 12 \left[\check{P}_{i,t} \left(\check{A}_{i,t} - \check{S}_{i,t}\right) - k\check{S}_{i,t} - \mathbf{1}_{i=i^r} \left(\theta^r \beta^5\right) \frac{\left(\check{Q}^r_{i^r,t}\right)^{1+1/\alpha^Q}}{1+1/\alpha^Q}\right], \tag{A11}$$

where $\bar{P}$ and $\bar{D}$ play no role.

Behavioral parameters

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).

Figure A1 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 article (see Joseph, Irwin, and Garcia, 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.

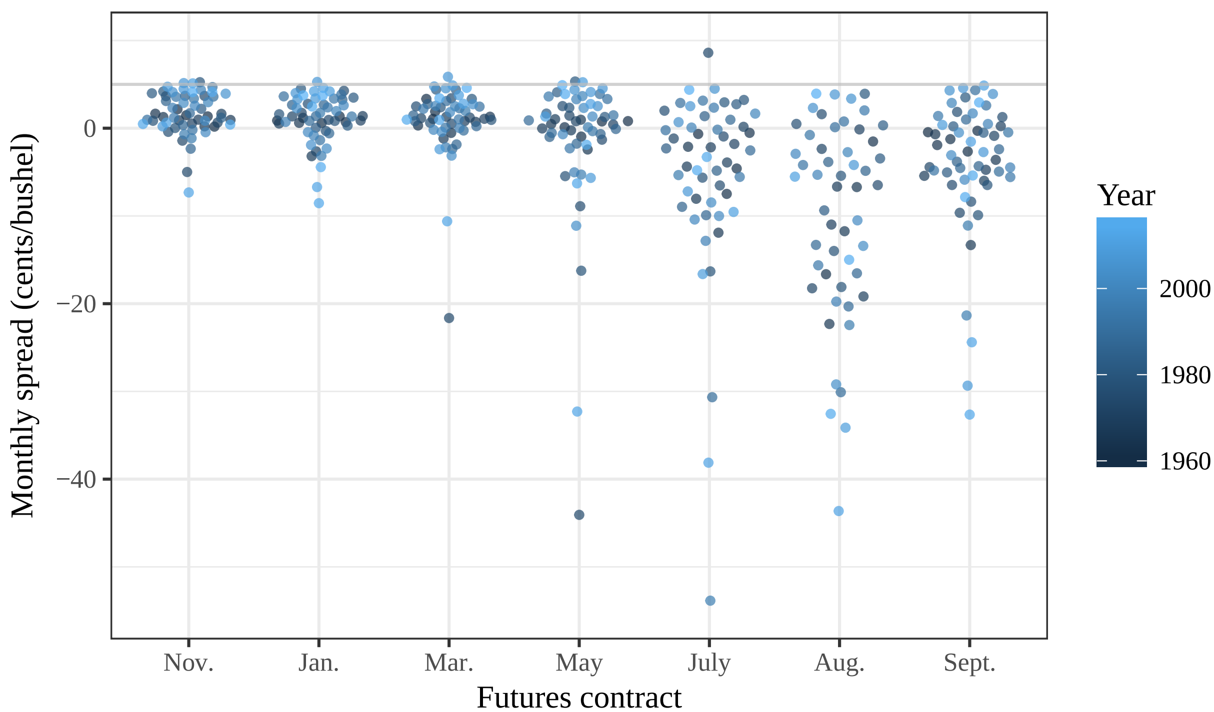

Figure A1. Interest-adjusted monthly price spread between a contract on its first delivery date and the following contract. Truncated y-axis: 9 spreads are below −55 cents per bushel.

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 A1. 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, Irwin, and Smith, 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 A2 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%, a value whose consequences are analyzed in the sensitivity analysis.

