The Wheat Market

Andrew Kao, Andrei Kim, and Yu-Jia Wang

August 2023

Research Question

In our research project, we would like to explore the effect of removing Ukraine from the market of wheat, in particular, the changes in prices and quantities produced by different countries.

Model — timing

t.0 — Global market price P for wheat is decided (forward contract)

— and firms decide how much quantity $\tilde{Q}$ to produce

t.1 — Supply shocks result in Q actually being produced

t.2 — Global marketplace opens: all firms sell to meet demand

t.3 — Warehouse buys/sells wheat to clear market; market closes

t+1 — Process repeats with new stock of wheat

Model — supply

Supply: We assume that there is a representative firm in each country c that collectively sell in a single (perfectly competitive) market. They each face an i.i.d. supply shock $\nu_{ct} \sim F_\nu$ with unit mean.

At each time t, firms maximize their expected profit:

$$\tilde{Q}_{ct} = \arg\max_{\tilde{Q}} \mathbb{E}[P_t Q - C_c(\tilde{Q})], \qquad Q = \tilde{Q}\nu_{ct}$$

where Q is the amount produced after the supply shock. We assume ν_{ct} is unobserved to both the firms and the econometrician. Costs are assumed convex and modeled as:

$$C_c(\tilde{Q}) = \gamma_c \tilde{Q}^\alpha$$

with $\alpha > 1$. For all $c, c' \in C$ and all t, we assume that $\nu_{ct} \perp \gamma_{c'}$. The FOC holds in equilibrium:

$$P_t \mathbb{E}[\nu_{ct}] - \alpha\gamma_c \tilde{Q}_{ct}^{\alpha-1} = 0 \implies P_t = \alpha\gamma_c \tilde{Q}_{ct}^{\alpha-1}$$

Model — demand and inventory

Demand: We assume that global demand is perfectly inelastic and a function of global population N_t plus a (known to everyone) demand shock ζ_t assuming $\mathbb{E} \log \zeta_t = 0$ and $\zeta_t \perp N_t$:

$$Q_t^D = \gamma N_t^{\delta} \zeta_t$$

We further assume that $\nu_{ct} \perp (\zeta_{t'}, N_{t'})$ for all $c \in C$ and all $t, t' \in \{1, \ldots, T\}$.

Warehouse: We assume that there is a warehouse with a stock of wheat. It is the only firm with storage technology. If there is excess supply/demand, it buys/sells the difference ($\pm Q_{wt}$) at price P_t to clear the market. The warehouse makes 0 profit in expectation.

Model — equilibrium

Let $C = \{c_1, \ldots, c_N\}$ be the set of representative firms. An equilibrium is a price P and a set of quantities Q_t^D (demand), $\tilde{Q}_{ct}, Q_{ct} \forall c \in C$ (chosen and realized quantities by firms) such that:

1. Firms and consumers optimize. Firms must optimally choose a quantity target $\tilde{Q}$ based on the FOC:

$$\forall c \in C, P_t = \alpha \gamma_c \tilde{Q}_{ct}^{\alpha - 1}$$

2. Markets clear in expectation:

$$Q_t^D = \mathbb{E}\left[\sum_{c \in C} Q_{ct}\right] = \sum_{c \in C} \tilde{Q}_{ct}$$

   and in reality:

$$Q_t^D + Q_{wt} = \sum_{c \in C} Q_{ct}$$

An equilibrium must hold in all time periods t.

Model — equilibrium (ctd.)

Fix any t. For each firm c, optimization implies

$$\tilde{Q}_{ct} = \left(\frac{P_t}{\alpha\gamma_c}\right)^{\frac{1}{\alpha-1}}$$

Market clearing then pins down the equilibrium price, i.e.

$$\gamma N_t^{\delta}\zeta_t = \sum_{c\in C}\left(\frac{P_t}{\alpha\gamma_c}\right)^{\frac{1}{\alpha-1}} \Rightarrow P_t = \alpha\left(\frac{\gamma N_t^{\delta}\zeta_t}{\sum_{c\in C}\gamma_c^{\frac{1}{1-\alpha}}}\right)^{\alpha-1}$$

It follows that the equilibrium quantities are given by

$$\tilde{Q}_{ct} = \frac{\gamma_c^{\frac{1}{1-\alpha}}}{\sum_{c'\in C}\gamma_{c'}^{\frac{1}{1-\alpha}}}\gamma N_t^{\delta}\zeta_t$$

for all $c \in C$.

Data

In our analysis, we consider the following data as observed:

- Global price P_t and demand for wheat Q_t^D, as well as quantities produced by each country Q_{ct} from 1992 until 2020. Source: FAO.
- World population N_t from 1992 until 2020. Source: World Bank.
- The distribution of supply shocks ν_{ct} (in particular, the expectation and variance of the $\log \nu_{ct}$) and information breaking down the operational costs for wheat farms (for arguing about the cost function and its independence of the shock). Source: USDA ARMS.

