# Gross Output Production Functions and Spatial Dependence

Hans Martinez

2023-11-03

## To Do

- Investigate if there exists an input-output matrix by industry sector for Colombia during the 80s.
- What are these metropolitan areas? Is there geographical information, such as latitude and altitude?
- Can other datasets be linked to these areas, like a Census?
- Simulations in GNR?

## Intro

Key Idea: Review the latest spatial inference estimators on gross output production functions framework using firm-level Colombian data (à la T. Conley, Gonçalves, and Hansen 2018).

Using several dissimilarity measures, I could compare the latest spatial inference methods, including different ways of clustering the standard errors. How? I could construct dissimilarity measures using firms' characteristics such as the share of skilled and unskilled labor, or the share of imports and exports. With the dissimilarity measures on hand, I can also compare MW's (Müller and Watson 2022b, 2022a) and Kernel estimators. In addition, I can compare traditional s.e. clustering[1] with the cluster-learning method by Cao et al. (2021).

To the best of my understanding, there is no guide regarding the best practices when dealing with the inference in the estimation of gross-output production functions. The literature has focused on obtaining production function and productivity unbiased estimates.

Alternatively, the productivity variance-covariance matrix might be of interest to researchers and policy-makers. For example, the productivity variance-covariance matrix across sectors could indicate how productivity shocks diffuse through

[1] Because these variables are available in the Colombian data, it is natural for researchers to cluster the standard errors by metropolitan area or by subsectors.

other sectors across the economy. This might be of interest to policymakers because they can impulse policies that have the largest positive effect on the whole economy. The Var-Covar matrix within sectors, using skilled/unskilled labor or local/foreign purchases/sales, might inform researchers about the productivity spillover effects between firms.

In the gross output production function framework, an output shock and a productivity shock form the error term. The output shock is independent of the inputs (usually, capital, labor, and intermediates), is not serially or cross-sectionally correlated, and has a mean zero.

Productivity, on the other hand, is observed by the firm when choosing input quantities, but unobserved by the econometrician. The well-known *simultaneity* problem. In addition, productivity is assumed to follow a Markov process. The rationale is that productivity is persistent. Researchers assume in practice a linear AR(1) functional form.

Because production functions are commonly estimated by the industry sector, both the error term and productivity have sector-specific variances. Researchers assume technology in the same industry is common for all industries. Hence, firms vary only in their productivity.

I can leave the error term as a random output shock and focus on productivity. If we think about technology and how it diffuses through the economy and how innovations in one sector spill over to other sectors and companies, we can back up the estimates with a solid theoretical model.

# 1 Location and dissimilarity measure

The firm-level Colombian data contains information on the firm's characteristics, inputs, and outputs. The data also indicates the firms' metropolitan area and country region.

I could construct several dissimilarity measures:

1. Labor market (T. G. Conley, Flyer, and Tsiang 2003): The Colombian data includes firm-level labor detailed by skilled and unskilled labor, as well as managers, local and foreign technicians, their wages, and benefits.
2. Industry sectors (T. G. Conley et al. 2003): In the case of Colombian data, the firm-level survey does not detail sales per sector. However, it does include the share of local and export sales and the share of local and imported inputs. It also includes the firm's four-digit sector code; industries are usually defined at the three-digit level.
3. Firm characteristics: The data also details firms' value of capital in land, buildings, machinery, and office equipment; it also contains some details on industrial expenses such as maintenance and industrial work by other establishments; and general expenses such as publicity and insurances.

## 2 Setting

Firms produce gross output $Y_{it}$ given a production function $Y_{it} = G(K_{it}, L_{it}, M_{it})e^{\omega_{it}+\varepsilon_{it}}$ and a productivity shock $\omega_{it}$ using capital $K_{it}$, labor $L_{it}$, and intermediates $M_{it}$.

$$y_{it} = g(k_{it}, l_{it}, m_{it}) + \omega_{it} + \varepsilon_{it}$$

where $\varepsilon_{it}$ is an output shock that is not part of the information set of the firm. The productivity shock $\omega_{it}$ is known when taking input decisions, giving rise to the well-known simultaneity problem in estimating production functions. It is usually assumed that the output shock is i.i.d., and that the productivity shock follows a Markov process[2]. In practice, researchers assume an AR(1) model

$$\omega_{it} = h(\omega_{it-1}) + \eta_{it}$$

where $\eta_{it}$ is iid

### 2.1 Estimating productivity

One popular method to estimate productivity is Gandhi, Navarro, and Rivers (2020). The authors start with the firm's cost-minimization problem