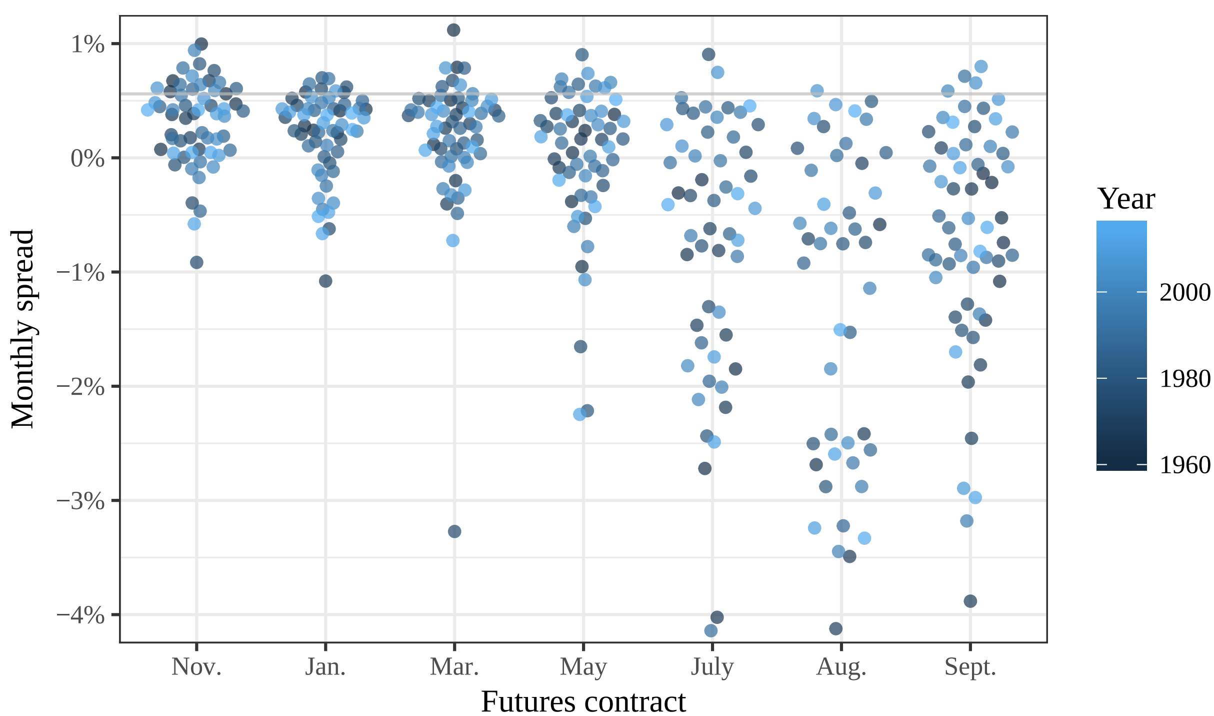

Figure A2. Interest-adjusted monthly price spread in percentage between a contract on its first delivery date and the following contract. Truncated y-axis: 17 spreads are below −4.2%.

Supply elasticity

To estimate supply elasticity, we take equation (14), multiply it by $\epsilon^r_{i^r+5,t}$, and take logs, which gives

$$\log\left(Q^r_{i^r,t}\epsilon^r_{i^r+5,t}\right) = \iota + \alpha^Q \log\left(\mathrm{E}_{i^r,t}\left(P_{i^r+5,t}\epsilon^r_{i^r+5,t}\right)\right) + \log\epsilon^r_{i^r+5,t} + \log\theta^r, \tag{A12}$$

where $\iota = \log\bar{D} - \alpha^Q \log\bar{P}$. There is generally no observables in the data corresponding to $\mathrm{E}_{i^r,t}\left(P_{i^r+5,t}\epsilon^r_{i^r+5,t}\right)$, so we substitute it by $\mathrm{E}_{i^r,t}\left(P_{i^r+5,t}\right)$ which can be mapped in the data to the futures prices at planting time for delivery at harvest.

