

# Brazilian Exports Imputation: a New Algorithm for Estimating the Municipal Production Directed at the Foreign Market

Alan Leal

Doutorando em Teoria  
Econômica na FEA/USP  
[www.alanleal-econ.com](http://www.alanleal-econ.com)

Michelle Márcia Vianna Martins

Professora da Universidade  
Federal de Viçosa

# Introdução

- Definições:
  - ① Comex Stat: Plataforma de dados do comércio exterior brasileiros.
  - ② Sistema Harmonizado: Categorização abrangente dos produtos negociados no mercado internacional.
- Fatos Estilizados:
  - ① Ao nível municipal, os dados brasileiros de comércio exterior se encontram disponíveis a nível SH4.
  - ② A nível estadual, os dados brasileiros de comércio exterior se encontram disponíveis a nível SH6.
  - ③ Contudo, esses dados não dizem respeito à origem do produto exportado - onde ele foi produzido-, mas de onde o movimento exportador (burocracia) se iniciou. Os dados focam mais nos traders do que nos produtores cuja produção é direcionada à exportação.
  - ④ Downscaling é o principal método usado atualmente para inputar exportações a nível municipal (veja Nonnemberg et al. (2024)).

# Dificuldades Principais

- Não temos as exportações brasileiras municipais a nível SH6. As exportações brasileiras municipais a nível SH4 não são diretamente relacionadas à produção exportadora, mas sim ao fato exportador.
- Como resolver isso? Empregamos um método abrangente que usa relações espaciais e downscaling para inputar as exportações municipais brasileiras ao nível SH6.
- O que este método é? Uma tentativa de inputar exportações municipais com as informações disponíveis.
- O que este método não é? Uma resposta definitiva e final ao problema. Esperamos incentivar acadêmicos a olhar para essa questão de policy tão relevante para tantos agentes.

- Exportações municipais a nível SH6 importam de um ponto de vista econômico:
  - 1 Onde a soja exportada para a União Europeia é produzida na prática? Para o compliance com a EUDR, essa questão importa.
  - 2 Em termos logísticos, importa também saber se a produção que está sendo toda "exportada" pelo município B está ocorrendo nesse município B ou nos municípios vizinhos.
  - 3 Em termos de facilitação de exportação, é relevante ao agente facilitador saber em quais municípios ele precisa atuar mais fortemente, com o risco do desconhecimento dos dados tornar seus esforços pouco efetivos.

# Como a academia trabalha usualmente com o Comex Stat?

- Exportações municipais a nível SH6 importam de um ponto de vista econômico:
  - 1 Acadêmicos geralmente não mencionam em seus trabalhos uma distinção entre exportações produzidas no município e exportações indo para o exterior, apenas como forma de registro.
  - 2 Nota Fiscal de Exportação e a Declaração Única da Exportação. Para agentes na policy, essa distinção é conhecida, ainda que não facilmente superável.

# Método I

- Define-se  $k$  como o número de bens no nível HS6 presentes em uma categoria de produtos no nível HS4, pode-se então aproximar a produção direcionada ao mercado externo em um determinado município no nível HS6 pela seguinte equação:

$$E_{i,HS6} = \underbrace{\alpha \mathbf{W} E_{i,HS4}}_{\text{Componente Espacial}} + \underbrace{r E_{i^*,HS6}^*}_{\text{Componente de Ajuste}} \quad (1)$$

Onde:

- $i$  é um município específico.
- $E_{i,HS6}$  são as exportações municipais estimadas no nível HS6.
- $E_{i^*,HS6}^*$  é definido como  $\sum_{i=1}^I \alpha (E_{i^*,HS4} - \mathbf{W} E_{i^*,HS4})$ .

## Método II

- $\alpha$  é o coeficiente de proporção em nível estadual definido como:

$$\alpha = \frac{E_{state,HS6}}{E_{state,HS4}} \quad (2)$$

onde E significa Exportações, portanto  $E_{state,HS6}$  significa as exportações estaduais de um determinado bem no nível HS6, enquanto  $E_{state,HS4}$  significa as exportações estaduais da categoria de bem no nível HS4.