Identification

Given $\{(P_t, \{Q_{ct}\}_{c \in C})\}_{t=1}^{T}$, firms' FOC and $Q_{ct} = \tilde{Q}_{ct}\nu_{ct}$ imply

$$\log P_t = \log(\alpha\gamma_c) + (\alpha - 1)(\log Q_{ct} - \log \nu_{ct}).$$

Taking within-transformation, we have

$$p_t = (\alpha - 1)q_{ct} + (1 - \alpha)\eta_{ct}$$

where

$$p_t = \log P_t - \frac{1}{T}\sum_{t=1}^{T}\log P_t,$$

$$q_{ct} = \log Q_{ct} - \frac{1}{T}\sum_{t=1}^{T}\log Q_{ct},$$

$$\eta_{ct} = \log \nu_{ct} - \frac{1}{T}\sum_{t=1}^{T}\log \nu_{ct}.$$

Identification

Regressing p_t on q_{ct}, we obtain the coefficient

$$\begin{aligned}
\beta =& \frac{\mathrm{Cov}(p_t, q_{ct})}{\mathrm{Var}(q_{ct})} \\
=& \alpha - 1 + (1-\alpha)\frac{\mathrm{Cov}(\eta_{ct}, q_{ct})}{\mathrm{Var}(q_{ct})} \\
=& \alpha - 1 + (1-\alpha)\frac{\mathrm{Var}(\eta_{ct}) + \mathrm{Cov}(\eta_{ct}, \tilde{q}_{ct})}{\mathrm{Var}(\tilde{q}_{ct}) + \mathrm{Var}(\eta_{ct}) + 2\mathrm{Cov}(\eta_{ct}, \tilde{q}_{ct})} \\
=& (\alpha - 1)\frac{\mathrm{Var}(\tilde{q}_{ct}) + \mathrm{Cov}(\eta_{ct}, \tilde{q}_{ct})}{\mathrm{Var}(\tilde{q}_{ct}) + \mathrm{Var}(\eta_{ct}) + 2\mathrm{Cov}(\eta_{ct}, \tilde{q}_{ct})}
\end{aligned}$$

where

$$\tilde{q}_{ct} = \log \tilde{Q}_{ct} - \frac{1}{T}\sum_{t=1}^{T} \log \tilde{Q}_{ct}.$$

Identification

From the equilibrium result and the orthogonality assumptions $\nu_{ct} \perp (\gamma_{c'}, N_{t'}, \zeta_{t'})$ for all $c, c' \in C$ and $t, t' \in \{1, \ldots, T\}$, we have $\nu_{ct} \perp \tilde{Q}_{c't'}$ for all $c, c' \in C$ and $t, t' \in \{1, \ldots, T\}$, therefore $\eta_{ct} \perp \tilde{q}_{ct}$ for all $c \in C$ and the regression coefficient becomes

$$\beta = (\alpha - 1)\frac{\mathrm{Var}(\tilde{q}_{ct})}{\mathrm{Var}(\tilde{q}_{ct}) + \mathrm{Var}(\eta_{ct})}$$

which is subject to the usual attenuation bias as in a regression model with classical measurement error. Rearranging, we obtain

$$\alpha = 1 + \beta\frac{\mathrm{Var}(q_{ct})}{\mathrm{Var}(q_{ct}) - \mathrm{Var}(\eta_{ct})}.$$

Note that $\mathrm{Var}(q_{ct})$ and β is identified from the joint distribution of $(P_t, \{Q_{ct}\}_{c\in C})$. Therefore α is identified provided that $\mathrm{Var}(\eta_{ct})$ is identified.

Identification

Since ν_{ct} is iid, we have

$$\mathrm{Var}(\eta_{ct}) = \mathrm{Var}\left(\log \nu_{ct} - \frac{1}{T}\sum_{t=1}^{T} \log \nu_{ct}\right) = \left(1 - \frac{1}{T}\right)\mathrm{Var}(\log \nu_{ct}).$$

From the USDA data $\{(\tilde{Q}_i, Q_i)\}_{i\in I}$, we identify

$$\mathrm{Var}(\log \nu_{ct}) = \mathrm{Var}(\log Q_i - \log \tilde{Q}_i)$$

assuming $Q_i = \tilde{Q}_i \nu_i$ where $\nu_i \overset{iid}{\sim} F_\nu$. We have thus identified α as

$$\alpha = 1 + \beta \frac{\mathrm{Var}(q_{ct})}{\mathrm{Var}(q_{ct}) - (1 - \frac{1}{T})\mathrm{Var}(\log Q_i - \log \tilde{Q}_i)}$$

via WFO data $\{(P_t, \{Q_{ct}\}_{c\in C})\}_{t=1}^{T}$ and USDA data $\{(\tilde{Q}_i, Q_i)\}_{i\in I}$.