Before moving to the rest of the estimation strategy, let first discuss the consequences of utilizing futures prices as proxy for the true producer incentives. The two variables have the following relationship

$$\log\left(\mathrm{E}_{i^r,t}\left(P_{i^r+5,t}\epsilon^r_{i^r+5,t}\right)\right) = \log\left(\mathrm{E}_{i^r,t}\left(P_{i^r+5,t}\right)\right) + \log\left(1 + \frac{\mathrm{cov}_{i^r,t}\left(P_{i^r+5,t},\epsilon^r_{i^r+5,t}\right)}{\mathrm{E}_{i^r,t}\left(P_{i^r+5,t}\right)}\right), \tag{A13}$$

where $\mathrm{cov}_{i^r,t}\left(P_{i^r+5,t},\epsilon^r_{i^r+5,t}\right)$ is the covariance between price at the harvest and production shock conditional on the information at planting. We can expect $\mathrm{cov}_{i^r,t}\left(P_{i^r+5,t},\epsilon^r_{i^r+5,t}\right) < 0$ because a positive yield shock will be associated to lower prices and conversely, but this covariance won't be constant. For high level of availability at planting, it can be expected that yield shocks will be mostly absorbed by stock variations and the covariance will be small. Conversely, for low levels of availability, stocks will be tight, the perceived demand elasticity in the market closer to the final demand elasticity, and the covariance more negative. This implies that we should expect $\log\left(\mathrm{E}_{i^r,t}\left(P_{i^r+5,t}\epsilon^r_{i^r+5,t}\right)\right) < \log\left(\mathrm{E}_{i^r,t}\left(P_{i^r+5,t}\right)\right)$, the two values being closer the higher the level of availability. The consequence is that using futures prices as a proxy for the producer incentives will lead to a downward bias in the estimation of supply elasticity. After the estimations, we show with numerical simulations that the bias is likely to be small, because the two variables are very close.

Estimating supply elasticity based on equation (A12) with OLS could lead to biased estimates because of endogeneity problems. So, 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.

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. From equation (A12), our dependent variable is the log quantity supplied, denoted $q_{rt} = \log\left(Q^r_{i^r+5,t}\epsilon^r_{i^r+5,t}\right)$. Our explanatory variables are the log of the harvest-time price expected at planting time, $p^q_{rt} = \log\left(\mathrm{E}_{i^r,t}P_{i^r+5,t}\right)$, corresponding empirically to the futures price, and a region-specific random shock, $\omega_{rt} = \log\epsilon^r_{i^r+5,t}$. 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, Janzen, and Smith, 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.[1] 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 nonlinear and 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. So, we 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, $p^d_{rt-1} = \log\left(P_{i^r+5,t-1}\right)$, 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.[2]

The resulting empirical model is

$$\text{(A14)} \qquad q_{rt} = \iota + \alpha^Q p^q_{rt} + \gamma\omega_{rt} + f^q_r(t) + \delta_r + u_{rt},$$

$$\text{(A15)} \qquad p^q_{rt} = \zeta + \kappa\omega_{rt} + \lambda\omega_{ht-1} + \nu p^d_{rt-1} + \xi\omega_{ht-1}p^d_{rt-1} + f^p_r(t) + \tau_r + v_{rt} \text{ for } r \notin h.$$

Each equation includes a region-specific time trend $f^j_r(t)$, modeled by restricted cubic spline ($j \in \{q, p\}$),[3] and region fixed effects, δ_r and τ_r. This empirical model is consistent with our theoretical model with the exception of the substitution of the producer incentives by the expected prices and 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.

Soybean production and yields are from FAOSTAT (FAO, 2018). 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

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

[2] Adding quadratic terms in the instruments to better account for the nonlinear relationship between futures and lagged yield shocks does not change the estimates but leads to weaker instruments.

[3] 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.

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 A1 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 -1.66, evaluated at the 10% percentile of sample price it is -0.46, and evaluated at the 90% percentile of sample price it is -2.66. 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, Kalkuhl, and von Braun'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).[4] 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, Janzen, and Smith'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.

