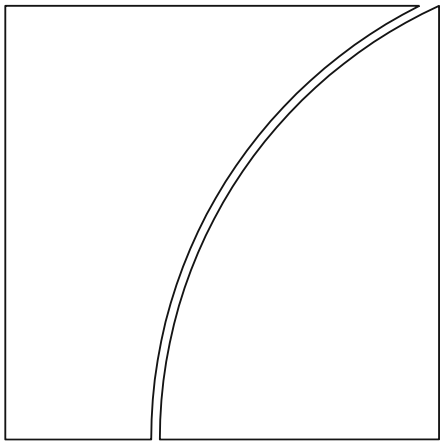


BIS Working Papers

No 1181

Synthetic controls with machine learning: application on the effect of labour deregulation on worker productivity in Brazil

by Douglas K. G. Araujo

Monetary and Economic Department

April 2024

JEL classification: B41, C32, C54, E24, J50, J83, O47

Keywords: causal inference, synthetic controls, machine learning, labour reforms, productivity

BIS Working Papers are written by members of the Monetary and Economic Department of the Bank for International Settlements, and from time to time by other economists, and are published by the Bank. The papers are on subjects of topical interest and are technical in character. The views expressed in them are those of their authors and not necessarily the views of the BIS.

This publication is available on the BIS website (www.bis.org).

ISSN 1020-0959 (print)
ISSN 1682-7678 (online)

Synthetic controls with machine learning: application on the effect of labour deregulation on worker productivity in Brazil*

Douglas K. G. Araujo[1]

[1]Bank for International Settlements, douglas.araujo@bis.org

Abstract

Synthetic control methods are a data-driven way to calculate counterfactuals from control individuals for the estimation of treatment effects in many settings of empirical importance. In canonical implementations, this weighting is linear and the key methodological steps of donor pool selection and covariate comparison between the treated entity and its synthetic control depend on some degree of subjective judgment. Thus current methods may not perform best in settings with large datasets or when the best synthetic control is obtained by a nonlinear combination of donor pool individuals. This paper proposes "machine controls", synthetic controls based on automated donor pool selection through clustering algorithms, supervised learning for flexible non-linear weighting of control entities and manifold learning to confirm numerically whether the synthetic control indeed resembles the target unit. The machine controls method is demonstrated with the effect of the 2017 labour deregulation on worker productivity in Brazil. Contrary to policymaker expectations at the time of enactment of the reform, there is no discernible effect on worker productivity. This result points to the deep challenges in increasing the level of productivity, and with it, economic welfare. JEL: B41, C32, C54, E24, J50, J83, O47. Keywords: causal inference, synthetic controls, machine learning, labour reforms, productivity.

1 Introduction

The synthetic control (SC) method (Abadie and Gardeazabal (2003), Abadie, Diamond, and Hainmueller (2010), Abadie, Diamond, and Hainmueller (2015))

*This work represents my opinion and not necessarily that of the Bank for International Settlements. I thank Ben Cohen, members of the BIS Informal Machine Learning Community and seminar participants at the Armerican University of Armenia, Bank for International Settlements, Bank of Canada, IÉSEG and International Monetary Fund for helpful comments and questions. An informal version of the empirical work is published as an example of the library gingado (Araujo (2023)). All errors are my own.

is widely used to enable the estimation of causal effects of interventions in only one or a few entities of a larger population. The key idea is to calculate a SC based on pre-intervention information from control individuals, which then serves as estimate for the treated unit's potential outcome under no intervention. The difference between the actual values and the SC after intervention is interpreted as a causal effect. One of the main advantages of SC methods is the data-driven weighting of control units, instead of simpler heuristics such as unweighted average or subjective judgment (ie, hand-picking of alternatives). But there remains considerable room for subjectivity in the selection of the control units from a larger population to form the "donor pool" and in the selection of variables used to assess the fit of the estimated control to the treated individual. In addition, the canonical linear regression format of the SC relies on an implicit assumption of linearity that might not hold in cases where a richer set of interactions between control units would better represent the target unit before treatment.

This paper argues that machine learning techniques obtain a more data-driven estimation of SC than current applications. Specific algorithms can select the control units from a larger population without human intervention, estimate the counterfactual with flexible functional forms, and allow for a comparison between the target and control units that is simpler to understand and analyse. This "machine controls" method is illustrated with an evaluation of the effects of Brazil's 2017 extensive labour deregulation reform on worker productivity, helping inform a long-standing policy debate (Vergeer and Kleinknecht (2010)). In this case, the reform failed to lead to a noticeable effect on average worker productivity, in spite of optimistic expectations by policymakers of 1.5-2.0% additional annual productivity growth.