- **W** é uma matriz de pesos espaciais  $n \times n$  normalizada por linha onde  $n$  é o número de municípios, tal que:

$$\sum_{j=1}^n w_{ij} = 1 \quad \forall i \quad (3)$$

$$E w_{ij} \geq 0 \quad \forall i, j$$

## Método III

- $\mathbf{r}$  é um vetor  $n \times 1$ , que soma 1. Neste trabalho, considera-se que seja a relevância de cada município na força de trabalho formal estadual no setor de soja.
- $E_{i^*,HS4}$  é um vetor  $n \times 1$  de exportações municipais observadas no nível HS4, onde \* indica que algumas entradas podem ser zero.



# Propriedades I

O método satisfaz várias propriedades-chave:

① **Consistência Dimensional:**

$$\mathbf{W}(n \times n)E_{i^*,HS4}(n \times 1) - r(n \times 1) \quad (4)$$

- ② **Propriedade de Separabilidade e Agregação:** Os componentes espacial e de ajuste somam (em todos  $k$  possíveis) um valor idêntico às exportações HS4 corretas para todo o estado. Isso acontece porque o componente de Ajuste é calculado com base na diferença entre os valores das exportações de produtos e suas estimativas. Isso nos permite calcular um valor mais direto para as exportações de um determinado bem no nível HS6. Como exemplo, Mato Grosso produz soja triturada para semeadura e não para a semeadura. O algoritmo proposto neste paper imputa as exportações municipais nesse nível (soja não e para semeadura), tal que a soma dessas exportações

## Propriedades II

municipais imputadas para os dois bens equivale às exportações de soja triturada pelo Mato Grosso. Isso configura a propriedade de agregação.

- 3 **Tratamento de Zeros:** O método naturalmente acomoda municípios com exportações zero através da estrutura de multiplicação matricial.

- 1 O único requisito que se postula sobre  $W$  é ser normalizado por linha e possuir apenas elementos não-negativos.
- 2 Neste trabalho, usamos matrizes do tipo queen e inversa da distância para demonstrar o método.
- 3 Sendo linha-normalizada, o pesquisador pode usar a matriz que julgar mais interessante para trabalhar com o bem o qual ele deseja imputar a exportação municipal.
- 4 O vetor  $r$  precisa somar 1 em todos os seus elementos. Usando o número de empregos no setor da soja para o exercício empírico demonstrando o método neste paper, mas o pesquisador pode utilizar a PAM por exemplo para obter estimativas para esse vetor. As possibilidades são amplas e o uso da RAIS neste exercício empírico não deve ser considerado como o único ou mais perfeito possível. O conhecimento setorial do pesquisado pode melhorar a capacidade do método de produzir resultados mais precisos.

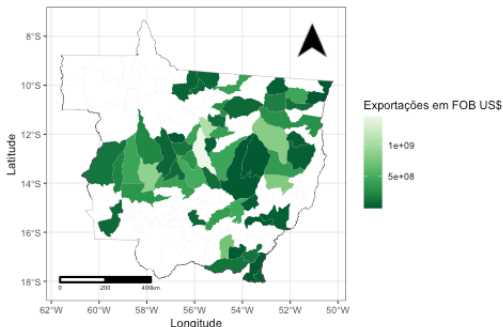
# Experimento de Monte Carlo

- 1 Como forma de validar e expandir nosso horizonte de análise - restrito empiricamente à soja no Mato Grosso-, implementamos um exercício de Monte Carlo para descobrir qual a matriz espacial que produz resultados mais precisos para diferentes estados.
- 2 Há heterogeneidade em relação à qual matriz espacial minimiza o erro.
- 3 Aleatorizamos o vetor de emprego e exportações municipais ao nível SH4 e SH6, com as exportações SH6 sendo menores que as exportações no SH4 correspondentes.
- 4 Valores positivos indicam maiores erros associados a matriz inversa da distância, enquanto valores negativos indicam maiores erros associados a matriz queen. A métrica do erro usada é o Mean Absolute Percentage Error (MAPE), adaptado pela seguinte fórmula para comparar as diferentes matrizes espaciais:

$$\frac{MAPE_i^{Inverse\ Distance} - MAPE_i^{Queen}}{MAPE_i^{Queen}}.$$

# A Soja no Mato Grosso como no Comex Stat - SH4

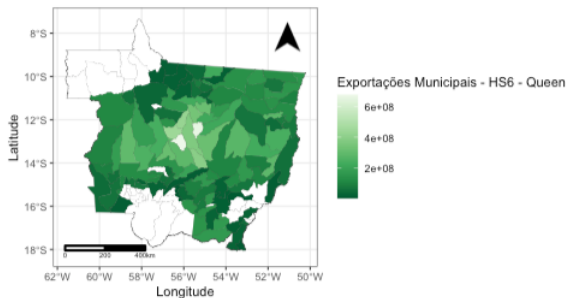
Figure: Exportações de soja do Mato Grosso no nível HS4



