Short note on the estimation of the parameters of the game

Charles-Albert Lehalle

Printed 17 August 2025

The goal of this note is to explain how to estimate the parameters of this game:

$$\max_{Q^n_{i,\cdot}} \left\{ \varphi^n_{i,\cdot} \sum_j Q^n_{i,j} M^n_{i,j} - \sum_j r_j \left[\left(\sum_{m\neq n} \varphi^m_{j,\cdot} \right) + \varphi^n_{j,\cdot} + \varphi^n_{i,\cdot} Q^n_{i,j} \right]^2 - c_n \sum_{j\neq i} \left(\varphi^n_{i,\cdot} Q^n_{i,j} \right)^2 \cdot g(T_{i,j}) \right\}. \tag{1}$$

1 Reminder: Quickly solving the control problem

(1) is not very difficult to solve. The control is a vector $(q_j)_{1\leq j\leq J}$ in the simplex corresponding the fraction of what is available at port i for good j that will be sent to port j (the configuration $j = i$ is possible; it is the amount of good that stays at i).

Using λ as Lagrange multiplier to keep q in the simplex, the optimal q has to cancel the augmented gradient (with the notation $\varphi^\bullet_{j,\cdot} = \sum_m \varphi^m_{j,\cdot}$):

$$\forall n, \forall i: \ \lambda^n_i = M^n_{i,j} - 2r_j(\varphi^\bullet_{j,\cdot} + \varphi^n_{i,\cdot} q_j) - 2c_n g(T_{i,j})\varphi^n_{i,\cdot} q_j \mathbf{1}_{j\neq i}.$$

It is then enough to express q_j as a function of the Lagrange multiplier and saturate the constraint:

$$Q^n_{i,j} = \frac{w^n_{i,j}}{2\varphi^n_{i,\cdot}} \left(M^n_{i,j} - 2r_j \varphi^\bullet_{j,\cdot} - \lambda^n_i \right) \Rightarrow \lambda^n_i = \frac{1}{\sum_\ell w^n_{i,\ell}} \sum_\ell w^n_{i,\ell} \left(M^n_{i,\ell} - 2r_\ell \varphi^\bullet_{\ell,\cdot} \right). \tag{2}$$

The Lagrange multiplier is a barycenter using $w^n_{i,j} = \left(r_j + c_n g(T_{i,j})\mathbf{1}_{j\neq i}\right)^{-1}$ as weights. The interpretation of the optimal destinations starting from the port i is the following:

- the higher the potential margins, i.e. $M^n_{i,j}$, the more quantity to send,
- the more crowded, i.e. $2r_j\varphi^\bullet_{j,\cdot}$, the less quantity to send,
- all that is adjusted by λ^n_i, that is an average of what would be sent to all ports without constraint,
- the averaging scheme weights less: the places where crowded costs more (without respect of what is sent, because of the term r_j), the places that suffers from high transportation costs (because of the term $c_n g(T_{i,j})$, except for the considered port i, since the transportation cost is zero).

CAL: For now I do not take care to keep the $Q^n_{i,j}$ non negative, but I think it is not a real problem. In practice we can take all that is negative and assign it to $Q^n_{i,i}$ (since it is cheaper to import goods from these ports than to export to them, the agent i can count on the other ports to send them for her). Of course we will have to address this for the convergence but I am confident.

Closing the loop to get the mean field. Solving this quadratic problem in the simplex gives the ith row of the transition matrix for good n. It is enough to stack them to get $[Q^n_{i,j}]_{i,j}$ that is the transition matrix (i.e. the control) for good n. Its first eigenvector is $\varphi^n_{i,\cdot}$. This can be done for all goods and we can go back to the optimal control problem until we get to a fixed point.

2 How to estimate the parameters of the game

2.1 Selection of time series

A list of meaningful goods, ports and countries has to be established. The criteria are

- Goods should be in the correct unit (check with ComTrade: the sum of what we see in Panjiva should be lower than the official total imports / exports in the considered country).
- Goods should be exchanged over the countries we have. For the countries we have (say we have countries A and B), the consistency of imports $A \to B$ with exports $B \to A$ has to be checked.
- Ports should be filtered and / or aggregated (according to the distance). We need ports that have enough shipping per week of our goods of interest.

In general, the selection procedure is: (1) reject the good if there is a too high inconsistency, (2) log it with descriptive statistics somewhere for the report.

A small report of what has been done to end up with our shortlist of goods has to be written.

2.2 Variables of interest

Thanks to this

- for each good n that we have in our final list, we should get a time series of matrices $Q^n_{i,j}(t)$.
- We should have an occupation index for each port, that will be our proxy for $\varphi^\bullet_{i,\cdot}$. First we will need the time series of $\varphi^\bullet_{i,\cdot}(t)$.
- Moreover, now that we know $\varphi^n_{i,\cdot}$ should be close to the first eigenvector of $Q^n_{i,j}$, some exploratory diagonalizations of local averages of $Q^n_{i,j}(t)$ (to stabilize it) should be tried and the time series of the first eigenvectors should be explored.
- We will need the distance $T_{i,j}$.

