Wheat[1]

Andrew Kao, Andrei Kim, and Yu-Jia Wang

August 2023

[1] À la Dube and Harish (2020).

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 $\zeta_t \perp N_t$:

$$Q_t^D = \phi N_t^\delta \zeta_t.$$

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. Eqbm P_t and $\tilde{Q}_{ct}$

Data

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

Variable	Notation	Structure	Source
global price	P_t	time series	FAO
global demand	Q_t^D	time series	FAO
country-level production	Q_{ct}	panel	FAO
farm-level yield	Q_i	cross-section	USDA ARMS
farm-level expected yield	$\tilde{Q}_i$	cross-section	USDA ARMS
global population	N_t	time series	WDI

Identification — fixed-effect model with CME ▸ Full derivation

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

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

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

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

where $\mathrm{Cov}(\eta_{ct}, \tilde{q}_{ct})$ follows from the orthogonality conditions.

Identification — bias correction

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:

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

Identification — recover cost parameters

With α identified, we recover γ_c for each country:

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

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

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

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

Estimation — plug-in estimators

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. 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 $\hat{\phi}, \hat{\delta}$ (demand) by OLS using population data.
3. Run fixed-effect regression to obtain $\hat{\beta}$.
4. Use procedure in prior slide 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 the 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.[2]

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

Is the model 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.

Is the model 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.
- Orthogonality: we cannot directly test whether $\nu_{ct} \perp \zeta_t$, but we can correlate the estimated residuals $\hat{\nu}_{ct}, \hat{\zeta}_t$. Although the $R^2 = 0.0002$ (in level and logs), $p = 0.001$ so the model is (statistically) rejected.

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 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_{ct}? Orthogonality, iid.
3. Entry costs: Countries sometimes produce 0 in the data.
4. Elastic demand.
5. Strategic inventory or storage costs (direct/iceberg).
6. Oligopology with competitive fringe.

Appendix I. Equilibrium in Closed Form

Back

Fix any $t \in \{1, \ldots, T\}$. For each firm $c \in C$, the FOC implies

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

Market clearing then pins down the equilibrium price:

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

It follows that the equilibrium quantities are given by

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

for all $c \in C$.

Appendix II. Derivation of Identification Results

Back

From firms' FOC:

$$
\begin{aligned}
P_t = \alpha\gamma_c \tilde{Q}_{ct}^{\alpha-1} &\implies \log P_t = \log(\alpha\gamma_c) + (\alpha-1)\log \tilde{Q}_{ct} \\
&\implies \log P_t = \log(\alpha\gamma_c) + (\alpha-1)\log\left(\frac{Q_{ct}}{\nu_{ct}}\right) \\
&\implies \log P_t = \log(\alpha\gamma_c) + (\alpha-1)(\log Q_{ct} - \log \nu_{ct}) \\
&\implies p_t = (\alpha-1)q_{ct} + (1-\alpha)\eta_{ct}
\end{aligned}
$$

where we take the within-transformations:

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

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

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

Appendix II. Derivation of Identification Results (ct'd)

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

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

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

Appendix II. Derivation of Identification Results (ct'd)

From the equilibrium result

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

and the orthogonality assumption

$$\nu_{ct} \perp (\gamma_{c'}, N_{t'}, \zeta_{t'}) \quad \forall c, c' \in C, \ \forall t, t' \in \{1, \ldots, T\},$$

we have

$$\nu_{ct} \perp \tilde{Q}_{c't'} \quad \forall c, c' \in C, \ \forall t, t' \in \{1, \ldots, T\},$$

and therefore

$$\eta_{ct} \perp \tilde{q}_{ct} \quad \forall c \in C, t \in \{1, \ldots, T\}.$$

The regression coefficient of p_t on q_{ct} thus becomes

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