FUNDAÇÃO GETULIO VARGAS
ESCOLA DE ECONOMIA DE SÃO PAULO

RAONI RIBEIRO AREDES DE OLIVEIRA

RANDOMIZATION INFERENCE IN SHIFT SHARE DESIGNS WITH AN APPLICATION IN BANKING

São Paulo

2023

RAONI RIBEIRO AREDES DE OLIVEIRA

RANDOMIZATION INFERENCE IN SHIFT SHARE DESIGNS WITH AN APPLICATION IN BANKING

Dissertação apresentada à Escola de Economia de São Paulo como pré-requisito à obtenção de título de mestre em Economia.

Campo do Conhecimento: Econometria, economia bancária.

Orientador: Bruno Ferman.

Coorientador: Daniel Ferreira Pereira Gonçalves da Mata.

São Paulo

2023

Oliveira, Raoni Ribeiro Aredes de.

Randomization inference in shift share designs with an application in banking / Raoni Ribeiro Aredes de Oliveira. - 2023.

52 f.

Orientador: Bruno Ferman.
Co-orientador: Daniel da Mata.

Dissertação (mestrado CME) – Fundação Getulio Vargas, Escola de Economia de São Paulo.

1. Econometria. 2. Estatística matemática. 3. Bancos - Brasil. 4. Desenvolvimento econômico. I. Ferman, Bruno. II. Mata, Daniel da. III. Dissertação (mestrado CME) – Escola de Economia de São Paulo. IV. Fundação Getulio Vargas. V. Título.

CDU 330.115(81)

Ficha Catalográfica elaborada por: Isabele Oliveira dos Santos Garcia CRB SP-010191/O
Biblioteca Karl A. Boedecker da Fundação Getulio Vargas - SP

RAONI RIBEIRO AREDES DE OLIVEIRA

RANDOMIZATION INFERENCE IN SHIFT SHARE DESIGNS WITH AN APPLICATION IN BANKING

Dissertação apresentada à Escola de Economia de São Paulo como pré-requisito à obtenção de título de mestre em Economia.

Campo do Conhecimento: Econometria, economia bancária.

Data de aprovação: 06/03/2023

Banca examinadora:

Prof. Dr. Bruno Ferman
FGV-EESP (Orientador)

Prof. Dr. Daniel Ferreira Pereira Gonçalves da Mata
FGV-EESP (Co-orientador)

Prof. Dr. Vitor Augusto Possebom
FGV-EESP

Prof. Dr. Pedro H. C. G. Sant'Anna
Vanderbilt University

Agradecimentos

A meus pais pelo apoio que me deram em todas as empreitadas da minha vida.

A Bruno Ferman pelas diversas oportunidades de aprendizado, tanto na execução dessa dissertação quanto acompanhando sua pesquisa e ensino.

A Daniel da Mata pela orientação acadêmica e por ter gentilmente cedido os códigos de seu trabalho que inspirou a parte empírica dessa dissertação, economizando diversos dias de coding e data collection, assim como ao Mário Dotta, que foi muito parceiro em me mostrar como esses funcionam.

Ao Banco Central, em especial aos colegas da UniBC e André Minella, por todo o apoio antes e durante esse afastamento para mestrado.

Aos colegas de mestrado pela companhia nessa empreitada. Em especial: Alei, pelos papos e simpatia sem igual, Giovanni, pelo bom humor e dicas de econometria, Maria, pelos papos diversos e dicas gastronômicas, e ao Murilo, pela afinidade, resenhas e rolês.

Aos amigos de longa data, Daniel Onoda, Eduardo Murai, Fernando Maiello, Matias Tomazelli, e Rodrigo da Costa, por ajudarem a colorir o canvas da vida há mais de 20 anos.

Aos amigos também de longa data (16 anos), mas sobretudo muitos quilômetros rodados juntos, Alexandre Gollub e Rodrigo Camilo, pela amizade que nenhuma distância mais separa.

Resumo

Esta Dissertação de Mestrado é composta de duas partes.

A primeira, apresentada no Capítulo 1, é uma contribuição teórica em econometria para Shift-Share Designs, baseada no trabalho feito em conjunto com Luis Alvarez and Bruno Ferman, “Randomization Inference Tests for Shift-Share Designs” (veja a referência Alvarez et al., 2022). Ela mostra que, escolhendo uma estatística adequadamente estudentizada para realizar Inferência por Randomização, nós somos capazes de (i) controlar o tamanho do teste em amostras finitas sob condições relativamente fortes, como efeitos homogêneos de tratamento e distribuição conhecida dos choques, e (ii) controlar o tamanho do teste assintoticamente sob condiçõesmenos exigentes, como heterogeneidade dos efeitos de tratamento e distribuição dos choques usada nas simulações (randomização) “bem comportados”, ainda que essa última seja diferente da distribuição verdadeira do processo gerador de dados.

A segunda parte, no Capítulo 2, é uma aplicação empírica dessa técnica à expansão da rede física de agências bancárias no Brasil durante o *boom* de commodities dos anos 2000 e 2010. Tenta-se medir até que ponto o aumento do número de agências foi uma resposta à maior atividade econômica. Trata-se uma pergunta interessante pois, à época, essa rede física foi um vetor de inclusão financeira. Entender até que ponto isso foi estimulado por um choque exógeno de demanda pode ser informativo para políticas públicas futuras.

Palavras-chave: shift-share designs; inferência; randomization inference; rede bancária; agências; Brazil.

Abstract

This Master Thesis is comprised of two parts.

The first one, presented in Chapter 1, is a theoretical advance in econometrics for Shift-share Designs, based on joint work with Luis Alvarez and Bruno Ferman, "Randomization Inference Tests for Shift-Share Designs" (see the reference Alvarez et al., 2022). It shows that, by choosing a properly studentized statistic for performing Randomization Inference, we are able to (i) control size in finite samples under relatively strong hypotheses, such as homogeneous treatment effects and known assignment process, and (ii) control size asymptotically under milder hypotheses, such as a "well-behaved" treatment heterogeneity and randomization distribution, even if it is different from the original assignment process.

The second part, in Chapter 2, is an empirical application of this technique to the expansion of the physical bank network in Brazil during the commodity boom of the 2000's and 2010's. It seeks to measure to what extent the increase in the number of branches in the period was a response to greater economic activity. This is of interest because, at the time, the physical network was a vector of financial inclusion. Understanding the extent to which this was spurred by exogenous demand shocks may be informative for future policy.

Keywords: shift-share designs; inference; randomization inference; bank branch network; Brazil.

JEL Classification: C15, C18, C21, C26, G21, O16.

List of Figures

List of Tables

Contents

1 Randomization Inference in Shift Share Designs

1.1 Motivation

Shift-share research designs consider instrumental variables that are constructed based on a common set of shocks that differentially affects multiple units, depending on their exposure to those shocks. Prominent examples of papers that used this methodology include Bartik (1991), Blanchard and Katz (1992), Card (2001), and Autor et al. (2013).

Adão et al. (2019) (henceforth, AKM) show that usual standard error formulas may substantially under-state the true variability of the shift-share estimator, if regions with similar exposures to the sector-level shocks also have correlated errors. AKM and Borusyak et al. (2021) (henceforth, BHJ) propose alternative estimators for the asymptotic variance of the shift-share estimator that are valid under arbitrary cross-regional correlation in the regression residuals. These methods rely on an asymptotic theory in which we have a large number of sectors, and the relevance of each sector becomes asymptotically negligible. While these methods provide reliable inference in many applications, they may lead to large over-rejection when such asymptotic theory does not provide a reasonable approximation to the empirical setting (Ferman, 2019). Borusyak and Hull (2020) (henceforth, BH) propose another alternative, based on the ideas of randomization inference (RI), that is valid even in finite samples. However, their approach relies on assumptions on the shock assignment mechanism that may be relatively harder to justify in some settings, such as, for example, knowledge of the distribution of the shocks, or that shocks are iid. Given the advantages and disadvantages of each approach, BH state that "*the choice between RI and asymptotic approaches involves tradeoffs.*"

This chapter exposes alternative inference methods based on RI that combine the advantages of the asymptotic methods proposed by AKM and BHJ, and of the RI method proposed by BH. The proposed inference methods are valid in finite samples under relatively stronger assumptions, including homogeneous treatment effects and correct specification of the distribution of the shocks up to a scale parameter. Finite sample validity also holds if, instead of correct specification of the the shocks, we can assume they are symmetric around a known mean, or identically distributed (iid). Moreover, under weaker assumptions on the treatment effects heterogeneity, and even when the distribution of the shocks is misspecified, these inference methods are *also* asymptotically valid when the number of sectors increases.[1] In brief, we provide inference methods for Shift-share Designs that are valid under relatively stronger assumptions in finite samples, but can be relaxed once the number of sectors increases. This alleviates the trade-offs between RI and asymptotic approaches mentioned by BH.

The approaches build on a large literature that studies the use of RI methods in other settings, and considers RI with studentized test statistics that are valid under stronger assumptions (or, alternatively, for inference on sharper null hypotheses) in finite samples, but also asymptotically valid under weaker assumptions (or, alternatively, for inference on less stringent null hypotheses). See, for example, Janssen (1997), Chapter 15 of Lehmann and Romano (2005), Chung and Romano (2013), Bugni et al. (2018), Wu and Ding (2021), Ferman (2021), Roth and Sant'Anna (2022).

1.2 Notation and Setup

For an outcome of interest Y, consider the structural model

$$Y(x, \epsilon; \beta) = \beta x + \varepsilon, \tag{1.1}$$

where $x \in \mathbb{R}$ denotes a treatment of interest, and ε are the remaining determinants of Y. We consider for simplicity the case without a constant and without other covariates. However, all the results remain valid for a more general setting.

We assume the existence of another set of variables called *shocks*, denoted by the vector $G = (g_1, \ldots, g_J)' \in \mathbb{R}^J$, also a variable called *shares*, denoted by $S = (s_1, \ldots, s_J)' \in \mathbb{R}_+^J$. From them, we define the *shift-share instrument*, denoted by Z, as follows

$$Z = S'G \tag{1.2}$$

Consider a sample of N units, who all share the same structural relation between the treatment X and the unobservables ε, that is

$$Y_i = Y(X_i, \varepsilon_i; \beta_i) \quad i = 1, \ldots, N, \tag{1.3}$$

[1] The RI tests we consider will be asymptotically conservative whenever the conditions stated by AKM in their Appendix A.1.6 hold. Those conditions limit the correlation between treatment effect heterogeneity and exposure weights.

with separable unobservables as in eq. (1.1). For each of them, we construct an observation of our instrument, using the same common shock G for every observation, but idiosyncratic shares, $S_i = (s_{i1}, \ldots, s_{iJ})'$:

$$Z_i = S_i' G = \sum_{j=1}^{J} s_{ij} g_j \tag{1.4}$$

We *do not* assume the observations are identically distributed nor independent. The full observable data can be written as $\left\{ \left\{ (Y_i, X_i, S_i) \right\}_{i=1}^{N}, G \right\}$.

We define a conditioning sigma-algebra, $\mathcal{A} := \{B, \mathbf{S}, \boldsymbol{\epsilon}\}$, with B, $\mathbf{S}$ and $\boldsymbol{\epsilon}$ denoting, respectively, the ensemble of the structural (causal) coefficients, of the sampled shares and of the unobservables, i.e.,

$$B := (\beta_i)_{i=1}^{N} \qquad \mathbf{S} := [S_1, S_2, \ldots, S_N]' \qquad \boldsymbol{\epsilon} := (\varepsilon_i)_{i=1}^{N} \tag{1.5}$$

We assume that the shocks G are drawn from a conditional distribution denoted by $\mathbb{P}(\cdot \mid \mathcal{A})$, such that $\operatorname{Supp}(\mathbb{P}) \subseteq \mathbb{R}^J$.

1.2.1 Identification and estimation

Along the lines of AKM, I use the Instrumental Variable (IV) as my main estimator:

$$\hat{\beta} = \hat{\beta}_{\text{IV}}(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) = \frac{\sum_{i=1}^{N} Z_i Y_i}{\sum_{i=1}^{N} Z_i X_i} \tag{1.6}$$

where $\mathbf{X} = (X_1, \ldots, X_N)'$, $\mathbf{Y} = (Y_1, \ldots, Y_N)'$ and $\mathbf{Z} = (Z_1, \ldots, Z_N)'$. The cornerstone identifying assumption is the following:

Assumption 1 (Shock exogeneity)**.** $\mathbb{E}[\, G \mid \mathcal{A}\,] = \mathbb{E}[\, G \mid B, \mathbf{S}, \boldsymbol{\epsilon}\,] = 0.$

Assumption 1 imposes that, conditional on exposures, shocks are mean-independent from unobserved determinants, with a common mean. More generally, I could have assumed that:

$$\exists\, \mu \in \mathbb{R} \quad \text{s.t.} \quad \forall\, j \in \{1, \ldots, J\}, \quad \mathbb{E}[\, g_j \mid \mathcal{A}\,] = \mu$$

In this case, the researcher may conduct inference by working with statistics that depend on *demeaned shocks* $\tilde{g}_j = g_j - \bar{g}$ (e.g. the shift-share estimator constructed with demeaned shocks). This is the solution proposed by Borusyak and Hull (2020) for inference in linear shift-share designs where the sum of exposures may be uneven across units.[2]

For the sequence, I will assume we have an estimator that is consistent for the IV estimand β given in eq. (1.7) below[3].

$$\beta := \frac{\sum_{i=1}^{N} \beta_i \mathbb{E}\left[\, X_i Z_i \mid \mathcal{A} \,\right]}{\sum_{i=1}^{N} \mathbb{E}\left[\, X_i Z_i \mid \mathcal{A} \,\right]} \tag{1.7}$$

Where needed to prove the results from this chapter, the relevant assumptions will be listed, but I consciously opt to leave out for now the technical assumptions for consistency and asymptotic normality of $\hat{\beta}$, inviting instead the interested reader to consult Adão et al. (2019) on the matter.