Using all the chosen and estimated parameters, we can assess the size of the bias resulting from using futures prices instead of the true producer incentives by simulating the model. The difference between the two variables is illustrated on figure A3 with the calibrated model. We can see that these two expectation terms are very close, especially for high availability, and as predicted the expected price is superior to the producer incentives. To analyze the bias quantitatively, let's now suppose that we can express the log of the producer incentive as a linear function of the log of the expected price:

$$\log\left(\mathrm{E}_{i^r,t}\left(P_{i^r+5,t}\epsilon^r_{i^r+5,t}\right)\right) = \upsilon + \rho \log\left(\mathrm{E}_{i^r,t}\left(P_{i^r+5,t}\right)\right) + \vartheta_{i^r,t}, \tag{A16}$$

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

Table A1. Estimates of Supply Elasticity

	2SLS			OLS		
	(1a)	(1b)	(1c)	(2a)	(2b)	(2c)
Supply elasticity α^Q	0.60***	0.25**	0.38***	0.51***	0.08	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)	(0.22)	(0.15)	(0.14)
First stage ω_{rt}	0.17	0.11	0.09			
	(0.16)	(0.14)	(0.15)			
First stage ω_{ht-1}	10.74*	9.99*	10.38*			
	(6.06)	(5.50)	(5.55)			
First stage p^d_{rt-1}	0.79***	0.59***	0.57***			
	(0.04)	(0.06)	(0.06)			
First stage $\omega_{ht-1}p^d_{rt-1}$	−2.04**	−1.89**	−1.96**			
	(0.98)	(0.89)	(0.90)			
First-stage F-statistics	116.00	38.82	35.70			
p-value for Hausman test	0.02	0.01	0.00			
p-value for overid. test	0.54	0.76	0.76			
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.

with $\rho < 1$. Replacing equation (A16) in equation (A12) gives

$$\text{(A17)} \qquad \log\left(Q^r_{i^r,t}\epsilon^r_{i^r+5,t}\right) = \iota' + \alpha^Q\rho\log\left(\mathrm{E}_{i^r,t}\left(P_{i^r+5,t}\right)\right) + \log\epsilon^r_{i^r+5,t} + \log\theta^r + \alpha^Q\vartheta_{i^r,t}.$$

Under the model's parameters, and after solving the model, we obtain $\rho = 0.955$. Thus, there is a 5% downward bias.

In our context, it appears that the size of the bias is smaller than the standard deviation of the estimates. This may not be the case in all settings. Intuitively, the bias should be a positive function of demand elasticity, storage cost, and the size of the production shocks. It should be evaluated depending on the situation.

Numerical methods[5]