Consider a panel time series with $t = (1, ..., T0, ..., T)$ periods and a population of individuals or units $j \in \mathcal{J}$. An intervention or treatment of interest happens at $T0$ to individual $j = 1$ and is continued for all subsequent periods.[1] Some outcome of interest, $Y_{j,t}$, equals the potential outcome under intervention $Y^I_{j,t}$ when $j = 1$ and $t \geq T0$ or alternatively under no intervention otherwise, $Y^N_{j,t}$. Some (but not necessarily all) $j \neq 1$ individuals in the population are informative about $Y^N_{1,t}$; $\mathbf{J} \subseteq \mathcal{J} \setminus \{1\}$ is the set containing these units. The following analyses assume $Y^N_{1,t}\, \forall\, t$ can be estimated by a potentially non-linear combination of the control units in $\mathbf{J}$ and an error term: $Y^N_{j,t} = \omega_Y(\mathbf{J}) + \eta_t$, with $\omega_Y(\mathbf{J}) : \mathbf{R}^{(Z)} \times \mathbf{R}^{(J)} \rightarrow \mathbf{R}^{(Y)}$ an estimator function taking as input control units' covariates $Z_{j,t}$, which might or might not include the actual outcome variable of interest, $Y_{j,t}$. In situations where the superpopulation is relatively large or contains a substantial number of units that are not informative to the estimation of $Y^N_{j,t}$, using only the units in $\mathbf{J}$ would yield a more precise SC. Formally, if $\mathbf{J} \subset \mathcal{J} \setminus \{1\}$, then $E[|\eta_t|] \leq E[|Y^N_{j,t} - \omega_Y(\mathbf{J} \cup \{\mathbf{j} \notin \mathbf{J}\})|]$. This might happen in situations, for example, where $\mathcal{J}$ is very large and including uninformative units would degrade

[1] Without loss of generality. In fact, the intervention could even happen to more than one individual.

model performance out-of-sample (even if slightly) or simply because estimates of ω_Y might not converge fast enough in a large set of controls even when they are informative.

In this work, a clustering algorithm uses pre-intervention data to select a donor pool $\hat{\mathbf{J}}$ of units that are more similar as a group to $j = 1$ than the wider population. Then, a supervised learning algorithm combines the elements of $\hat{\mathbf{J}}$ using the function $\hat{\omega}_Y(\hat{\mathbf{J}})$, also with pre-intervention data only, as in:

$$\underset{\hat{\omega}(\hat{\mathbf{J}})}{\operatorname{argmin}} \, Z_{1,t<T0} - \hat{\omega}_Z(\hat{\mathbf{J}}). \tag{1}$$

For example, if $Z_{j,t} = Y_{j,t}$ (Doudchenko and Imbens (2016)), then the SC is $\hat{Y}^N_{1,t} = \hat{\omega}_Y(\hat{\mathbf{J}})$. If $\hat{\omega}_Y(\hat{\mathbf{J}})$ is able to fundamentally approximate $Y_{1,t}$ before the intervention, then in periods after $T0$ the SC is sufficient to identify the effect of treatment $\tau_t = Y_{1,t} - \hat{Y}^N_{1,t}$. The SC's quality is assessed by comparing its pre-intervention values $Z_{\text{SC},t<T0}$ with the treated unit, in contrast with other $Z_{j\in\hat{\mathbf{J}},t<T0}$, to show that $d(Z_{1,t<T0}, Z_{\text{SC},t<T0}) < d(Z_{1,t<T0}, Z_{j\in\hat{\mathbf{J}},t<T0})$ for some distance measure d.

Estimating $\hat{\mathbf{J}}$ rather than using all elements of $\mathcal{J} \setminus \{1\}$ is important in many practical applications to minimise interpolation biases (Abadie, Diamond, and Hainmueller (2010)). In big data settings, the selection is also important to ensure the estimation lies within the computational budget. But the selection process to find $\hat{\mathbf{J}}$ is typically subjective, which allows the analyst to incorporate expert knowledge into the selection at the cost of introducing controls that might not have the best predictive power for $\hat{Y}^N_{1,t}$, leaving important information behind or potentially including uninformative units. This problem is somewhat mitigated by the estimation of the SC itself, which need only have a good (but not necessarily the best) pre-treatment fit; and the often sparse estimations deal with uninformative controls.[2] A worse problem from subjective selection is that it opens a flank for accusations that the pool is cherry-picked to achieve a certain result. Typically this is addressed proactively by a clear narrative explaining the selection of $\hat{\mathbf{J}}$. But a more structural and data-driven way to address both issues is to use algorithms that estimate $\hat{\mathbf{J}}$ based on some similarity measure from data, without human subjectivity.