All the results remain valid if we consider the reduced-form equation, in which case we set $X_i = Z_i$. I also present some results that are valid only for the reduced-form case (such as the statistic T_2, further on in Section 1.3.2).

1.3 Proposed statistics

This section introduces the new studentized statistics. First, a few background assumptions

Assumption 2. *We wish to test the null hypothesis $H_0 : \beta = b$ against either a unilateral or bilateral alternative. For this, we use a test statistic $\hat{T} = \bar{T}(g, \boldsymbol{S}, \boldsymbol{X}, \boldsymbol{Y})$, such that large values of $\hat{T}$ constitute evidence against the null.*

We also define the “null-imposed residuals”, $\mathbf{e}_b$, as

$$\mathbf{e}_b = \mathbf{Y} - \mathbf{X}b \tag{1.8}$$

Under Assumption 2, homogeneous causal effects yields $\mathbf{e}_b = \boldsymbol{\epsilon}$, while heterogeneous effects yields $\mathbf{e}_b = \boldsymbol{\epsilon} + (B - b\cdot\mathbf{1}_N)\,\mathbf{X}$, with β as in eq. (1.7), $B = [\beta_1, \ldots, \beta_N]'$, and $\mathbf{1}_N$ representing a N-sized vector of ones.

---

[2] The results remain valid in such setting. Specifically, valid finite sample inference under correct specification of the shock assignment mechanism (Proposition 1) would solely require that shocks are correctly specified up to a common location shift. Another alternative would be to control for the sum of the exposures, as proposed by BHJ.

[3] This hinges on Assumption 1.

On our test statistics $\hat{T} = \bar{T}(g, \boldsymbol{S}, \boldsymbol{X}, \boldsymbol{Y})$, we make the following assumption:

Assumption 3. *Under the null $H_0 : \beta = b$, there exits some other map T such that*

$$\bar{T}(g, \mathbf{S}, \mathbf{Y}, \mathbf{X}) = T(g, \mathbf{S}, \mathbf{Y} - \mathbf{X}b) = T(g, \mathbf{S}, \mathbf{e}_b)$$

That is, under H_0, the test-statistic depends on $\mathbf{Y}$ and $\mathbf{X}$ solely through the null-imposed residuals, $\mathbf{e}_b$.

Considering the estimator $\hat{\beta}$ from eq. (1.6), Adão et al. (2019) show that, when shocks are independent and the importance of each sector becomes asymptotically negligible,

$$\forall\, c \in \mathbb{R}\,,\ \lim_{N,J\to\infty} \Pr\left(V^{-\frac{1}{2}}(\hat{\beta} - \beta) \leq c \;\middle|\; B, \mathbf{S}, \boldsymbol{\epsilon} \right) \xrightarrow{p} \Phi(c)$$
$$\text{whith} \qquad V = \frac{\left(\sum_{i=1}^{N} \varepsilon_i S_i\right)' \operatorname{Var}\left(g \mid B, \mathbf{S}, \boldsymbol{\epsilon}\right) \left(\sum_{i=1}^{N} \varepsilon_i S_i\right)}{\left(\sum_{i=1}^{N} Z_i X_i\right)^2}$$

While this approach provides reliable inference in empirical applications with many sectors, we may have relevant size distortions when there are few or concentrated sectors (see, for example, Ferman, 2019).

Adão et al. (2019) propose a consistent estimator for V using the usual regression residuals $\hat{\varepsilon}_i := Y_i - \hat{\beta} X_i$, as follows:[4]

$$\hat{V} = \frac{\sum_{j=1}^{J} g_j^2 \left(\sum_{i=1}^{N} \hat{\varepsilon}_i s_{ij}\right)^2}{\left(\sum_{i=1}^{N} Z_i X_i\right)^2} = \frac{\sum_{j=1}^{J} g_j^2 \left(\sum_{i=1}^{N} \left(Y_i - \hat{\beta} X_i\right) s_{ij}\right)^2}{\left(\sum_{i=1}^{N} Z_i X_i\right)^2} \tag{1.9}$$

They also propose a consistent estimator with the null-imposed, using the null-imposed residuals from eq. (1.8)

$$\hat{V}_b = \frac{\sum_{j=1}^{J} g_j^2 \left(\sum_{i=1}^{N} e_{b,i}\, s_{ij}\right)^2}{\left(\sum_{i=1}^{N} Z_i X_i\right)^2} \tag{1.10}$$

In their paper, BH propose to do inference based on the following statistic

$$\hat{T}_0 = \bar{T}_0(G, \boldsymbol{S}, \boldsymbol{X}, \boldsymbol{Y}) = \frac{1}{N} \sum_{i=1}^{N} (S_i' G)(Y_i - bX_i), \tag{1.11}$$

[4] Crucially, they assume conditional independence of the shocks, $g_j \mid B, \mathbf{S}, \boldsymbol{\epsilon}$, among other regularity conditions which I temporarily omit. Please refer to Adão et al. (2019) for more details.

which depends on $(\boldsymbol{X}, \boldsymbol{Y})$ solely through the null-imposed residuals $\boldsymbol{e}_b$, so that Assumption 3 holds.

Our contribution will be to take a different approach, with two studentized statistics, as follows.

1.3.1 Null-imposed statistic

We use the null-imposed variance estimator from (1.10) to create our first test statistic, T_1:

$$\hat{T}_1 = \bar{T}_1(G, \boldsymbol{S}, \boldsymbol{X}, \boldsymbol{Y}) = \frac{\hat{\beta} - b}{\sqrt{\hat{V}_b}} = \frac{\sum_{i=1}^N (S_i' G)\, e_{b,i}}{\sqrt{\sum_{j=1}^J \left(\sum_{i=1}^N e_{b,i} s_{ij}\right)^2 g_j^2}}$$

Notice that $\bar{T}_1(g, \boldsymbol{S}, \boldsymbol{X}, \boldsymbol{Y})$ follows Assumption 3, so that we can write it as

$$\hat{T}_1 = T_1(G, \boldsymbol{S}, \mathbf{e}_b) = \frac{\sum_{i=1}^N (S_i' G)\, e_{b,i}}{\sqrt{\sum_{j=1}^J \left(\sum_{i=1}^N e_{b,i} s_{ij}\right)^2 g_j^2}} \tag{1.12}$$

1.3.2 Unrestricted statistic

Using (1.9), we define a test statistic T_2 as

$$\hat{T}_2 = \bar{T}_2(g, \boldsymbol{S}, \boldsymbol{X}, \boldsymbol{Y}) = \frac{\hat{\beta} - b}{\sqrt{\hat{V}}} = \frac{\sum_{i=1}^N (S_i' G)\, e_{b,i}}{\sqrt{\sum_{j=1}^J \left(\sum_{i=1}^N \left((b - \hat{\beta}) X_i + e_{b,i}\right) s_{ij}\right)^2 g_j^2}}$$

Notice that, if $X_i \neq Z_i$, $\hat{T}_2$ will depend on $\mathbf{X}$, violating Assumption 3. We therefore restrict its use to the reduced form, with $X_i = Z_i$, so that

$$\hat{T}_2 = T_2(g, \boldsymbol{S}, \mathbf{e}_b) = \frac{\sum_{i=1}^N (S_i' G)\, e_{b,i}}{\sqrt{\sum_{j=1}^J \left(\sum_{i=1}^N \left((b - \hat{\beta})(S_i' G) + e_{b,i}\right) s_{ij}\right)^2 g_j^2}} \tag{1.13}$$

I have omitted the dependency of $\hat{\beta}$ on the data for conciseness, but keep in mind that, when $Z_i = X_i$, we have dependency only on G, $\boldsymbol{S}$, and $\mathbf{e}_b$, as desired:

$$\hat{\beta} = \hat{\beta}_{\text{IV}}(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) = \hat{\beta}_{\text{IV}}(\mathbf{S}G,\ b\mathbf{S}G + \mathbf{e}_b,\ \mathbf{S}G)$$

1.3.3 Algorithm

The algorithm for conducting inference with T_k, with $k = 1, 2$, is as follows:

(a) From a chosen distribution $\mathbb{H}\left(\cdot \mid \mathbf{S}, \boldsymbol{e}_b\right)$, draw L simulated shocks $(G^*_\ell)_{\ell=1}^L$

(b) Calculate $(T^*_{k,\ell})_{\ell=1}^L$ using $T_k(G^*_\ell, \boldsymbol{S}, \mathbf{e}_b)$

(c) Conduct inference by comparing $\hat{T}_k = T_k(G, \boldsymbol{S}, \mathbf{e}_b)$ with the appropriate quantiles of $(T^*_{k,\ell})_{\ell=1}^L$

Notice that

- For T_1, step 2 above is the same as constructing L simulated instruments $\mathbf{Z}_\ell = \mathbf{S}G^*_\ell$, running the IV regression, and calculating $(\hat{\beta} - b)/\hat{V}_b$.
- For T_2, forcefully a reduced-form case, step 2 is the same as constructing L simulated regressors $\mathbf{Z}^*_\ell = \mathbf{S}G^*_\ell$, generating $\mathbf{Y}^*_\ell = b\mathbf{Z}^*_\ell + \mathbf{e}_b$, running the OLS regression of Y on Z, and calculating $(\hat{\beta} - b)/\hat{V}$.
- For the usual null of $b = 0$, calculations of T_2 involve fewer steps, as $\mathbf{Y}^*_\ell = \mathbf{e}_b$ for every ℓ.

1.4 Finite-sample results

Suppose now that the researcher has a guess on the shock assignment mechanism, i.e on the conditional probabilities $\mathbb{P}\left(\cdot \mid \mathcal{F}\right)$. We denote this guess by $\mathbb{H}$, which specifies a c.d.f. for each $(\mathbf{s}, \mathbf{e})$ in the support of $(\mathbf{S}, \boldsymbol{\epsilon})$. In this case, for any given statistic T under H_0 from Assumption 2 and homogeneous treatment effects $\beta_{ij} = b$, the researcher is able to compute critical values by analysing the quantiles of

$$H_b\left(c \mid \mathbf{S}, \mathbf{X}, \mathbf{Y}\right) := \int \mathbf{1}_{\left\{\hat{T} \,\leq\, c\right\}} \mathbb{H}\left(dg \mid \mathbf{S}, \boldsymbol{e}_b\right) \tag{1.14}$$

Such quantity is easily estimable by simulation. Indeed, if we are able to draw L independent draws G^*_ℓ, $\ell = 1, \ldots, L$, from $\mathbb{H}\left(\cdot|\boldsymbol{S}, \boldsymbol{e}_b\right)$, then $H_b\left(c \mid \boldsymbol{S}, \boldsymbol{X}, \boldsymbol{Y}\right)$ may be estimated as

$$\hat{H}_b\left(c \mid \mathbf{S}, \mathbf{X}, \mathbf{Y}\right) = \frac{1}{L+1}\left(1 + \sum_{\ell=1}^{L} \mathbf{1}_{\{T(G^*_\ell, \mathbf{S}, \mathbf{e}_b) \leq c\}}\right) \tag{1.15}$$

If the shock-assignment process is correctly specified, then the procedure above provides valid inference, which is summarised in Proposition 1 below.

Proposition 1. *Under Assumptions 2 and 3, if* $\mathbb{H} = \mathbb{P}$*, then, under the null* $H_0 : \beta = b$ *and homogeneous causal effects* $\beta_i = \beta$*,*

$$H_b\left(c \mid \mathbf{S}, \mathbf{X}, \mathbf{Y}\right) = \mathbb{P}\left(\hat{T} \leq c \mid \mathbf{S}, \boldsymbol{\epsilon}\right)$$

Proof. See Appendix A.1.1. □

Here, Assumption 3 was necessary because we have assumed knowledge of the conditional distribution $\mathbb{P}$ only up to knowledge of the pair $(\mathbf{S}, \boldsymbol{\epsilon})$, and not for every combination of $(\mathbf{Y}, \mathbf{X})$ such that $\mathbf{Y} - \mathbf{X}b = \boldsymbol{\epsilon}$. This allows us to use knowledge of the residuals under the null, along with a sensible guess $\mathbb{H}$, to calculate these quantiles.

Although fundamental, Proposition 1 is much more useful for the following corollaries.

Corollary 1.1. *Under the same assumptions as Proposition 1 and for* $\alpha \in (0, 1)$*, a test that rejects the null of* $H_0 \colon \beta = b$ *if* $\hat{T}$ *exceeds the* $1 - \alpha$ *quantile of* $H_b\left(c \mid \mathbf{S}, \mathbf{X}, \mathbf{Y}\right)$ *is level* α*, conditional on* $(\mathbf{S}, \mathbf{X}, \mathbf{Y})$*.*

Proof. See Appendix A.1.2. □

Corollary 1.2. *For* $\alpha \in (0, 1)$*, let*

$$k_\alpha := 1 + \frac{2}{L+1} - \frac{\lfloor \alpha(L+1) \rfloor}{L+1}$$

Under the same assumptions as Proposition 1, a test that rejects the null if $\hat{T}$ *exceeds the* k_α *quantile of* $\hat{H}_b$ *is conditionally level* α*.*

Proof. See Appendix A.1.3. □

Corollary 1.2 imposes some restrictions on the sample size which we do not address here because they are rather trivial. For instance, for $\alpha = 0.05$ and $L < 40$, we have $k_\alpha \geq 1$, yielding a trivial test that never rejects. However, even a very conservative $\alpha = 0.001$ needs a sample of at least 2000 simulations to be non-trivial. As of 2023, those sample sizes can be simulated even with very modest hardware.