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

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

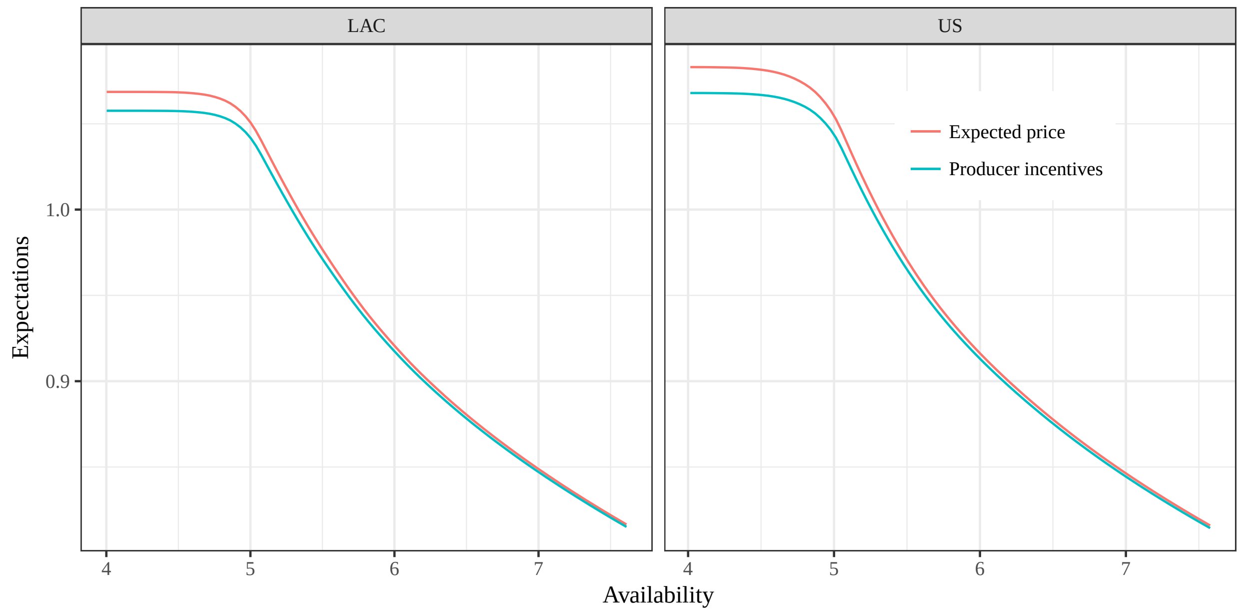

Figure A3. Expected prices and producer incentives at planting time as function of availability

R2018b 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 article is available at DOI: 10.15454/YWAOKM). The method presented is for a typical rational expectations model with informational subperiods, and not for the specific model equations. Following Fackler (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:

$$s_i = g_{i-1}(s_{i-1}, x_{i-1}, e_i), \tag{A18}$$

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

$$f_i(s_i, x_i, z_i) \leq 0, = 0 \text{ if } x_i > \underline{x}_i. \tag{A19}$$

Response variables can have lower bounds, $\underline{x}_i$. In cases where response variables have no bounds, equation (A19) 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

$$z_i = \mathrm{E}_i \left[h_i \left(s_i, x_i, e_{i+1}, s_{i+1}, x_{i+1} \right) \right]. \tag{A20}$$

There is one difference between the generic model defined by equations (A18)–(A20) 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, $\mathrm{E}_{i^r,t}(P_{i^r+5,t}\epsilon^r_{i^r+5,t})$ in equation (3). We can use the law of iterated expectations to make the connection between the two expressions. For example:

$$\mathrm{E}_{i^r,t}(P_{i^r+5,t}\epsilon^r_{i^r+5,t}) = \mathrm{E}_{i^r,t}(\mathrm{E}_{i^r+1,t}(\mathrm{E}_{i^r+2,t}(\mathrm{E}_{i^r+3,t}(\mathrm{E}_{i^r+4,t}(P_{i^r+5,t}\epsilon^r_{i^r+5,t}))))). \tag{A21}$$

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,

$$x_i \approx \Phi_i\left(s_i, \theta_i\right), \tag{A22}$$

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 (A20) 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 (A18) and (A20)–(A22), we can express the equilibrium equation (A19) as

$$f_i\left(s_i, x_i, \sum_l w^l_i h_i\left(s_i, x_i, e^l_{i+1}, g_i\left(s_i, x_i, e^l_{i+1}\right), \Phi_{i+1}\left(g_i\left(s_i, x_i, e^l_{i+1}\right), \theta_{i+1}\right)\right)\right) \le 0, = 0 \text{ if } x_i > \underline{x}_i. \tag{A23}$$

For a given spline approximation, θ_{i+1}, and a given s_i, equation (A23) 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^n_1 with $n = 1$, based on a first-guess (for example, the steady-state values).

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

Step 2.1. Equation solving step. For each point of the grid of state variables, s^j_i, solve equation (A23) for x^j_i using a mixed complementarity solver:

$$f_i\left(s^j_i, x^j_i, \sum_l w^l_i h_i\left(s^j_i, x^j_i, e^l_{i+1}, g_i\left(s^j_i, x^j_i, e^l_{i+1}\right), \Phi_{i+1}\left(g_i\left(s^j_i, x^j_i, e^l_{i+1}\right), \theta^n_{i+1}\right)\right)\right) \le 0, = 0 \text{ if } x^j_i > \underline{x}_i. \tag{A24}$$

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

Step 3. Terminal step. If $n = 1$ or $\|\theta^{n+1} - \theta^n\|_2 \ge 10^{-8}$ 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 (A23) 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 (11), 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.