Identification

With α identified, we recover γ_c for each country via

$$\gamma_c = \exp\left\{\mathbb{E}[\log P_t] - \log\alpha + (\alpha - 1)(\mathbb{E}[\log \nu_{ct}|c] - \mathbb{E}[\log Q_{ct}|c])\right\}$$

where $\mathbb{E}[\log \nu_{ct}|c]$ is identified from the USDA data $\{(\tilde{Q}_i, Q_i)\}_{i\in I}$ as

$$\mathbb{E}[\log \nu_{ct}|c] = \mathbb{E}\left[\log\left(\frac{Q_i}{\tilde{Q}_i}\right)\right] = \mathbb{E}[\log Q_i] - \mathbb{E}[\log \tilde{Q}_i],$$

and the rest of the quantities are identified from moments of the WFO data $\{(P_t, \{Q_{ct}\}_{c\in C})\}_{t=1}^T$.

Estimation

We apply the plug-in principles for estimation. First, we estimate supply elasticity α by

$$\hat{\alpha} = 1 + \hat{\beta}\frac{\widehat{\mathrm{Var}}(q_{ct})}{\widehat{\mathrm{Var}}(q_{ct}) - (1 - \frac{1}{T})\widehat{\mathrm{Var}}(\log Q_i - \log \tilde{Q}_i)}$$

where $\hat{\beta}$ is the sample Mundlak estimator of $\log P_t$ on $\log Q_{ct}$, and $\widehat{\mathrm{Var}}(\cdot)$ is the standard sample variance estimator of the relevant quantities. We then estimate the cost parameters γ_c for each representative firm $c \in C$ by

$$\hat{\gamma}_c = \exp\left\{ \frac{1}{T}\sum_{t=1}^{T}\left[\log P_t - \log\hat{\alpha} + (\hat{\alpha} - 1)(\widehat{\mathbb{E}}[\log \nu_{ct}] - \log Q_{ct})\right]\right\}$$

where

$$\widehat{\mathbb{E}}[\log \nu_{ct}] = \frac{1}{|I|}\sum_{i}(\log Q_i - \log \tilde{Q}_i).$$

By standard asymptotic arguments, our estimates $\hat{\alpha}$ and $\{\hat{\gamma}_c\}_{c\in C}$ are consistent as $T \to \infty$ and $|I| \to \infty$.

Calculations — counterfactual price of wheat

We compute the counterfactual price following this procedure:

1. Estimate $\widehat{\mathbb{E}}[\log \nu_{ct}], \widehat{Var}(\log \nu_{ct})$ from USDA data.
2. Estimate $\hat{\gamma}, \hat{\delta}$ (demand) by OLS using population data.
3. Run fixed effects regression to obtain $\hat{\beta}$.
4. Use procedure in prior slides to estimate supply elasticity and fixed effects $\hat{\alpha}$ and $\hat{\gamma}_c$.
5. Estimate wheat demand in 2022 (using world population) and drop Ukraine from list of countries.
6. Solve the linear programming problem from firm FOC and market clearing, yielding (expected) equilibrium price.
7. Bootstrap steps 1-6 to form a confidence set.[1]

[1] If $\hat{\alpha} \leq 1$ then we stop at step 4.

Why the model is not refuted by data

- Supply shock: the assumption that $\mathbb{E}[\nu_{ct}] = 1$ is not (statistically) rejected. 95% CI: $\mathbb{E}[\nu_{ct}] \in (0.962, 1.081)$ with point estimate 1.022. This accounts for variation in supply and explains why supply is not equal to demand.
- Cost function: supply costs are mostly contracted before quantity is realized (ex/ seed, fertilizer, machinery cost, are 75% of operating costs) so $\tilde{Q}$, not Q, enters the cost function. $\alpha > 1$ is also confirmed by our point estimate.
- Demand: R^2 of regressing world population on demand is 0.95 (ζ_t accounts for the remaining variation). We do not include price because, if included, price has a coefficient that's not significantly different from 0.

Findings

Parameters:

1. $\hat{\alpha} = 1.0022$, 95% CI: [0.989, 1.015] (rejects model 39.4% of the time)
2. $\hat{Q}_{D,2022} = 760.83$, 95% CI: [749.1, 769.6]
3. $\hat{\gamma}_{Ukraine} = 208.1$, 95% CI: [184.6, 223.6]
   $\hat{\gamma}_{China} = 210.5$, 95% CI: [153.8,222.9]
   $\hat{\gamma}_{US} = 210.8$, 95% CI: [158.5,222.7]

Counterfactuals:

1. Price of wheat: $211.65 per ton, 95% CI: [158.43, 211.19]
2. Δ price of wheat: $0.0005 per ton, 95% CI: [0, 0.78]
3. New Caledonia produces 0.725 million tons of extra wheat (out of a demand shortfall of 0.763 million tons)

Why are they informative?

The confidence interval for the change in price of wheat is very informative (insofar as it is tight). It is also informs us that the model is probably wrong.

Limitations and future work

Here are things we considered doing but have left for future work:

1. Adjustment costs/land (and land quality)
2. Strong assumptions on ν?[2] Entry costs potential solution.
3. Demand function that responds to price
4. Strategic inventory
5. Oligopology with competitive fringe

[2] In the data countries that sometimes produce 0 and sometimes not implies $\nu = 0$ with positive probability; in practice we drop from regression.