1.4.1 Group transformations

Instead of assuming that the shock-assignment mechanism is known, an alternative would be to consider a group $\boldsymbol{H}$ of transformations on $\mathbb{R}^J$ such that, under the null, for any $h \in \boldsymbol{H}$, $\bar{T}(h(G), \mathbf{S}, \boldsymbol{\epsilon})|\mathbf{S}, \boldsymbol{\epsilon} \overset{d}{=} \bar{T}(G, \boldsymbol{S}, \boldsymbol{\epsilon})|\mathbf{S}, \boldsymbol{\epsilon}$. In these settings, it follows from well-established results on randomization tests (Lehmann and Romano, 2005, Theorem 15.21) that the procedure described in Proposition 1 (and its Corollaries) remains valid if simulated shocks are constructed as $G^* = \boldsymbol{h}(G)$, where $\boldsymbol{h} \sim \text{Uniform}(\boldsymbol{H})$, independently from the data. For example, if, conditional on $(\mathbf{S}, \boldsymbol{\epsilon})$, shocks were assumed independently drawn from symmetric distributions with known common symmetry point m, then one could take the group of transformations to be recentred sign changes, i.e. $h(g) = \kappa \odot (g - m \cdot \iota_J) + m \cdot \iota_J$ for $\kappa \in \{-1, 1\}^J$, where $\odot$ denotes entry-by-entry multiplication and ι_J is a J dimensional vector of ones.[5] BH consider this kind of simulations in their Appendix D4. Similarly, if, conditional on $(\mathbf{S}, \boldsymbol{\epsilon})$, shocks were assumed to be iid, then one could take the group to be the set of permutations of a J-dimensional vector, as also discussed by BH.

1.4.2 Scale invariance

We observe that, when inference is based on the test-statistics $\hat{T}_1$ or $\hat{T}_2$, the requirement in Proposition 1 may be weakened to: the distribution of shocks G is correctly specified, up to multiplication of g by a positive scalar. Indeed, test statistics $\hat{T}_1$ and $\hat{T}_2$ are invariant to multiplication of the shocks by a common positive constant. This contrasts with the test statistic $\hat{T}_0$, which requires the researcher to correctly specify the scale of shocks.

1.5 Asymptotic results

This section shows the asymptotic properties of the methods proposed in section 1.3.3. This is done as J is allowed to grow. The cross-section size, N, is implicitly allowed to grow as well, as in $N = N(J)$.

[5] If the symmetry point were estimated (for example, by using the sample mean as an estimator of m), then the simulation procedure would no longer retain finite sample validity. In this case, conservative inference could be conducted by computing p-values under different choices of m, as m varies over a valid confidence set, and then taking the supremum and adding one minus the confidence of the confidence set to it (Berger and Boos, 1994). See Proposition S6 in BH for details.

Consider a sequence of structural equations indexed by J, with a cross section of size $N(J)$. In other words, $i = 1, \ldots, N(J)$.

$$Y_{J,i} = \beta_{J,i} X_{J,i} + \varepsilon_{J,i} \tag{1.16}$$

Let the IV estimand be denoted by β_J:

$$\beta_J := \frac{\sum_{i=1}^{N(J)} \beta_{J,i} \mathbb{E}\left[\, X_{J,i}\, Z_{J,i} \,\right]}{\sum_{i=1}^{N(J)} \mathbb{E}\left[\, X_{J,i}\, Z_{J,i} \,\right]} \tag{1.17}$$

with

$$Z_{J,i} = S'_{J,i}\, G_J = \sum_{j=1}^{J} s_{J,ij} g_{J,j}$$

Inserting it into eq. (1.16) yields

$$Y_{J,i} = \beta_J X_{J,i} + (\beta_{J,i} - \beta_J)\, X_{J,i} + \varepsilon_{J,i}$$

Let B_J, $\mathbf{S}_J$ and $\boldsymbol{\epsilon}_J$ by the analogues of those in equation (1.5), i.e.

$$B_J := (\beta_{J,i})_{i=1}^{N(J)} \qquad \mathbf{S}_J := [S_{J,1}, S_{J,2}, \ldots, S_{J,N}]' \qquad \boldsymbol{\epsilon}_J := (\varepsilon_{J,i})_{i=1}^{N}$$

For Propositions 2 (asymptotic normality of T_1^*) and 3 (T_2^*) that follow I will mostly omit the indexing on J, in the interest of clarity. Keep in mind that *all* quantities are a function of J, the total number of shocks, in those propositions.

Proposition 2. *(Asymptotic normality of T_1) Let $\nu := \sum_{j=1}^{J} \left(\sum_{i=1}^{N} s_{ij}\right)^2$, and define the σ-algebra (conditioning set) $\mathcal{F} := \{B_J, \mathbf{S}_J, \boldsymbol{\epsilon}_J, G_J, \mathbf{X}_J\}$. Also, let $e_{b,i} = Y_i - bX_i$ as in eq. (1.8). If we have that $N(J) \to \infty$ as $J \to \infty$, and $G^* \mid \mathcal{F}$ sampled with $g_j^* \perp\!\!\!\perp g_k^* \mid \mathcal{F}$ for $j \neq k$, plus*

(i) $\frac{1}{\sqrt{\nu}} \sum_{j=1}^{J} \left(\sum_{i=1}^{N} s_{ij} e_{b,i}\right) \mathbb{E}\left[\, g_j^* \mid \mathcal{F} \,\right] \xrightarrow{p} 0$

(ii) $\frac{1}{\nu} \sum_{j=1}^{J} \left(\sum_{i=1}^{N} s_{ij} e_{b,i}\right)^2 \mathbb{E}\left[\, (g_j^*)^2 \mid \mathcal{F} \,\right] \xrightarrow{p} \sigma^2 > 0$

(iii) $\frac{1}{\nu^2} \sum_{j=1}^{J} \left(\sum_{i=1}^{N} s_{ij} e_{b,i}\right)^4 \mathbb{E}\left[\, (g_j^*)^4 \mid \mathcal{F} \,\right] \xrightarrow{p} 0$

Then, for $T_1^* = T_1(G^*, \boldsymbol{S}, \mathbf{e}_b)$,

$$\forall\, c \in \mathbb{R}, \quad \lim_{J\to\infty} \Pr\left(\, T_1^* \leq c \mid \mathcal{F} \,\right) \xrightarrow{p} \Phi(c)$$

Proof. See Appendix A.2.1. □

Proposition 2 provides high-level conditions for conditional asymptotic normality of the simulated statistic. In Appendix B.1, I show that these conditions are satisfied for three examples of simulation distributions:

(i) when we sample with replacement from the (recentered) empirical distribution of shocks;

(ii) when we consider shocks iid $N(0,1)$, independently from $\mathbf{X}$ and G_J; and

(iii) when we consider sign-changes of observed shocks.[6]

The crucial point is that we consider a studentized test statistic, so the simulated test statistic is asymptotically $N(0,1)$. In contrast, if we considered alternative test statistics, such as T_0, then we would not reach this conclusion. Studentizing the test statistic using robust standard errors (e.g. using Eicker-Huber-White standars errors of White, 1980) would also generally not work.

Next, we analyze the test statistic T_2. In this case, since the shift-share estimator is being recomputed across samples and then used in the calculation of the standard error, we need to ensure that $\hat{\beta}^*$, the simulated shift-share estimator from the algorithm in section 1.3.3, is consistent at a given rate. In addition to the assumptions in Proposition 2, I require a "strong simulated shock" assumption that ensures that the variance of the simulated shift-share regressor does not vanish asymptotically; as well as conditions that ensure the estimation error of the standard error vanishes.

Proposition 3. *(Asymptotic normality of T_2) Under $X_i = Z_i$, consider*

$$Z_i = S_i' G = \sum_{j=1}^{J} s_{ij}\, g_j$$
$$e_{b,i} = Y_i - bX_i = \varepsilon_i + (\beta_i - b)X_i \qquad \textit{as in eq. (1.8)}$$

Under the conditions (i), (ii), (iii) from Proposition 2, plus

[6] In the Appendix, we consider sign changes without recentering shocks ($m = 0$). We note, however, that convergence would hold for any choice of recentering parameter m, including the case in which it is misspecified, and the case in which m is replaced by an estimator such as the sample mean of shocks; provided we work with the shift-share estimator that uses demeaned shocks (as per footnote 2).

(iv) $\frac{1}{N}\sum_{i=1}^{N}(Z_i)^2 \xrightarrow{p} \pi^\star > 0$

(v) $\frac{1}{\nu}\sum_{j=1}^{J}\left[\sum_{p=1}^{N}\sum_{q=1}^{N} s_{pj}s_{qj}\left(Z_p e_{b,q} + Z_q e_{b,p}\right)(g_j^*)^2\right] = o_p\left(\frac{N}{\sqrt{\nu}}\right)$

(vi) $\frac{1}{\nu}\sum_{j=1}^{J}\left[\sum_{p=1}^{N}\sum_{q=1}^{N} s_{pj}s_{qj}Z_pZ_q(g_j^*)^2\right] = o_p\left(\frac{N^2}{\nu}\right)$

Then, for $T_2^* = T_2(G^*, \boldsymbol{S}, \mathbf{e}_b)$,

$$\forall\, c \in \mathbb{R}, \quad \lim_{J\to\infty} \Pr\left(T_2^* \leq c \mid \mathcal{F}\right) \xrightarrow{p} \Phi(c)$$

Proof. See Appendix A.2.2. □

Appendix B.2 shows a discussion assumptions (iv)-(vi) of the proposition in the context of our three examples of simulation distributions.

1.5.1 Treatment Effect Heterogeneity

Notice that Propositions 2 and 3 do not assume homogeneous treatment effects, but merely limit heterogeneity through $e_{b,i} = (b - \beta_i) X_i + \varepsilon_i$. As a byproduct, whenever inference based on $\hat{T}_{1,2}$ and normal critical values provides asymptotically conservative inference, the simulation-based approach will also lead to asymptotically conservative inference. This is summarized in the Lemma below.

Lemma 1. *Suppose that, under the null $\beta_J = b_J$, there exists $v \geq 1$ such that, for every $c \in \mathbb{R}$, $\mathbb{P}[\hat{T}_1 \leq c] \to \Phi(v \cdot c)$. Assume that the conditions in Proposition 2 hold under the null. Then the simulation-based approach to inference of section 1.3.3 using the T_1 statistic will be asymptotically conservative for any significance level $\alpha < 0.5$, in the sense that, under the null, the probability of rejecting the null converges to a number smaller than the nominal significance level.*

If there exists $\tilde{v} \geq 1$ such that, for every $c \in \mathbb{R}$, $\mathbb{P}[\hat{T}_2 \leq c] \to \Phi(\tilde{v} \cdot c)$ and we assume that the conditions in Proposition 3 hold under the null, the same is true for randomization inference using the T_2 statistic.

When there is no treatment effect heterogeneity, it follows that, under the conditions in AKM and BHJ, $v = 1$. These conditions include that shocks are independent, that the number of sectors increase, and that the relevance of sectors are asymptotically negligible (AKM and BHJ consider alternatives that relax the assumption that shocks are independent, and we discuss that in Remark 2 below). In this case, inference based on the simulation approach is asymptotically size α. More generally, when there is treatment effect heterogeneity, AKM provide sufficient conditions for inference based on $\hat{T}_1$ and $\hat{T}_2$ and normal critical-values being conservative (i.e. $\upsilon \geq 1$)[7]. **These conditions limit the correlation between treatment effect heterogeneity and exposure weights.** In this case, the simulation-based approach will also lead to conservative inference. Notice that, in contrast to the finite sample results, which require homogeneous treatment effects, the method may be able to provide asymptotically conservative inference under treatment effect heterogeneity.

1.5.2 Other considerations

Remark 1. (Power) Notice that the statement of Propositions 2 and 3 do not require the null to be true. Specifically, if their conditions can be shown to be valid under a given sequence of alternatives[8] then it follows that the distribution of the simulated statistic converges to a standard normal along such sequence. In this case, the power of the null-imposed t-test and our simulation-based approach coincide asymptotically along this sequence.

Remark 2. (Clustered shock assignment) Suppose that instead of assuming that shocks are independent, we consider that we have clusters of shocks that are independent, but that there may be correlation between shocks within the same cluster. In this case, the results from Propositions 2 and 3, and from Lemma 1 should remain valid if we studentized the test statistic using AKM and BHJ standard errors with clusters of shocks, provided the number of clusters is large. We may also consider using a distribution for the simulated shocks that allows for correlation within clusters.

1.6 Discussion & Conclusion

Considering the problem of inference in shift-share research designs, there are two main existing approaches that allow for unrestricted spatial correlation. The RI approach is valid even with relatively few or concentrated shocks, but relies on relatively strong assumptions on the shock assignment process and on treatment effect heterogeneity. In

[7] See Section A.1.6 of their online supplement

[8] See Appendices B.1 and B.2 .