Fonte: Elaboração dos autores com base no Comex Stat.

# A Soja no Mato Grosso - SH6 - Queen I

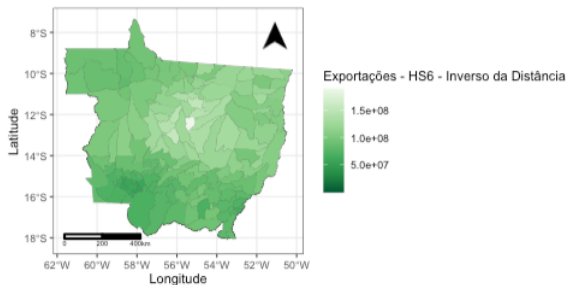
Figure: Exportações de soja do Mato Grosso (HS6) - Queen



Fonte: Elaboração dos autores com base no Comex Stat.

# A Soja no Mato Grosso - SH6 - Inverso da Distância

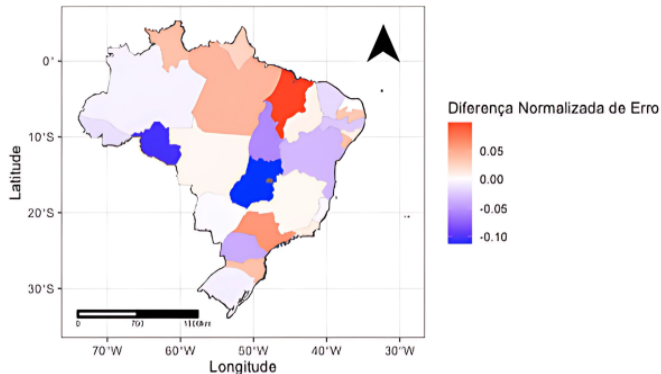
Figure: Exportações de soja do Mato Grosso (HS6) - Distância Inversa



Fonte: Elaboração dos autores com base no Comex Stat.

# Exercício de Monte Carlo com Dados Simulados I

Figure: Experimento de Monte Carlo Estadual para Diferentes Matrizes Espaciais



Fonte: Elaboração dos autores com base no Comex Stat.



# Conclusão I

- 1 O método de imputação das exportações municipais aqui não é final, nem perfeito.
- 2 Estamos lidando com dados inexistentes para todos os agentes econômicos, logo o movimento implementado por esse método é uma tentativa de reduzir incerteza para diversos agentes.
- 3 O método tem requerimentos matemáticos fortes, mas é consistente na soma para diferentes níveis de agregação e permite ao pesquisador escolher com bastante liberdade quais variáveis são mais relevantes em determinar o peso no processo de determinação dos níveis das exportações dirigidas ao mercado externo.

## Conclusão II

- 4 O exercício de Monte Carlo, cujo código se encontra disponível no paper, indica que há certa variabilidade entre os estados brasileiros de qual matriz  $W$  - queen ou inversa da distância - o pesquisador deve considerar. Exercícios de simulação são bastante relevantes para consolidar entendimentos sobre as escolhas paramétricas do modelo.
- 5 Não há método que resolva tudo, ainda mais considerando o cenário profundamente incerto e a maior dimensionalidade exigida para trabalhar com dados municipais de exportação ao nível SH6. Portanto, o uso do método deve ser cauteloso.

# References I

Nonnemberg, M. J. B., Martins, M. M. V., Cechin, A., Vianna, R. d. S., Cruz, C. C. P., Silva, F. A., Bispo, S. Q. A., and Santos, F. E. d. L. (2024). Efeitos da lei europeia contra desmatamento sobre as exportações brasileiras (publicação expressa).