2.3 Remark on the units

Remark that equation (2) is almost unit-less:

$$
\left\{
\begin{array}{rcl}
Q^n_{i,j} & = & \dfrac{w^n_{i,j}}{2\varphi^n_{i,\cdot}}\left(M^n_{i,j} - 2r_j\varphi^\bullet_{j,\cdot} - \lambda^n_i\right) \\
\lambda^n_i & = & \dfrac{1}{\sum_\ell w^n_{i,\ell}} \displaystyle\sum_\ell w^n_{i,\ell}\left(M^n_{i,\ell} - 2r_\ell\varphi^\bullet_{\ell,\cdot}\right) \\
w^n_{i,j} & = & \left(r_j + c_n g(T_{i,j})\mathbf{1}_{j\neq i}\right)^{-1}
\end{array}
\right. \tag{3}
$$

gives the same result if one rescale M, r and c by the same factor u.

2.4 Natural relations between the parameters

The main equation, corresponding to unfold (3), is the following:

$$
Q^n_{i,j}\varphi^n_{i,\cdot} = w^n_{i,j}\,\frac{1}{2}\left(M^n_{i,j} - \sum_\ell \frac{w^n_{i,\ell}}{\sum_{\ell'} w^n_{i,\ell'}} M^n_{i,\ell}\right) - w^n_{i,j}\,\left(r_j\varphi^\bullet_{j,\cdot} - \sum_\ell \frac{w^n_{i,\ell}}{\sum_{\ell'} w^n_{i,\ell'}} r_\ell\varphi^\bullet_{\ell,\cdot}\right). \tag{4}
$$

Keep in mind that[1]

[1] we could take $\varphi^\bullet_{j,\cdot}$ as the anticipation of what is planned to be sent and not an anticipated congestion term.

- the weights $w^n_{i,j}$ are *independent of the time, and carry the attractiveness of the route* i to j for good n in terms of costs;
- it will be difficult to know exactly the margins (they are unitless), hence is we set

$$\mathfrak{M}^n_{i,j} := \frac{1}{2}\left(M^n_{i,j} - \sum_\ell \frac{w^n_{i,\ell}}{\sum_{\ell'} w^n_{i,\ell}} M^n_{i,\ell}\right),$$

  the best we can try is to statistically identify the weighted margins: $\mathfrak{M}^n_{i,j}$.

This means that we can write the relation

$$\begin{aligned} Q^n_{i,j}\varphi^n_{i,\cdot} &= w^n_{i,j}\ \mathfrak{M}^n_{i,j} + \sum_\ell w^n_{i,j}\left(\frac{w^n_{i,\ell}}{\sum_{\ell'} w^n_{i,\ell'}} - \mathbf{1}_{j=\ell}\right) r_\ell\ \varphi^\bullet_{\ell,\cdot} \\ &= A^n_{i,j} \qquad + \sum_\ell B^n_{i,j,\ell} \qquad\qquad\qquad\qquad \varphi^\bullet_{\ell,\cdot} \end{aligned} \tag{5}$$

That is a linear relation of the quantities to be sent (i.e. $Q^n_{i,j}\varphi^n_{i,\cdot}$) over the available quantities $\varphi^\bullet_{\ell,\cdot}$, where the constant has information on the margins, and the coefficients have information on the weights.

How to do this linear regression. Make a database of[2]

- the quantity of n to be sent (exported) from i to j: $\psi^n_{i,j}(t) := Q^n_{i,j}(t)\varphi^n_{i,\cdot}(t)$, and its average

$$\widehat{\psi^n_{i,j}}(t) := \frac{1}{T}\sum_{\tau=1}^{T} \psi^n_{i,j}(t+\tau).$$

- and the average (anticipated) congestion at ports ℓ

$$\widehat{\varphi^\bullet_{\ell,\cdot}}(t+\delta T) := \frac{1}{T}\sum_{\tau=1}^{T} \varphi^\bullet_{\ell,\cdot}(t+\delta T+\tau),$$

  where δT is the average time to a port, and T a period covering enough date to be sure it corresponds to the anticipations of the agents.