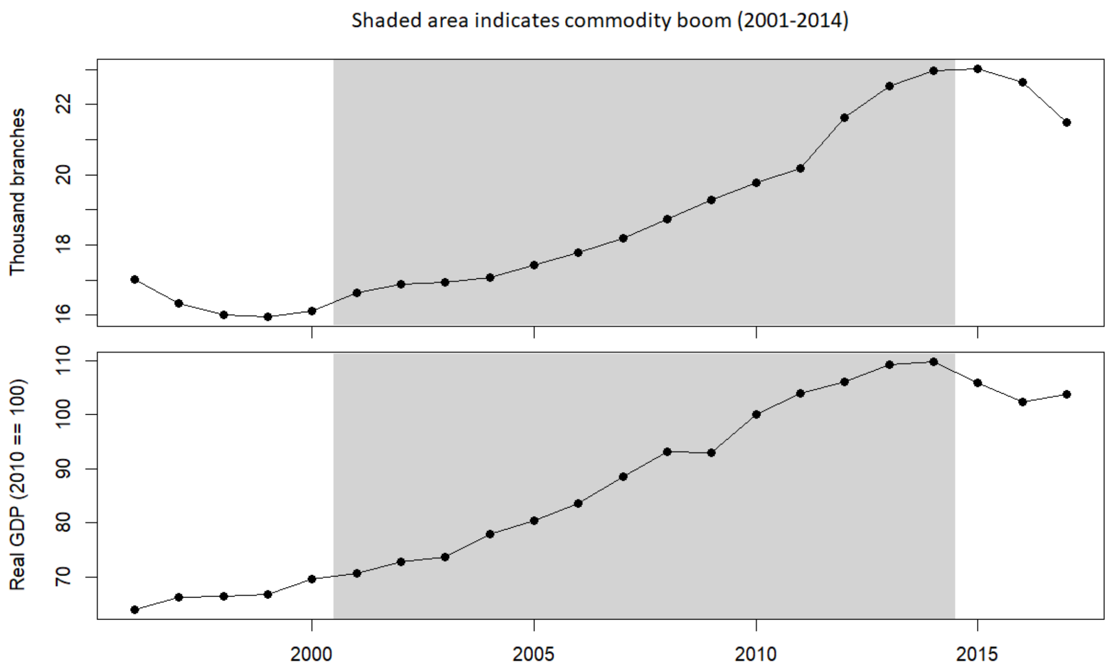
contrast, the asymptotic approach relies on weaker assumptions on the shock assignment process and on treatment effect heterogeneity, but asymptotic approximations may be inaccurate in some applications.

This chapter proposes alternative RI methods that combine the advantages of both approaches. More specifically, these inference methods are exact under relatively strong assumptions, and also asymptotically valid under weaker assumptions. The latter is achieved through studentization, which ensures convergence of the simulated distribution of the test statistic to a standard normal under mild regularity conditions.

2 Economic growth and bank branches

2.1 Motivation

Figure 1 – Bank Branches and Economic Activity



The goal of this chapter is to try to measure to what extent the bank expansion was a response to greater economic activity in the years 2002-2014. Between December 2002 and December 2014, the number of physical bank offices (*"agências"*) in Brazil grew by 36%, from 16,996 to 23,090 total, while total population increased by 13% in the same period. This increase in the bank network happened alongside a sizeable expansion of the Brazilian economy. In the 10 years between 2004 and 2013, Brazil grew 48%, averaging 4% per year, driven greatly by booming commodity prices abroad.

The presence of physical bank branches, in the beginning of the 20th century, was a good indicator of the offer of financial services to the general population. This, in turn is a driver of financial inclusion, considered a cornerstone of economic development (Demirgüç-Kunt et al., 2022, 2017). In a recent paper, Célerier and Matray (2019) show that the expansion of bank branches in the US between 1995 and 2004 increased low-income household financial inclusion, with positive effects on wealth accumulation, investment in durable assets, and probability of facing financial strain.

In Brazil, Resolution 2,640 from 1999 regulated the use of banking correspondents. These are are non-financial companies, for example retail outlets and postal offices, with which banks were allowed to make contractual arrangements for the offer of their financial services. Assunção (2013) argues that the use of correspondents brought to effectively zero the marginal cost of servicing a new municipality, as it eliminated the cost of installing a physical branch.

However, correspondents are not a perfect substitute for the bank itself. They are mainly used for transactional banking services, such as paying bills and depositing money. Although they can receive credit applications, deciding upon them remained the job of the contracting bank, and remained mostly an operation done in a physical branch (4 out of every 5 operations, cf. Table 2.1).

Table 1 – Credit Operations by Channel in 2011. Source: BCB

Branches	ATM	Phone	Correspondent	Internet	Mobile
458,888	75,624	7,393	12,987	33,607	214
78%	13%	1%	2%	6%	0%

Going back to the economic principles of financial intermediation, banks perform not only a liquidity transformation, but also manage information asymmetries between the savers and the potential borrowers. In that aspect, a closer relationship between the bank and the applicants might help, in particular in places where credit history is scarce – which is precisely the case when talking about financial inclusion.

Ergungor (2010), for example, showed that the presence of a bank branch in a low-income neighbourhood increases mortgage origination and decreases spreads, and this effect is vanishing with the distance of the borrower to the branch. He argues precisely for the better ability to collect "soft information" about applicants when the bank has a physical presence.

To summarise, understanding how the bank services offer reacts to the economic environment is a relevant policy question, as it may help guiding the Brazilian financial inclusion agenda.

2.2 Empirical Approach

Although the relation between growth and the provision of financial services may look uncontroversial at first, measurement of causality demands some care.

Within a country, banks are drawn to fast growing regions as these have a higher demand for financial services and borrowers exhibit lower credit risk, as incomes are high and unemployment is low. So growth may cause a bank expansion.

On the other hand, banking services such as transaction accounts, cash management services, and credit may aid economic growth, as they increase consumption and investment in the short run. So bank expansion may cause growth.

There is also the problem of confounders: legislative reforms might have prompted growth and financial development at the same time. A good example is the Bankruptcy Law of 2005, but other might exist. Greater human capital might also influence both.

Our empirical strategy takes into account that the Brazilian municipalities were differentially exposed to the global commodities price boom of the beginning of this century. Since we are looking within the same country, we can place banking regulatory issues aside, as they are uniform across regions. This allows us to compare different locations on the basis of an external (and arguably exogenous) demand shock.

We propose a Shift-share instrument, as first adopted by Bartik (1991), and recently formalized by Adão et al. (2019), Goldsmith-Pinkham et al. (2020) and Borusyak et al. (2021). In particular, we borrow the methodology from Da Mata and Dotta (2021), as will be explained in the sequence.

My main specification is the following Instrumental Variable setup, where eq. (2.1) is the structural (causal) equation, and eq. (2.2) is the first stage:

$$\Delta b_i = \alpha + \beta \Delta \ln Y_i + \gamma S_i + \varepsilon_i \tag{2.1}$$

$$\Delta \ln Y_i = \tilde{\alpha} + \delta Z_i + \tilde{\gamma} S_i + \zeta_i \tag{2.2}$$

The outcome, Δb_i, is the variation in bank branches $B_{i,\text{year}}$ per 10k pop.: $\Delta b_i = 10,000(B_{i,2015} - B_{i,2002})/\text{pop}_{2010}$. Our treatment, $\Delta \ln Y_i$, is the variation in real local GDP Y_i: $\Delta \ln Y_i = \ln(Y_{2014}) - \ln(Y_{2002})$.

The variable Z_i denotes our instrument, given by $Z_i = \sum_j s_{ij} G_j$, with $G_j = \ln(P_j^{2002-2013}) - \ln(P_j^{1995-2001})$ being the relative real change in average prices for each

Figure 2 – Normalized prices of considered commodities (in Nominal USD)

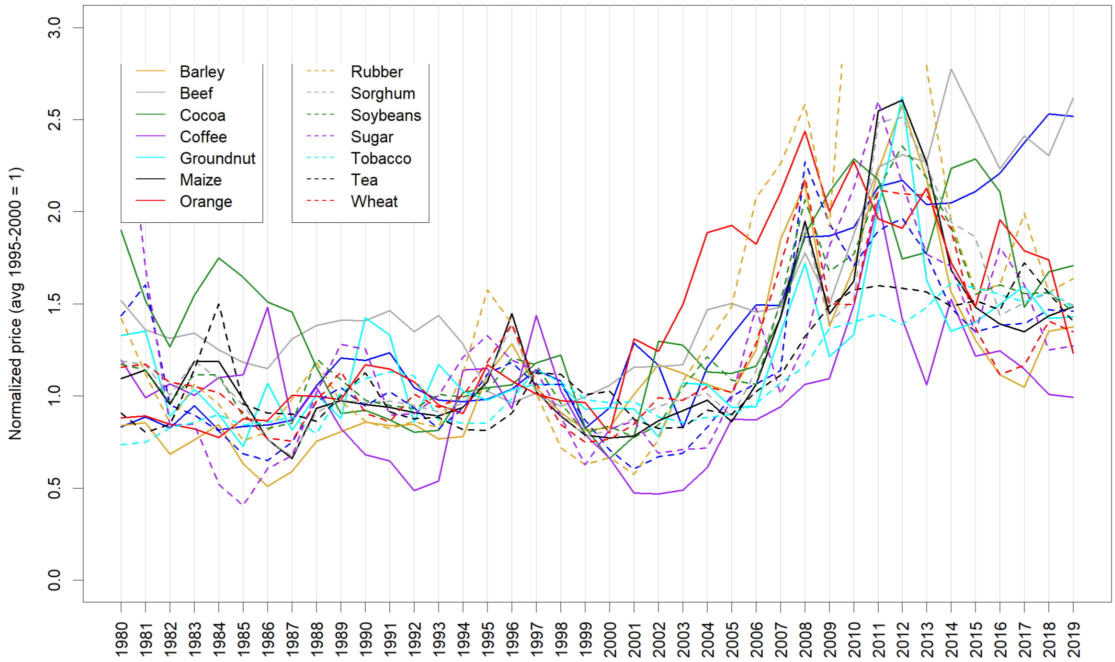
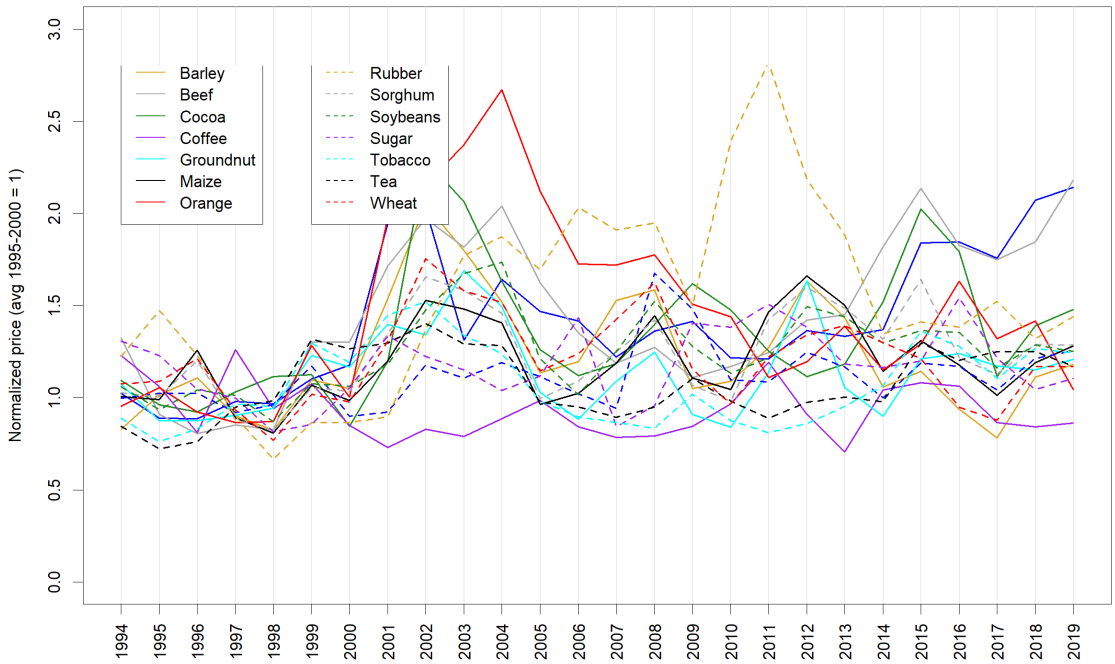


Figure 3 – Normalized prices of considered commodities (in 2010 BLR)



commodity, comparing the periods $1995 - 2001$ (pre-boom) and $P_j^{2002-2013}$ (boom). The s_{ij} are the exposures of each municipality i to a commodity j, which will be explained in the next section.

Finally, we add the sum of the exposures for each municipality, $S_i = \sum_j s_{ij}$, to account for the fact that the municipalities are exposed to shocks beyond the considered crops, that may have a shift-share structure. An example could be import penetration in manufactured goods. For more thorough exposition of why this is necessary, please refer to section "SSIVs with Incomplete Shares" in Borusyak et al. (2021).

2.2.1 Commodity exposure

The exposures s_{ij} are based on the work of Da Mata and Dotta (2021), with modifications to better suit our current work.

The potential yield of crop j in municipality i is denoted A_{ij}. This means that, if the municipality i devoted its entire land to that culture, it would produce A_{ij} metric tonnes of commodity j per year. This figure takes into account only climatic conditions (temperature, rain, and humidity). I avoid using the potential yield taking land into consideration because land modifications, such as irrigation, levelling or terrassing, might be endogenous to other characteristics of the municipality, and correlated with growth.