Supplementary figures and tables

Table A2. Effect of News Shocks on Price Standard Deviation (All Results in Percentages)

			Changes compared to no news		
Period	Shock	No news	All News	US News	LAC News
Oct./Nov.	σ_1^{US}	17.25	−15.1	−17.7	4.3
Nov./Dec.	σ_2^{LAC}	17.28	−11.7	−17.7	6.8
Dec./Jan.	σ_3^{LAC}	17.30	−6.7	−17.7	10.4
Jan./Feb.	σ_4^{LAC}	17.33	−3.4	−17.7	12.8
Feb./Mar.	σ_5^{LAC}	17.42	1.4	−18.0	16.3
Mar./Apr.	σ_6^{LAC}	15.43	−7.0	5.9	−11.3
Apr./May		15.46	−7.0	5.9	−11.3
May/June		15.48	−7.0	5.9	−11.3
June/July	σ_9^{US}	15.51	−3.8	8.4	−11.3
July/Aug.	σ_{10}^{US}	15.53	2.2	13.1	−11.3
Aug./Sept.	σ_{11}^{US}	15.56	8.1	17.9	−11.3
Sept./Oct.	σ_{12}^{US}	15.70	14.7	23.1	−12.0
All[a]		16.35	−2.9	−1.0	−0.6

Source: Statistics calculated over 1,000,000 sample observations from the asymptotic distribution simulated with the model.
Note: [a] Statistics calculated over the pooled prices from all periods.

August 2016

WASDE - 556 - 28

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

2014/15		Beginning Stocks	Production	Imports	Domestic Crush	Domestic Total	Exports	Ending Stocks
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 A4. Example of a WASDE report on soybean: Page 28 of the August 12th 2016 pdf report

References

Fackler, P.L. 2005. "A MATLAB Solver for Nonlinear Rational Expectations Models." *Computational Economics* 26:173–181.

FAO. 2018. "FAOSTAT Database." Retrieved May 28, 2018 from http://www.fao.org/faostat/en.

Garcia, P., S.H. Irwin, and A. Smith. 2015. "Futures Market Failure?" *American Journal of Agricultural Economics* 97:40–64.

Gouel, C. 2018. "RECS: MATLAB Solver for Rational Expectations Models with Complementarity Equations – Version 0.7." DOI: 10.5281/zenodo.1136095.

Gouel, C., and S. Jean. 2015. "Optimal Food Price Stabilization in a Small Open Developing Country." *World Bank Economic Review* 29:72–101.

Haile, M.G., M. Kalkuhl, and J. von Braun. 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* 98:172–190.

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

Hendricks, N.P., J.P. Janzen, and A. Smith. 2015. "Futures Prices in Supply Analysis: Are Instrumental Variables Necessary?" *American Journal of Agricultural Economics* 97:22–39.

Joseph, K., S.H. Irwin, and P. Garcia. 2016. "Commodity Storage under Backwardation: Does the Working Curve Still Work?" *Applied Economic Perspectives and Policy* 38:152–173.

Kaldor, N. 1939. "Speculation and Economic Stability." *The Review of Economic Studies* 7:1–27.

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

Miranda, M.J., and J.W. Glauber. 1995. "Solving Stochastic Models of Competitive Storage and Trade by Chebychev Collocation Methods." *Agricultural and Resource Economics Review* 24:70–77.

Roberts, M.J., and W. Schlenker. 2013. "Identifying Supply and Demand Elasticities of Agricultural Commodities: Implications for the US Ethanol Mandate." *The American Economic Review* 103:2265–2295.

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

Winne, J.D., and G. Peersman. 2016. "Macroeconomic Effects of Disruptions in Global Food Commodity Markets: Evidence for the United States." *Brookings Papers on Economic Activity* 47:183–286.

Working, H. 1949. "The Theory of Price of Storage." *The American Economic Review* 39:1254–1262.

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