$$max_{M_{it}} P_t \mathbb{E}[Y_{it}] - \rho_t M_{it} \tag{1}$$

Taking the first order condition of Equation 1 with respect to intermediates, in logs, and rearranging, we can get to the following expression

$$s_{it} = D^{\mathcal{E}}(k_{it}, l_{it}, m_{it}) + \varepsilon_{it}$$

where $s_{it} = \ln(\frac{\rho_t M_{it}}{P_t Y_{it}})$, $D(\cdot)$ is the output elasticity of the intermediates, and $\mathcal{E}$ denotes the constant resulting from firms taking the expectation over $\varepsilon_{it}$. The authors show both $D(\cdot)$ and $\mathcal{E}$ are identified under common assumptions.

Then, to recover the Markov process of productivity, note that

$$\mathcal{Y}_{it} \equiv y_{it} - \varepsilon_{it} - \int D(k_{it}, l_{it}, m_{it}) \mathrm{d}m_{it} = -\mathcal{C}(k_{it}, l_{it}) + \omega_{it}$$

The authors show that the Markov process of productivity is identified up to a constant by replacing $\omega_{it}$, as follows

[2] In T. G. Conley et al. (2003), $\omega_{it}$ is serially uncorrelated, independent of $\varepsilon_{it}$, expectation zero, but it is correlated across sectors as a function of economic distance. $\varepsilon_{it}$ is serially uncorrelated, independent across sectors with sector-specific variance.

$$\mathcal{Y}_{it} = -\mathcal{C}(k_{it}, l_{it}) + h(\mathcal{Y}_{it-1} + \mathcal{C}(k_{it-1}, l_{it-1})) + \eta_{it}$$

# 3 GNR's Monte Carlo Simulations

Gandhi, Navarro, and Rivers (2020) conduct several Monte Carlo experiments to evaluate the performance of using time-series variation as a source of identification. The authors show that relying on aggregate time-series variation provides weak identification in practice.

In the simulations, firms solve a dynamic optimization problem that is solved numerically. The authors simplify the production function to only depend on capital $K_{it}$ and intermediate inputs $M_{it}$. Different functional forms of the production function $G(K_{it}, L_{it})e^{\omega_{it}+\varepsilon_{it}}$, such as Cobb-Douglas, CES, and translog are considered.

$\varepsilon_{it}$ is distributed $N(0, \sigma_\varepsilon)$ and $\omega_{it}$ follows an AR(1) process

$$\omega_{it} = \delta_0 + \delta\omega_{it-1} + \eta_{it}$$

where $\eta_{it}$ is distributed $N(0, \sigma_\omega)$.

Investment $I_{it}$ determines the next period's capital stock via the law of motion

$$K_{it_1} + (1 - \kappa_i) + I_{it}$$

The price of output is normalized to one. The prices of investment is fixed. Finally, the price of intermediates follows an AR(1) process.

All parameters were selected to roughly match the estimates from the Chilean and Colombian datasets.

The problem is solved numerically by value function iteration.

# References


Cao, Jianfei, Christian Hansen, Damian Kozbur, and Lucciano Villacorta. 2021. "Inference for Dependent Data with Learned Clusters," July. http://arxiv.org/abs/2107.14677.

Conley, Timothy G, Bill Dupor, Gadi Barlevy, Susanto Basu, Gerald Carlino, Xiao-Hong Chen, Adeline Delavande, et al. 2003. "A Spatial Analysis of Sectoral Complementarity." *Journal of Political Economy.*

Conley, Timothy G, Fredrick Flyer, and Grace R Tsiang. 2003. "Spillovers from Local Market Human Capital and the Spatial Distribution of Productivity in Malaysia." *Advances in Economic Analysis and Policy3* 3. http://www.bepress.com/bejeap.

Conley, Timothy, Silvia Gonçalves, and Christian Hansen. 2018. "Inference with Dependent Data in Accounting and Finance Applications." *Journal of Accounting Research* 56 (September): 1139–1203. https://doi.org/10.1111/1475-679X.12219.

Gandhi, Amit, Salvador Navarro, and David A Rivers. 2020. "On the Identification of Gross Output Production Functions." *Journal of Political Economy* 128 (June): 2973–3016.

Müller, Ulrich K, and Mark W Watson. 2022a. "SPATIAL CORRELATION ROBUST INFERENCE." *Econometrica* 90: 2901–35. https://doi.org/10.3982/ECTA19465.

———. 2022b. "Spatial Correlation Robust Inference in Linear Regression and Panel Models." *Journal of Business & Economic Statistics* 00: 1–15. https://doi.org/10.1080/07350015.2022.2127737.