To account for the fact that municipalities may diversify their agriculture, we introduce a share of production w_{ij} for each municipality-crop. As will become clearer when we address the econometric identification issues, we use historical shares of production from 1996 to 2000, but project these shares onto the potential yields using a fractional multinomial logit model, to obtain $\hat{w}_{ij}$ as follows

$$\hat{w}_{ij} = \mathbb{E}\left[\, w_{ij} \mid \mathbf{A}_i \,\right] = \frac{\exp\left(\mathbf{A}_i'\beta_j\right)}{1 + \sum_{k=1}^{J-1} \exp\left(\mathbf{A}_i'\beta_j\right)} \tag{2.3}$$

We are then able to calculate a quantity akin to a potential production in a diversified agriculture, in tonnes, denoted by Q_{ij}:

$$Q_{ij} = \hat{w}_{ij} \cdot A_{ij} \tag{2.4}$$

For each crop, we are able to calculate a share q_{ij}, relative to the other commodities, within each municipality i. We do so in terms of economic value, for which we use a measure of the value of each crop from historical prices. Let $\bar{p}_j$ represent the historical average price for each crop from 1995 to 2000. We get

$$q_{ij} = \frac{Q_{ij}\bar{p}_j}{\sum_j Q_{ij}\bar{p}_j} \tag{2.5}$$

The shares from equation (2.5) add up to one by construction, so they are then scaled separately for each i to form our actual shares s_{ij}, which add up to the share of agriculture in that municipality's GDP, $\bar{a}_i$, as below

$$s_{ij} = \bar{a}_i q_{ij} \tag{2.6}$$

$$\text{with} \quad \bar{a}_i = \frac{Y_{i,2001}^{\text{Agro}}}{Y_{i,2001}} \tag{2.7}$$

Chapter 2.3, "Data", shows the distribution of those shares.

2.2.2 Econometric Identification

Our specification is identified in the terms proposed by (Adão et al., 2019), that is, for $S := (s_{ij})_{i,j=1}^{N,J}$ and $\epsilon := (\varepsilon_i)_{i=1}^{N}$, we assume that we have mean independence of the price shocks. In a simplified way, this means

$$\mathbb{E}\left[\, G_j \mid S, \epsilon \,\right] = 0$$

This is a sufficient condition[1], hence one that guarantees identification. This *econometric* assumption is backed by two economic assumptions. First, that I can approximate the commodity prices by stationary time series in the roughly 20 years period considered. Therefore, comparing differences of average prices of any two periods would yield, on average, a expected value of zero. Second, this is true regardless of the ensemble of exposures of Brazilian municipalities, S, and of the counterfactual outcomes in the absence of the commodity boom, ε.

We have also taken measures to make the exposures to the commodity prices, or shares, as exogenous as possible. If they are correlated with unobserved regional characteristics, this could bias our results (remember the instrument is built by combining

[1] Adão et al. (2019) introduce other technical conditions for consistency and inference of which some are necessary, but I focus on the mean independence of the shocks for those carry the most meaningful economic assumptions.

the shocks G_j with the shares S). Those measures included using potential yields looking only at climatic characteristics, not terrain, and projecting the historical production shares on them (the $\hat{w}_{ij}$).

Since I advocate for the exogeneity of an instrument, it helps to think about the exclusion restriction. For this, I assume that the commodity prices abroad don't cause bank branch opening other than through greater demand in the affected municipalities. This seems reasonable as commercial banks in Brazil don't have large exposures to market risk, and this risk is mostly concentrated in interest and currency products.

2.2.3 Inference

As highlighted in the recent literature (see Adão et al., 2019; Borusyak et al., 2021), inference in shift share designs must be conducted with care. In those settings, the treatment variable is correlated across units according to the share structure. In other words, municipalities with similar shares are affected similarly by our identifying variation[2]. We can usually expect that the unobserved shocks are also correlated in a similar manner, creating inconsistency in the traditional Eicker-Huber-White Heteroskedasticity-robust standard errors (White, 1980). As this correlation happens regardless of geographical distance, cluster-robust estimators and other spatial-correlation-robust estimators (such as Conley, 1999), do not solve the issue. This usually leads to over-rejection by the inference procedure.

To overcome this, we adopt Shift-share consistent standard errors from Adão et al. (2019). These are valid only asymptotically[3], so we complement our inference procedure with the randomization techniques suggested by Alvarez et al. (2022) and exposed in Chapter 1 of this Master Thesis.

2.3 Data

Our data from potential yields A_{ij} comes from FAO-GAEZ. They express, for each crop in a square grid cell of 9km of width, the maximum yield taking into account only climatic conditions (temperature, rain, and humidity), therefore assuming the best possi-

[2] I.e., variation of GDP growth that is explained by the exposure to the exogenous commodity shock

[3] And the effective sample size is a function of the number of shocks, 16 agricultural products in this setting, and not on our much higher cross-section size, with 5570 municipalities.

ble use of the land (Fischer et al., 2021). These grids are aggregated for each municipality, based on territorial limits from the Brazilian Bureau of Statistics (IBGE).

To calculate the share of production for each municipality (the $\hat{w}_{ij}$ in eq. (2.3)), we use information on crops from 1996 to 2000 taken from the surveys *Pesquisa Pecuária Municipal* and *Pesquisa Agrícola Municipal* conducted by the IBGE, as done in Da Mata and Dotta (2021). They included livestock, permanent and temporary crops.

We use the same 15 crops as Da Mata and Dotta (2021): banana, barley, citrus, cocoa, coffee, groundnut, maize, rice, rubber, sorghum, soybeans, sugarcane, tea, tobacco, and wheat. In addition, we use their calculation of potential yield of bovines from the total yield of grass, using the amount of dry grass needed to raise one bovine.

For the share of agricultural GDP we use data from the national accounts from IBGE, for the year 2001.

In addition, we set the commodity exposure index to 2010 constant (real) BRL values using international commodity prices in U.S. Dollars from the World Bank (The Pink Sheet), Brazil's consumer price index (IPCA index), and exchange rate data from Central Bank of Brazil's SGS.

Data for population comes from the 2010 population census, and bank branches are taken from the Monthly Bank Statistics by Municipality (ESTBAN) from the Central Bank of Brazil. Our figure for branches excludes banking correspondents. These are mainly used for transactional financial services such as paying bills and depositing and withdrawing money. Although they are an important vector for financial inclusion, we focus in physical branches for two reasons. First, there is no data on correspondents going back in time before December 2007. Second, we believe our research question is relevant nevertheless because, in the period of interest, physical branches where the main channel for initiating credit operations: table 2.1 shows that, in 2011, almost 4 of every 5 credit operation was done in a physical branch. So, opening a physical branch means greater financial possibilities, even if that municipality was already serviced by a correspondent.

2.3.1 Summary measures on the commodity exposure

Figures 4, 5 and 6, present a summary on the commodity exposure and shocks. They summarise a few points:

- There is substantial heterogeneity in the share of GDP from agriculture for the municipalities in our sample.
- Even considering the non-negligible average annual inflation of 6.4% from 2002 to 2013, there is still significant real commodity price variation in the sample
- All 16 agricultural products have participation in the composition of the instrument (within one order of magnitude)

Figure 4 – Proportion of Agricultures in GDP, all Municipalities

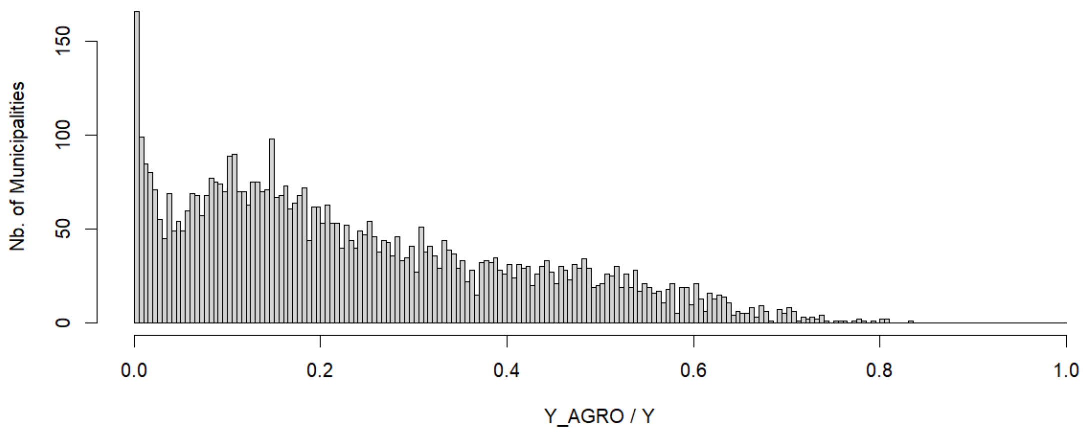


Figure 5 – Real Variation in Average Commodity Prices (ex-IPCA): 2002-2013 vs. 1995-2001

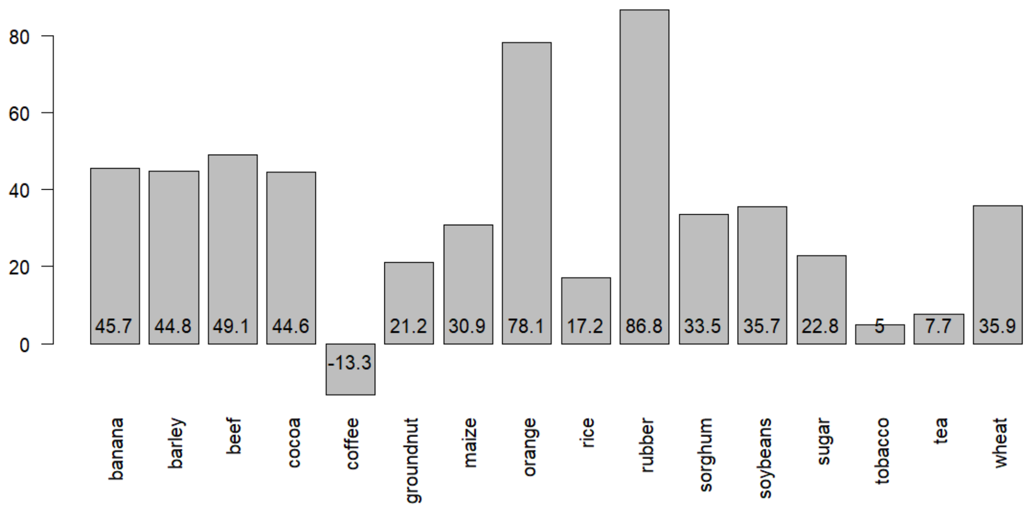
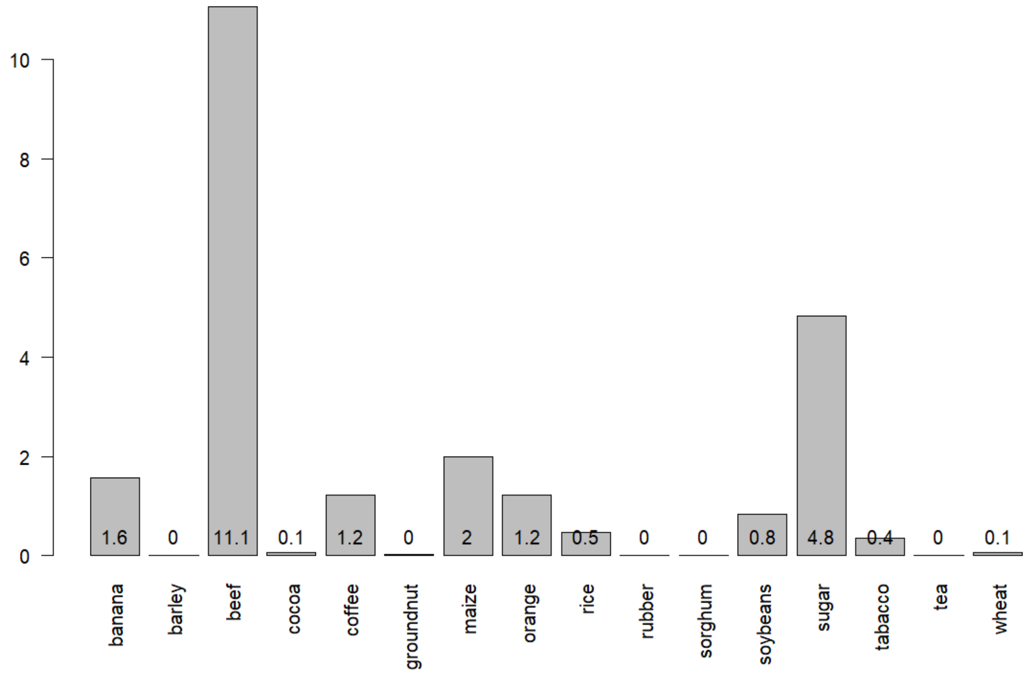


Figure 6 – Average Share for each crop (%)



2.4 Results

For the regressions, I have used all municipalities for which I had complete data, which amount to 5560 from the current 5570. To summarise, I use 1991-2001 vs 2002-2013 average commodity price changes to instrument the 2002 to 2014 real variation in GDP, with the of goal inquiring about its effect on the variation in the number of branches per 10 thousand population between 2002 and 2015. Our outcome is measured relative to the population from the 2010 census, and the number of bank branches in 2002 and 2015 are the annual averages of the monthly figures.

2.4.1 First stage

The first stage equation is in equation (2.2), reproduced here for simplicity:

$$\Delta \ln Y_i = \tilde{\alpha} + \delta Z_i + \tilde{\gamma} S_i + \zeta_i \tag{2.2}$$

A positive coefficient is expected, as municipalities more exposed to the commodities demand shock should experience increased economic growth. This is indeed what we get in Table 2. Using Shift Share consistent standard errors of Adão et al. (2019) in lieu of standard robust estimator still gives significant results.

Randomization inference results are presented in Table 3 for the two statistics proposed in Chapter 2, namely T_1, using the null-imposed residuals[4], and T_2, using the usual residuals. I use three alternative distributions for the simulations of the commodity shocks:

- *Perm.* stands for permutations of the original 16 shocks
- $\mathcal{N}(0,1)$ stands for drawing standard normal i.i.d. shocks
- $\mathcal{N}(0,\Sigma)$ stands for drawing joint normal shocks with common variance 1 and covariance 0.7 .

The distribution with correlated shocks tries to address worries about different crops having correlated prices in the period. A discussion is done in section 2.4.3.