In practice we can take one month for T. It would allow to have as many observations as $J \times (J-1) \times N \times \#\{\text{dates}\}$ and obtain:

$$\widehat{\psi^n_{i,j}}(t) = a^n_{i,j} + \sum_\ell b^n_{i,j,\ell}\ \widehat{\varphi^\bullet_{\ell,\cdot}}(t+\delta T) + \epsilon^n_{i,j}(t). \tag{6}$$

Going to margins and costs. Then have to solve

$$a^n_{i,j} \simeq w^n_{i,j}\ \mathfrak{M}^n_{i,j} = \frac{v^n_j - v^n_i}{r_j + c_n g(T_{i,j})\mathbf{1}_{j\neq i}}. \tag{7}$$

That could be solved by maximum likelihood, for instance with a Gaussian assumption over the values v^n_i and v^n_j, i.e. minimizing

$$\left\| a^n_{i,j}\left(r_j + c_n g(T_{i,j})\mathbf{1}_{j\neq i}\right) - (v^n_j - v^n_i)\right\|^2$$

where $a^n_{i,j}$ and $g(T_{i,j})$ are known.[3]

The outcome would be the margins and the weights (since c_n and r_j will be known).

[2] in the setup where $\varphi^\bullet_{j,\cdot}$ is the anticipation of what is planned to be sent and not an anticipated congestion term, there would be a $\widehat{\varphi^\bullet_{\ell,\cdot}}(t)$ term in the dataset.

[3] possibly trying g as a linear or quadratic function.

2.5 How to study the parameters

One first task, already mentioned, should be to do local averages of $Q^n_{i,j}(t)$ (to stabilize it) should be tried and the time series of the first eigenvectors should be explored. This should be our $\hat{\varphi}^n_{i,\cdot}(t)$.

On another hand, if we focus then on the first equation of the system:[4]

$$(8)\qquad Q^n_{i,j}\cdot\varphi^n_{i,\cdot} = \frac{M^n_{i,j} - 2r_j\varphi^\bullet_{j,\cdot} - \lambda^n_i}{2(r_j + c_n g(T_{i,j})\mathbf{1}_{j\neq i})} \simeq \frac{M^n_{i,j}/c_n}{2g(T_{i,j})\mathbf{1}_{j\neq i}} - \frac{2r_j\varphi^\bullet_{j,\cdot} + \lambda^n_i}{2(r_j + c_n g(T_{i,j})\mathbf{1}_{j\neq i})} + \mathcal{O}\left(\frac{r^+}{c^+}\right).$$

If we focus in the first term that is in $(M^n_{i,j}/c_n)/g(T_{i,j})$ (note that $\varphi^n_{i,\cdot}$ is independent of the port of destination): it implies that the outgoing flows should linearly increase with the margin and decrease with a monotonous function of the distance. That is something that can be studied empirically:

- since the denominator is independent of the good n, we should see a similar relation between the $Q^n_{i,j}\cdot\varphi^n_{i,\cdot}$ and $T_{i,j}$ when $j\neq i$, *more or less independently of the good.* It means that averaging over goods should give us information on the shape of $g(\cdot)$. A local regression when $j\neq i$ (for instance a kernel regression) on the scatter plot of $1/(1/N\sum_n Q^n_{i,j}(t)\cdot\hat{\varphi}^n_{i,\cdot}(t))$ (or the median instead of the mean) vs. $T_{i,j}$ should display this relation.

- Once it is done, one will get $\hat{g}(T_{i,j})$ that is an estimate of $g(T_{i,j})$. Then one could have a look at the time series of $Q^n_{i,j}(t)\cdot\hat{\varphi}^n_{i,\cdot}(t)\cdot\hat{g}(T_{i,j})$ when $j\neq i$, since

$$(9)\qquad Q^n_{i,j}\cdot\varphi^n_{i,\cdot}\cdot g(T_{i,j}) \simeq \frac{M^n_{i,j}}{2c_n}\cdot\mathbf{1}_{j\neq i} - \frac{2\varphi^\bullet_{j,\cdot} + \lambda^n_i/r_j}{2(1/g(T_{i,j}) + c_n/r_j\cdot\mathbf{1}_{j\neq i})} + \mathcal{O}\left(\frac{r_n}{c_n}\right).$$

  Remember that $M^n_{i,j} = v^n_j - v^n_i$, hence if it is observed on a lot of pairs (i,j), one should get estimates of the relative values $v^n_j/(2c_n)$ and $v^n_i/(2c_n)$ and come back to an estimate for $\hat{M}^n_{i,j}$.

- This process can be augmented by focusing on the $i=j$ cases, since (8) reads (note that $\lambda^n_i/(2r_i)$ is negative when $i=j$ since $M_{i,i}=0$)

$$(10)\qquad Q^n_{i,i}\cdot\varphi^n_{i,\cdot} = (-\lambda^n_i/(2r_i)) - \varphi^\bullet_{i,\cdot}.$$

  But the difficulty there is to estimate $Q^n_{i,i}$, i.e. *the quantity that does not ship.* The main component will be the cost of crowding at port i, that may be difficult to estimate too.

This short paragraph shows that they are different ways to explore the dataset targeting to obtain rough estimates for the important quantities.

[4] considering that the congestion costs at ports $r^+ = \sup_j r_j$ are smaller than the transportation costs $c^+ = \sup_n c_n$ of goods, we could express the approximation as a function of r^+/c^+.