Table 2 – First stage results with asymptotic inference

	(1)	(2)
δ	3.995***	3.995***
	(0.393)	(0.713)
Num.Obs.	5560	5560
R2	0.042	0.042
Std.Errors	HC1	AKM

Notes: Inference for the first stage using asymptotic methods and two different variance estimators: Eicker-Huber-White with d.o.f. adjustment (HC1, column 1), and Shift-Share Consistent (AKM, column 2).
Guide: + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Results from RI in Table 3 point to an overall non-significant first stage coefficient, δ, with p-values between 20% and 30%. If we believe our randomization distributions are informative of the true data generating process, these results undermine the whole IV strategy.

[4] We use a standard null of $H_0 : \delta = 0$

Table 3 – p-values for different inference methods on the first stage

		Perm.	$\mathcal{N}(0,1)$	$\mathcal{N}(0,\Sigma)$	HC1	AKM
δ	T_1 statistic	0.286	0.279	0.276	0.000	0.000
	T_2 statistic	0.204	0.204	0.202		

Notes: Inference for the first stage using randomization inference with statistics proposed in Chapter 1 in separate rows. Columns: *Perm.* uses permutations of the original 16 shocks; $\mathcal{N}(0,1)$ uses standard normal i.i.d. shocks; $\mathcal{N}(0,\Sigma)$ uses joint normal shocks with common variance 1 and covariance 0.7. For comparison, columns *HC1* and *AKM* present asymptotic p-values from estimators in table 2.

2.4.2 Second stage

The structural and first stage equations are as in eqs. (2.1) and (2.2), and reproduced below:

$$\Delta b_i = \alpha + \beta \Delta \ln Y_i + \gamma S_i + \varepsilon_i \tag{2.1}$$

$$\Delta \ln Y_i = \tilde{\alpha} + \delta Z_i + \tilde{\gamma} S_i + \zeta_i \tag{2.2}$$

Results using asymptotic inference are in Table 4, while those using randomization inference are in Table 5. This time, adding a consistent inference procedure in column (2) of Table 4 robs all significance from the main coefficient. Column (3) presents a naïve OLS estimate with robust standard errors, for comparison.

While asymptotic inference methods show a significant IV coefficient (at 5%), when we perform randomization inference, we see a total loss of significance, with p-values going from 4% to the vicinity of 40%.

2.4.3 Some comments on the randomization inference

As the procedure of Alvarez et al. (2022) controls for scale, and the specification in eqs. (2.1) and (2.2) controls for location of the shocks, using a correlated joint normal distribution for simulating the shocks is equivalent to using a standard normal i.i.d. shock. The result with the correlated shocks has been kept to show that a simpler randomization inference already captures worries about correlated prices.

Table 4 – IV and naïve OLS results with asymptotic inference

	(1)	(2)	(3)
β	1.855***	1.855*	0.072**
	(0.291)	0.908	(0.025)
Num.Obs.	5560	5560	5560
Std.Errors	HC1	AKM	HC1

Notes: Inference for the IV estimate using asymptotic methods and two different variance estimators: Eicker-Huber-White with d.o.f. adjustment (HC1, column 1), and Shift-Share Consistent (AKM, column 2). Column 3 shows a naïve OLS estimate of β, along with robust standard-errors.
Guide: + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5 – p-values for different inference methods on IV estimate

		Perm.	$\mathcal{N}(0,1)$	$\mathcal{N}(0,\Sigma)$	HC1	AKM
β	T_1 statistic	0.389	0.389	0.388	0.000	0.041

Notes: Inference for the first stage using randomization inference with statistics proposed in Chapter 1. *Perm.* uses permutations of the original 16 shocks; $\mathcal{N}(0,1)$ uses standard normal i.i.d. shocks; $\mathcal{N}(0,\Sigma)$ uses joint normal shocks with common variance 1 and covariance 0.7. For comparison, columns *HC1* and *AKM* present asymptotic p-values from estimators in table 4.

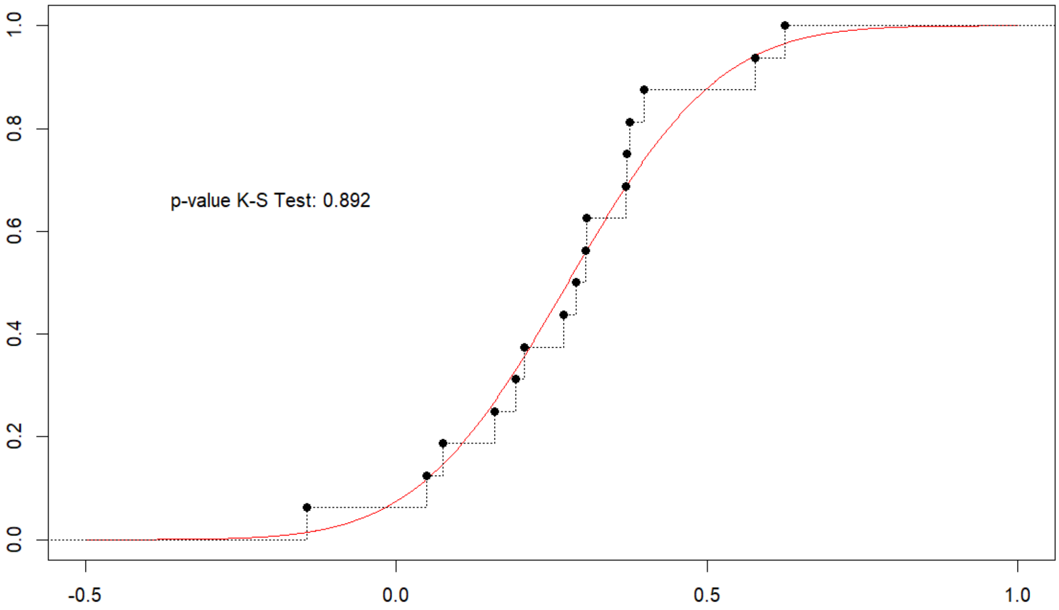
Also, notice that the permutation of the shocks has given very similar results to the standard normal i.i.d. shocks. Although a very small sample, the original shocks look a lot like a draw from a normal distribution. Power issues aside, a standard Kolmogorov-Smirnov test comparing with a normal distribution with this sample's mean and variance cannot reject its normality. This is illustrated in the plot below:

2.5 Discussion & Conclusion

I have tried to assess whether the bank branch expansion in Brazil during the commodity supercycle responded to economic activity. Although the point estimate shows an average in increase of 1.3 new bank branches for the median municipality in the period[5],

[5] The coefficient is 1.855, the median municipality had about 10,940 population in 2010, and the median growth was 89%, which gives 0.64 in *ln* terms. Multiplication gives 1.292.

Figure 7 – E.D.F. of shocks vs. Normal C.D.F. with same mean and variance



we cannot confidently separate this from noise.

In particular, we have seen an example where asymptotically consistent inference methods point towards a significant estimate, whereas randomization inference indicate an irrelevant instrument and an insignificant causal estimate.

I conclude that further research is needed to arrive at a more precise estimate. It is worth noting that the next steps should also look at different measures of income and activity, as agricultural GDP growth might be a result of big farms whose financing is made elsewhere.

Bibliography

Adão, R., M. Kolesar, and E. Morales (2019, 08). Shift-Share Designs: Theory and Inference. *The Quarterly Journal of Economics 134*(4), 1949–2010.

Alvarez, L., B. Ferman, and R. Oliveira (2022). Randomization inference tests for shift-share designs. Available at ArXiv.

Assunção, J. (2013). Eliminating entry barriers for the provision of banking services: Evidence from 'banking correspondents' in brazil. *Journal of Banking & Finance 37*(8), 2806–2811.

Autor, D. H., D. Dorn, and G. H. Hanson (2013, October). The china syndrome: Local labor market effects of import competition in the United States. *American Economic Review 103*(6), 2121–68.

Bartik, T. J. (1991). *Who Benefits from State and Local Economic Development Policies?* W.E. Upjohn Institute.

Berger, R. L. and D. D. Boos (1994). P values maximized over a confidence set for the nuisance parameter. *Journal of the American Statistical Association 89*(427), 1012–1016.

Billingsley, P. (1995). *Probability and Measure 3rd Edition.* Wiley series in probability and mathematical statistics. Wiley-Interscience.

Blanchard, O. and L. Katz (1992). Regional evolutions. *Brookings Papers on Economic Activity 23*(1), 1–76.

Borusyak, K. and P. Hull (2020, September). Non-random exposure to exogenous shocks: Theory and applications. Working Paper 27845, National Bureau of Economic Research.

Borusyak, K., P. Hull, and X. Jaravel (2021, 06). Quasi-experimental shift-share research designs. *The Review of Economic Studies 89*(1), 181–213.

Bugni, F. A., I. A. Canay, and A. M. Shaikh (2018). Inference under covariate-adaptive randomization. *Journal of the American Statistical Association 113*(524), 1784–1796. PMID: 30906087.

Card, D. (2001). Immigrant inflows, native outflows, and the local labor market impacts of higher immigration. *Journal of Labor Economics 19*(1), 22–64.

Chung, E. and J. P. Romano (2013). Exact and asymptotically robust permutation tests. *The Annals of Statistics 41*(2), 484–507.

Conley, T. G. (1999). GMM estimation with cross-sectional dependence. *Journal of Econometrics 92*(1), 1–45.

Célerier, C. and A. Matray (2019, 04). Bank-Branch Supply, Financial Inclusion, and Wealth Accumulation. *The Review of Financial Studies 32*(12), 4767–4809.

Da Mata, D. and M. Dotta (2021, August). Commodity booms and the environment. Available at SSRN.

Demirgüç-Kunt, A., L. Klapper, and D. Singer (April 2017). Financial inclusion and inclusive growth: A review of recent empirical evidence. Policy research working paper no. 8040, World Bank Group.

Demirgüç-Kunt, A., L. Klapper, D. Singer, and S. Ansar (2022). The global findex database 2021: Financial inclusion, digital payments, and resilience in the age of covid-19. Technical report, Washington, DC: World Bank.

Durrett, R. (2019, May). *Probability: Theory and Examples 5th Edition*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Ergungor, O. E. (October 2010). Bank branch presence and access to credit in low- to moderate-income neighborhoods. *Journal of Money, Credit and Banking 42*(7), 1321–1349.

Ferman, B. (2019). Assessing inference methods.

Ferman, B. (2021). Matching estimators with few treated and many control observations. *Journal of Econometrics 225*(2), 295–307.

Fischer, G., F. Nachtergaele, H. van Velthuizen, F. Chiozza, G. Franceschini, M. Henry, D. Muchoney, and S. Tramberend (2021). Global agro-ecological zones v4 - model documentation. Model documentation, Food and Agriculture Organization of the United Nations and International Institute for Applied Systems Analysis.

Goldsmith-Pinkham, P., I. Sorkin, and H. Swift (2020, August). Bartik instruments: What, when, why, and how. *American Economic Review 110*(8), 2586–2624.

Janssen, A. (1997). Studentized permutation tests for non-i.i.d. hypotheses and the generalized Behrens-Fisher problem. *Statistics & Probability Letters 36*(1), 9–21.

Lehmann, E. L. and J. P. Romano (2005). *Testing Statistical Hypotheses* (3 ed.). Springer Texts in Statistics. New York, NY: Springer New York, NY.

Roth, J. and P. H. Sant'Anna (2022, 08). Efficient Estimation for Staggered Rollout Designs. Draft.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica 48*(4), 817–838.

Wu, J. and P. Ding (2021). Randomization tests for weak null hypotheses in randomized experiments. *Journal of the American Statistical Association 116*(536), 1898–1913.

Appendix

APPENDIX A – Proofs for the Theoretical Results

A.1 Finite-sample results

A.1.1 Proof: Proposition 1

Proof. Under Assumption 3, both $\hat{T}$ and H_b depend on $(\mathbf{Y}, \mathbf{X})$ solely through $\mathbf{e}_b$, so

$$H_b\left(c \mid \mathbf{S}, \mathbf{X}, \mathbf{Y}\right) = \int \mathbf{1}_{\{T(g, \mathbf{S}, \mathbf{e}_b)\ \leq\ c\}}\ \mathbb{H}\left(dg \mid \mathbf{S}, \mathbf{e}_b\right)$$

Under H_0 and homogeneous effects, we have that $\mathbf{e}_b = \boldsymbol{\epsilon}$. Further substituting $\mathbb{H}$ with $\mathbb{P}$ leads to

$$H_b\left(c \mid \mathbf{S}, \mathbf{X}, \mathbf{Y}\right) = \int \mathbf{1}_{\{T(g, \mathbf{S}, \boldsymbol{\epsilon})\ \leq\ c\}}\ \mathbb{P}\left(dg \mid \mathbf{S}, \boldsymbol{\epsilon}\right)$$

Which is, by definition, equal to $\mathbb{P}\left(\hat{T} \leq c \mid \mathbf{S}, \boldsymbol{\epsilon}\right)$ □

A.1.2 Proof: Corollary 1.1

Proof. The quantile function for a variable with p.d.f. F is given by

$$Q_F(u) := \inf\{z \in \mathbb{R} : F(z) \geq u\}$$

Rejection of H_0 occurs if, and only if, $T(g, \mathbf{S}, \boldsymbol{\epsilon}) > Q_{H_b}(1-\alpha)$. Therefore, the probability of rejection is

$$\Pr\left(\text{Reject } H_0\right) = \Pr\left(T(g, \mathbf{S}, \boldsymbol{\epsilon}) > Q_{H_b}(1-\alpha)\right) = 1 - F_{\mathbb{P}}\left(Q_{H_b}(1-\alpha)\right)$$

Using that $F\left(Q_F(u)\right) \geq u$, and that, from Proposition 1, $Q_{H_b} = Q_{\mathbb{P}}$, we have

$$\Pr\left(\text{Reject } H_0\right) \leq 1 - (1-\alpha) = \alpha$$

□

A.1.3 Proof: Corollary 1.2

Proof. Rejection of H_0 occurs if, and only if, $T(g, \mathbf{S}, \boldsymbol{\epsilon}) > Q_{\hat{H}_b}(k_\alpha)$. Since the quantile is a lower bound for the set $\hat{H}_b \geq k_\alpha$, having a $\hat{T}$ that exceeds it this implies that

$$\hat{H}_b\left(\hat{T} \mid \mathbf{S}, \mathbf{X}, \mathbf{Y}\right) \geq k_\alpha = 1 + \frac{2}{L+1} - \frac{\lfloor \alpha(L+1) \rfloor}{L+1}$$

$$\Longrightarrow \sum_{\ell=1}^{L} \mathbf{1}_{\left\{T(g_\ell^*, \mathbf{S}, \mathbf{e}_b) \leq \hat{T}\right\}} \geq L + 2 - \lfloor \alpha(L+1) \rfloor$$

meaning that $\hat{T} = T(G, \mathbf{S}, \mathbf{e}_b)$ is strictly greater than the m-th order statistic (i.e., m-th smallest draw) of $T(g_\ell^*, \mathbf{S}, \mathbf{e}_b)$, with $m = L + 1 - \lfloor \alpha(L+1) \rfloor$. [1]

Denoting by $T^*_{(m)}$ the m-th largest order statistic from the L independent draws $\left(T(g_\ell^*, \mathbf{S}, \mathbf{e}_b)\right)_{\ell=1}^{L}$, we have that

$$\Pr\left(\text{Reject } H_0 \mid \mathcal{A}\right) \leq \Pr\left(T^*_{(m)} < \hat{T} \mid \mathcal{A}\right) = 1 - \Pr\left(\hat{T} \leq T^*_{(m)} \mid \mathcal{A}\right)$$

with $\mathcal{A} = (\mathbf{S}, \boldsymbol{\epsilon})$ as usual. Since the L draws $G^* := (g_\ell^*)_{\ell=1}^{L}$ are independent from $\mathbf{S}$ and $\boldsymbol{\epsilon}$, we can use the Law of Iterated Expectations and write

$$\begin{aligned} \Pr\left(\hat{T} \leq T^*_{(m)} \mid \mathcal{A}\right) &= \mathbb{E}\left[\Pr\left(\hat{T} \leq T^*_{(m)} \mid \mathcal{A}, G^*\right) \,\middle|\, \mathcal{A}\right] \\ &= \mathbb{E}\left[H_b\left(T^*_{(m)} \mid \mathbf{S}, \mathbf{X}, \mathbf{Y}\right) \,\middle|\, \mathcal{A}\right] \end{aligned}$$

Since $G^* \mid \mathcal{A}$ are i.i.d. by assumption, we have that $U_{(m)} := H_b\left(T^*_{(m)} \mid \mathbf{S}, \mathbf{X}, \mathbf{Y}\right)$ is distributed as the m-th order statistic from an i.i.d. sample of size L from a uniform distribution. This gives us

$$\mathbb{E}\left[U_{(m)} \,\middle|\, \mathcal{A}\right] = \frac{m}{L+1} = \frac{L+1-\lfloor \alpha(L+1) \rfloor}{L+1} = 1 - \frac{\lfloor \alpha(L+1) \rfloor}{L+1} \geq 1 - \alpha$$

Therefore,

$$\Pr\left(\text{Reject } H_0 \mid \mathcal{A}\right) \leq 1 - \mathbb{E}\left[U_{(m)} \,\middle|\, \mathcal{A}\right] \leq \alpha$$

□

[1] It is at least as large as the $(m+1)$-th smallest draw, but this does not help because we would get inequalities of the type $\ldots \leq \ldots \geq \ldots$

A.2 Asymptotic results

A.2.1 Proof: Proposition 2

Proof. Notice that $T_1(g, \boldsymbol{S}, \mathbf{e}_b)$ from eq. (1.12) can be written

$$T_1(g, \boldsymbol{S}, \mathbf{e}_b) = \frac{\sum_{j=1}^{J}\sum_{i=1}^{N} s_{ij} e_{b,i} g_j}{\sqrt{\sum_{j=1}^{J}\left(\sum_{i=1}^{N} e_{b,i} s_{ij}\right)^2 g_j^2}} = \frac{N_J/\sqrt{\nu_J}}{\sqrt{D_J^2/\nu_J}}$$

We begin by evaluating the consistency of the denominator. For conciseness, I denote $\mathbb{E}\left[\,\cdot \mid \mathcal{F}\,\right]$ by $\mathcal{E}\left[\,\cdot\,\right]$. Since the s_{ij} and $e_{b,i} = (\beta_i - b)X_i + \varepsilon_i$ are fixed when conditioning on $\mathcal{F} = \{B_J, \mathbf{S}_J, \boldsymbol{\epsilon}_J, g_J, \mathbf{X}_J\}$, we have that

$$\mathcal{E}\left(D_J^2\right) = \sum_{j=1}^{J}\left(\sum_{i=1}^{N} e_{b,i} s_{ij}\right)^2 \mathcal{E}\left[\,g_j^2\,\right]$$

$$\Longrightarrow D_j^2 - \mathcal{E}\left(D_J^2\right) = \sum_{j=1}^{J}\left(\sum_{i=1}^{N} e_{b,i} s_{ij}\right)^2 \left(g_j^2 - \mathcal{E}\left[\,g_j^2\,\right]\right)$$

Continuing, from the mutual independence (i.e., across j) of the $g_j \mid \mathcal{F}$, the cross terms $\left(g_j^2 - \mathcal{E}\left[\,g_j^2\,\right]\right)(g_k^2 - \mathcal{E}\left[\,g_k^2\,\right])$ with $j \neq k$ vanish, and we write

$$\begin{aligned}
\mathcal{E}\left[\left|\,D_J^2 - \mathcal{E}\left(D_J^2\right)\,\right|^2\right] &= \mathcal{E}\left[\left(\sum_{j=1}^{J}\left(\sum_{i=1}^{N} e_{b,i} s_{ij}\right)^2 \left(g_j^2 - \mathcal{E}\left[\,g_j^2\,\right]\right)\right)^2\right] \\
&= \sum_{j=1}^{J}\left(\sum_{i=1}^{N} e_{b,i} s_{ij}\right)^4 \mathcal{E}\left[\left(g_j^2 - \mathcal{E}\left[\,g_j^2\,\right]\right)^2\right] \\
&= \sum_{j=1}^{J}\left(\sum_{i=1}^{N} e_{b,i} s_{ij}\right)^4 \left(\mathcal{E}\left[\,g_j^4\,\right] - \mathcal{E}\left[\,g_j^2\,\right]^2\right) \\
&\leq \sum_{j=1}^{J}\left(\sum_{i=1}^{N} e_{b,i} s_{ij}\right)^4 \mathcal{E}\left[\,g_j^4\,\right]
\end{aligned}$$

Using this result and Assumption (iii), we can write that

$$\mathcal{E}\left[\left|\frac{D_J^2}{\nu_J} - \mathcal{E}\left(\frac{D_J^2}{\nu_J}\right)\right|^2\right] = \mathbb{E}\left[\left|\frac{D_J^2}{\nu_J} - \mathbb{E}\left[\frac{D_J^2}{\nu_J}\,\middle|\,\mathcal{F}\right]\right|^2 \,\middle|\, \mathcal{F}\right] \xrightarrow{p} 0$$

We can then use Markov's Inequality (conditionally) to establish that

$$\Pr\left(\left|\frac{D_J^2}{\nu_J} - \mathbb{E}\left[\frac{D_J^2}{\nu_J}\,\middle|\,\mathcal{F}\right]\right| \geq \varepsilon \,\middle|\, \mathcal{F}\right) \xrightarrow{p} 0$$

Using the Law of Iterated Expectations, we have that

$$\Pr\left(\left|\frac{D_J^2}{\nu_J} - \mathbb{E}\left[\frac{D_J^2}{\nu_J} \,\middle|\, \mathcal{F}\right]\right| \geq \varepsilon\right) = \mathbb{E}\left[\Pr\left(\left|\frac{D_J^2}{\nu_J} - \mathbb{E}\left[\frac{D_J^2}{\nu_J} \,\middle|\, \mathcal{F}\right]\right| \geq \varepsilon \,\middle|\, \mathcal{F}\right)\right]$$

Finally, with the Bounded Convergence Theorem (because a probability is bounded by 1), we can say that[2]

$$\Pr\left(\left|\frac{D_J^2}{\nu_J} - \mathbb{E}\left[\frac{D_J^2}{\nu_J} \,\middle|\, \mathcal{F}\right]\right| \geq \varepsilon\right) \to 0$$

Thus establishing consistency of the denominator for its (conditional) mean for every value in $\mathcal{F}$. Using assumption (ii), we conclude that

$$\frac{D_J^2}{\nu_J} = \frac{1}{\nu_J} \sum_{j=1}^{J} \left(\sum_{i=1}^{N} e_{b,i} s_{ij}\right)^2 g_j^2 \xrightarrow{p} \sigma^2 > 0 \tag{A.1}$$

For the numerator, I define $\lambda_j := \nu_J^{-\frac{1}{2}} g_j \sum_{i=1}^{N} s_{ij} e_{b,i}$, such that we can write $N_J/\sqrt{\nu_J} = \sum_{j=1}^{J} \lambda_j$. The idea is to use a triangular array CLT on these λ_j. (Lindeberg-Feller, as in Durrett, 2019, Theorem 3.4.10). Passing to a subsequence $(J(k))_{k\in\mathbb{N}}$ if needed, we can restate our assumptions (i)-(iii) as almost sure convergence (refer to Billingsley, 1995, Theorem 20.5, statement (ii)).[3]

From Assumption (i), we can write

$$\sum_{j=1}^{J} \left(\lambda_j - \mathcal{E}\left[\lambda_j\right]\right) = \sum_{j=1}^{J} \lambda_j + o_p(1) \qquad \text{a.s.}$$

and the left-hand side has obviously zero-mean for every j (name it condition A.2.1.A). Now, from result in eq. (A.1), we have that

$$\sum_{j=1}^{J} \mathcal{E}\left[\lambda_j^2\right] \xrightarrow{a.s.} \sigma^2 > 0 \qquad \text{(condition A.2.1.B)}$$

Finally, for the Lindeberg condition (name it condition A.2.1.C), we can use assumption (iii) again, writing, for an arbitrary $\delta > 0$,

$$\sum_{j=1}^{J} \mathcal{E}\left[\lambda_j^2 \cdot \mathbf{1}\{\lambda_j > \delta\}\right] \leq \frac{1}{\delta^2} \sum_{j=1}^{J} \mathcal{E}\left[\lambda_j^4\right] =$$

$$= \frac{1}{\delta^2} \frac{1}{\nu^2} \sum_{j=1}^{J} \left(\sum_{i=1}^{N} s_{ij} e_{b,i}\right)^4 \mathbb{E}\left[(G_j^*)^4 \mid \mathcal{F}\right] \xrightarrow{a.s.} 0 \quad \text{(cond. A.2.1.C)}$$

[2] Notice this statement is non-probabilistic as a result of the BDD, meaning that the r.v. $\left| D_J^2/\nu_J - \mathbb{E}\left[D_J^2/\nu_J \mid \mathcal{F}\right] \right| \xrightarrow{p} 0$ unconditionally, i.e., a.s. for every value of $\mathcal{F}$.

[3] I cite: $X(n) \xrightarrow{p} X_0$ if, and only if, for every subsequence $N(k)$, there is a subsequence $J(i)$ such that $X_{(N\circ J)(i)} \xrightarrow{a.s.} X_0$

Using conditions A.2.1.(A,B,C), we are able to apply the CLT on $\sum_{j=1}^{J} \lambda_j = N_J/\sqrt{\nu_J}$ and say that

$$\forall\, c \in \mathbb{R} \quad \Pr\left(\frac{1}{\sigma} \sum_{j=1}^{J} \lambda_j \leq c \,\middle|\, \mathcal{F} \right) \xrightarrow{a.s.} \Phi(c) \tag{A.2}$$

Since we had a subsequence (c.f. footnote 3), we can state that, unconditionally

$$\forall\, c \in \mathbb{R} \quad \Pr\left(T_1(G^*, \boldsymbol{S}, \mathbf{e}_b) \leq c \right) \xrightarrow{p} \Phi(c) \tag{A.3}$$

□

A.2.2 Proof: Proposition 3

Proof. Notice that $T_2(G, \boldsymbol{S}, \mathbf{e}_b)$ from eq. (1.13) can be written

$$T_2(G, \boldsymbol{S}, \mathbf{e}_b) = \frac{\sum_{j=1}^{J} \sum_{i=1}^{N} s_{ij} e_{b,i} g_j}{\sqrt{\sum_{j=1}^{J} \left(\sum_{i=1}^{N} \left((b - \hat{\beta})(S_i' G) + e_{b,i} \right) s_{ij} \right)^2 g_j^2}} = \frac{N_J/\sqrt{\nu_J}}{\sqrt{B_J^2/\nu_J}}$$

Since we are under all the assumptions of Proposition 2, we use the result from eq. (A.2) in the previous proof directly[4]:

$$\forall\, c \in \mathbb{R} \quad \Pr\left(\frac{1}{\sigma} \frac{N_J}{\sqrt{\nu_J}} \leq c \,\middle|\, \mathcal{F} \right) \xrightarrow{a.s.} \Phi(c)$$

Then, simple algebra reveals that the denominator can be written as

$$\frac{B_J^2}{\nu_J} = C_0 + C_1 + 2C_2$$

[4] Passing to a subsequence if needed, see footnote 3.

with

$$C_0 = \frac{1}{\nu_J}\sum_{j=1}^{J}\left(\sum_{i=1}^{N} e_{b,i}s_{ij}\right)^2 g_j^2$$

$$\begin{aligned} C_1 &= \frac{1}{\nu_J}\sum_{j=1}^{J}\left(\sum_{i=1}^{N}(b-\hat{\beta})(S_i'G)s_{ij}\right)^2 g_j^2 \\ &= \frac{(b-\hat{\beta})^2}{\nu_J}\sum_{j=1}^{J} g_j^2 \sum_{p=1}^{N}\sum_{q=1}^{N} Z_pZ_q s_{pj}s_{qj} \end{aligned}$$

$$\begin{aligned} C_2 &= \frac{1}{\nu_J}\sum_{j=1}^{J}\left(\sum_{i=1}^{N} s_{ij}(b-\hat{\beta})(S_i'G)\right)\left(\sum_{i=1}^{N} s_{ij}e_{b,i}\right) g_j^2 \\ &= \frac{(b-\hat{\beta})}{\nu_J}\sum_{j=1}^{J} g_j^2 \sum_{p=1}^{N}\sum_{q=1}^{N} s_{pj}s_{qj}Z_pe_{b,q} \end{aligned}$$

Looking at the term $\left(b-\hat{\beta}\right)$, we write

$$\hat{\beta} - b = \frac{\sum_{i=1}^{N} e_{b,i}X_i}{\sum_{j=1}^{J} X_i^2} = \frac{\nu_J}{N}\cdot\frac{\left(\sum_{i=1}^{N} e_{b,i}X_i\right)/\nu_J}{\left(\sum_{j=1}^{J} X_i^2\right)/N}$$

But, again from the proof of Proposition 2 in section A.2.1, and from $X_i = Z_i$, we know that

$$\frac{1}{\nu_J}\sum_{i=1}^{N} e_{b,i}X_i = \frac{1}{\nu_J}\sum_{j=1}^{J} g_j \sum_{i=1}^{N} e_{b,i}s_{ij} = \frac{N_J}{\nu_J} = O_p(1)$$

Using assumption (iv), $\frac{1}{N}\sum_{j=1}^{J} X_i^2 = O_p(1)$, so that $\frac{N}{\sqrt{\nu_J}}\left(b-\hat{\beta}\right) = O_p(1)$, meaning that

$$b-\hat{\beta} = O_p\left(\frac{\sqrt{\nu_J}}{N}\right) \quad \text{and} \quad \left(b-\hat{\beta}\right)^2 = O_p\left(\frac{\nu_J}{N^2}\right)$$

This means that, using assumptions (v) and (vi), we have that

$$\begin{aligned} C_1 &= O_p\left(\frac{\nu_J}{N^2}\right)\cdot o_p\left(\frac{N^2}{\nu_J}\right) = o_p(1) && \text{by ass. (vi)} \\ 2C_2 &= O_p\left(\frac{\sqrt{\nu_J}}{N}\right)\cdot o_p\left(\frac{N}{\sqrt{\nu_J}}\right) = o_p(1) && \text{by ass. (v)} \end{aligned}$$

meaning that

$$\frac{B_J^2}{\nu_J} = C_0 + o_p(1)$$

Finally, we use the result in eq. (A.1) from the previous proof to write that

$$C_0 = \frac{1}{\nu_J} \sum_{j=1}^{J} \left(\sum_{i=1}^{N} e_{b,i} s_{ij} \right)^2 g_j^2 \ \xrightarrow{p} \ \sigma^2 > 0$$

This allows us to conclude that

$$\forall\, c \in \mathbb{R} \quad \Pr\left(T_2(G^*, \boldsymbol{S}, \mathbf{e}_b) \leq c \right) \xrightarrow{p} \Phi(c)$$

□

APPENDIX B – Examples of simulation distributions

Notation here is as follows:

$$\begin{aligned} Y_{J,i} &= \beta_J X_{J,i} + (\beta_{J,i} - \beta_J) X_{J,i} + \varepsilon_{J,i} \\ &= \beta_J X_{J,i} + \eta_{J,i} + \varepsilon_{J,i} \end{aligned}$$

with the obvious definition

$$\eta_{J,i} := (\beta_{J,i} - \beta_J) X_{J,i}$$

B.1 In the context of Proposition 2

We verify the conditions of Proposition 2 in three examples.

B.1.1 Nonparametric bootstrap

In this case, $g^*_{j,J} \overset{iid}{\sim} \hat{F}_g$, where $\hat{F}_g$ is the empirical distribution of recentered shocks, i.e. $\hat{F}_g(c) = \frac{1}{J}\sum_{j=1}^J \mathbf{1}_{\{g_{j,J} - \bar{g}_J \leq c\}}$ and $\bar{g}_J = \frac{1}{J}\sum_{j=1}^J g_{j,J}$. In this case, $\mathbb{E}_*[g^*_{j,J}] = 0$, which ensures condition (i). As for the second condition, since $\mathbb{E}_*|g^*_{j,J}|^2 = \frac{1}{J}\sum_{j=1}^J |g_{j,J} - \bar{g}_J|^2$, requirement (ii) in the Proposition will be satisfied if $\frac{1}{J}\sum_{j=1}^J |g_{j,J} - \bar{g}_J|^2$ converges in probability to a positive nonrandom limit and $\sum_{j=1}^J \frac{1}{v_J}\left(\sum_{i=1}^N [\epsilon_{i,J} + \eta_{i,J} + (\beta_J - b_J)X_{i,J}]\boldsymbol{s}_{i,j,J}\right)^2$ converges in probability to a positive nonrandom limit. Finaly, requirement (iii) is satisfied if $\mathbb{E}_*|g^*_{j,J}|^4 = \frac{1}{J}\sum_{j=1}^J |g_{j,J} - \bar{g}_J|^4$ converges in probability and

$$\sum_{j=1}^J \frac{1}{v_J^2}\left(\sum_{i=1}^N [\epsilon_{i,J} + \eta_{i,J} + (\beta_J - b_J)X_{i,J}]\boldsymbol{s}_{i,j,J}\right)^4 \xrightarrow{p} 0$$

B.1.2 Normal distribution

Suppose $g_J^* \sim N(0, \mathbb{I}_{J\times J})$, independently from $\boldsymbol{X}, g_J$. In this case, $\mathbb{E}_*[g_{j,J}^*] = 0$, for $j = 1, \ldots, J$, which ensures condition (i). As for the second condition in the Proposition,

$$\sum_{j=1}^{J} \frac{1}{v_J}\Big(\sum_{i=1}^{N}[\epsilon_{i,J} + \eta_{i,J} + (\beta_J - b_J)X_{i,J}]\boldsymbol{s}_{i,j,J}\Big)^2 \mathbb{E}_* g_{j,J}^{*2} = \sum_{j=1}^{J} \frac{1}{v_J}\Big(\sum_{i=1}^{N}[\epsilon_{i,J} + \eta_{i,J} + (\beta_J - b_J)X_{i,J}]\boldsymbol{s}_{i,j,J}\Big)^2$$

which converges in probability if the latter term converges. As discussed in Appendix A of AKM, convergence in probability of $\sum_{j=1}^{J} \frac{1}{v_J}\Big(\sum_{i=1}^{N}[\epsilon_{i,J} + \eta_{i,J} + (\beta_J - b_J)X_{i,J}]\boldsymbol{s}_{i,j,J}\Big)^2$ to a positive limit requires the existence of at least one "non-negligible" shock in most units, where by non-negligible shock in a unit we mean its exposure weight is bounded away from zero. In an "extreme" case, where $N = J$ and each unit is affected by a single distinct shock with unit exposure, this term simplifies to $\frac{1}{N}\sum_{i=1}^{N}[\epsilon_{i,J}+\eta_{i,J}+(\beta_J-b_J)X_{i,J}]^2$, which is expected to converge to a positive limit under mild conditions. Finally, the third condition simplifies to

$$\sum_{j=1}^{J} \frac{1}{v_J^2}\left(\sum_{i=1}^{N}[\epsilon_{i,J} + \eta_{i,J} + (\beta_J - b_J)X_{i,J}]\boldsymbol{s}_{i,j,J}\right)^4 \mathbb{E}_* g_{j,J}^{*4} = 3\sum_{j=1}^{J} \frac{1}{v_J^2}\left(\sum_{i=1}^{N}[\epsilon_{i,J} + \eta_{i,J} + (\beta_J - b_J)X_{i,J}]\boldsymbol{s}_{i,j,J}\right)^4$$

which converges to zero if the latter term converges. In the single-shock-exposure setting, this term simplifies to $\frac{1}{N^2}\sum_{i=1}^{N}[\epsilon_{i,J} + \eta_{i,J} + (\beta_J - b_J)X_{i,J}]^4$, which converges in probability to zero under mild conditions.

B.1.3 Sign changes

In this case, $g_J^* = \pi^* \odot g_J$, where $\odot$ denotes entry-by-entry multiplication, and $\pi^* \sim \text{Uniform}(\{-1,1\}^J)$, independently from $(g_J, \boldsymbol{X})$. By construction, $\mathbb{E}_*[g_{j,J}^*] = 0$. As for the second condition,

$$\sum_{j=1}^{J} \frac{1}{v_J}\left(\sum_{i=1}^{N}[\epsilon_{i,J} + \eta_{i,J} + (\beta_J - b_J)X_{i,J}]\boldsymbol{s}_{i,j,J}\right)^2 \mathbb{E}_* g_{j,J}^{*2} = \sum_{j=1}^{J} \frac{1}{v_J}\left(\sum_{i=1}^{N}[\epsilon_{i,J} + \eta_{i,J} + (\beta_J - b_J)X_{i,J}]\boldsymbol{s}_{i,j,J}\right)^2 g_{j,J}^2$$

which converges in probability to a positive constant if:

(a) $\sum_{j=1}^{J} \frac{1}{v_J} \left(\sum_{i=1}^{N} [\epsilon_{i,J} + \eta_{i,J} + (\beta_J - b_J) X_{i,J}] \boldsymbol{s}_{i,j,J}\right)^2 \mathbb{E}[g_{j,J}^2]$ converges to a positive constant; and

(b) $\mathbb{E}\left|\sum_{j=1}^{J} \frac{1}{v_J} \left(\sum_{i=1}^{N} [\epsilon_{i,J} + \eta_{i,J} + (\beta_J - b_J) X_{i,J}] \boldsymbol{s}_{i,j,J}\right)^2 (g_{j,J}^2 - \mathbb{E}[g_{j,j}])^2\right|^2$ converges to zero

A condition like (a) is required for existing inference methods in shift-share designs to work (see the discussion surounding Assumption A.1. in AKM). Condition (b) is satisfied if shocks $g_{j,J}$ are independent, $\mathbb{E}|g_{j,J}|^4$ is uniformly bounded and

$$\sum_{j=1}^{J} \frac{1}{v_J^2} \left(\sum_{i=1}^{N} [\epsilon_{i,J} + \eta_{i,J} + (\beta_J - b_J) X_{i,J}] \boldsymbol{s}_{i,j,J}\right)^4 \xrightarrow{p} 0$$

Finally, condition (iii) in the Proposition is satisfied if $\mathbb{E}|g_{j,J}|^4$ is uniformly bounded and $\sum_{j=1}^{J} \frac{1}{v_J^2} \left(\sum_{i=1}^{N} [\epsilon_{i,J} + \eta_{i,J} + (\beta_J - b_J) X_{i,J}] \boldsymbol{s}_{i,j,J}\right)^4$ converges to zero.

B.2 In the context of Proposition 3

We now discuss Assumptions (iv)-(vi) of Proposition 3 in the context of the three examples in the previous section.

B.2.1 Verification of condition (iv) of Proposition 3

When the distribution of simulated shocks is standard normal,

$$\frac{1}{N} \sum_{i=1}^{N} \mathbb{E}[(\boldsymbol{s}_{i,J}' g_J^*)^2] = \frac{\sum_{i=1}^{N} \sum_{j=1}^{J} s_{i,j,J}^2}{N}$$

which we require to converge to a positive constant. Such condition is analogous to Assumption A1.(ii) in AKM. Moreover, we note that, in the single-exposure case, this quantity is exactly equal to one. To conclude that the denominator of the simulated shift-share regression estimator converges in probability to a positive constant, it is sufficient to require that $\mathrm{Var}\left(\frac{1}{N} \sum_{i=1}^{N} (\boldsymbol{s}_{i,J}' g_J^*)^2\right) = o(1)$. Observe that, in the single-exposure case, this variance is given by

$$\frac{3N + N(N-1)}{N^2} - 1$$

which converges to zero as $N \to \infty$. Similar arguments establish convergence of the denominator of the shift-share regression estimator in bootstrap and sign changes examples.

B.2.2 Verification of conditions (v-vi) of Proposition 3

Assumptions (v-vi) implicitly restrict moments of the simulated shocks and the relation between exposure weights and the rate of growth of N. Indeed, in the single-exposure case, requirement (v) subsumes to

$$\frac{2}{N^{3/2}}\sum_{i=1}^{N} g_{j,J}^{*3}(\epsilon_{j,J}+\eta_{j,J}+(\beta_J-b_J)X_{i,J})=o_p(1)$$

which is expected to hold under mild conditions in our three main examples. Similarly, in the single exposure case, condition (vi) subsumes to,

$$\frac{1}{N^2}\sum_{i=1}^{N} g_{j,J}^{*4}=o_p(1),$$

which is also expected to hold.