These are known as clustering algorithms,[3] because they identify units that are the most similar to each other as a group within a larger population. These techniques protect the analyst against criticisms related to the real or perceived subjectiveness of the selection of control units, however careful and well-justified they were chosen. In fact, one advantage I observe in practice is that these

[2] When the target unit lies within the convex hull of the control units, there are infinite possible combinations of controls that optimise the SC (Abadie (2021)); penalising the selection of controls is generally sufficient to obtain a single optimal SC (Abadie and L'hour (2021)).

[3] More broadly, any type of estimator algorithm yielding some form of feature selection or feature ranking could be used to select control units.

methods often find elements of $\hat{\mathbf{J}}$ that can seem somewhat surprising to experts. Another advantage is that these clusters can find units similar to the treated individual using a (potentially high-dimensional, or including newer data such as text embeddings) data space that may or may not be the same as Z. In cases where variables driving treatment propensity are defined in theory, these could be explicitly incorporated to select controls that had similar propensities but were not subject to treatment.

The next step involves weighting the elements of $\hat{\mathbf{J}}$. The canonical SC method is a constrained linear regression on control units with non-negative coefficients. Doudchenko and Imbens (2016) discuss other more flexible linear models that include non-negative coefficients or a constant term. Still, they are linear models and thus may not be suitable in a variety of practical applications, in particular where $\#\hat{\mathbf{J}} \gg T0$. Also here machine learning techniques can help. Supervised learning methods, such as random forests (Breiman (2001)) or neural networks (Goodfellow, Bengio, and Courville (2016)), are flexible enough to accomodate non-linear data generating processes[4] and can regularise well even for high-dimensional inputs. These methods are increasingly deployed in economic estimation tasks due to their often superior predictive ability in practice (Athey and Imbens (2019)), although this may come at the cost of lower transparency of the estimation, especially when compared with the sparsity offered by canonical SC methods offering transparency and scrutiny by experts, who can inspect the individual contributions of control units and make a judgment on their adequateness (Abadie 2021).

A (likely non-linear) combination of controls $\hat{\omega}_Z(\hat{\mathbf{J}})$ from Equation 1 calibrated exclusively from pre-intervention data serves to estimate the post-intervention values of $\hat{Y}^N_{1,t}$. One advantage of supervised learning techniques is the potential to use alternative data. For example, texts such as central bank speeches (BIS (2024)) or monetary policy statements (Evdokimova et al. (2023)) can help estimate the effect of an idiosyncratic fiscal shocks on monetary policy, while satelite images (eg, Yeh et al. (2020)) enable the study of events such as war on economic activity in settings where availability of comparable control data might be scarce.

In this scenario where non-linearities and even non-traditional data can be used to estimate SCs, it is crucial to measure how well the SC matches the individual under treatment before $T0$. I see the process of checking the quality of an SC as a third step in its estimation process, after selecting $\hat{\mathbf{J}}$ and estimating $\hat{\omega}_Z(\hat{\mathbf{J}})$. Like the first two steps, the evaluation of SCs in canonical applications also affords considerable space for subjectivity in the choice of covariates on which the target entity and the SC are supposed to match. A different set

[4] J. Chen (2023) uses an analogy between SCs and online learning to show that the performance of SCs in practice is not dependendent on the underlying outcome generating process, generalising the analysis most commonly found in the literature of data generating processes modelled after linear factor model or a vector autoregression (Abadie, Diamond, and Hainmueller (2010)).

of machine learning techniques, usually referred to as manifold learning, can make this process more data-dependent and simultaneously easier to analyse. At its core, manifold learning algorithms are based on the manifold hypothesis, which dictates that many observed real life data processes are actually high-dimensional representations of a much lower-dimensional manifold. Intuitively, this is compatible for example with the observation that multiple economic series are highly corelated. Manifold learning techniques find lower-dimensionality representations of input data to a lower dimensional space that still retains most of the information about how each individual or unit is different from one another. In other words, these algorithms compare $j \in \hat{\mathbf{J}} \cup \{1\} \cup \{\widehat{\omega}_Z(\hat{\mathbf{J}})\}$ on a more fundamental level through their locations on the underlying data manifold. Their distance in this estimated manifold offers a practical test of the goodness-of-fit: ideally the SC should be closer to the treated unit than all other (actual) units.

Thus a completely data-driven estimation of SCs can be achieved by combining these methods listed above. These machine learning-enhanced SC, or "machine control" methods for short, can be helpful in settings with at least one of the following characteristics: (a) a large number of relevant data points for each sampled entity are available, (b) a large number of potential donor pool entities are available, (c) the best combination of donors to form a control is not necessarily linear and (d) the variable of interest is measured with alternative, potentially highly-dimensional data. Such an automated and flexible process might offer a higher level of credibility for SC-based inference out of the box due to the greater degree its results depend on data. They also allow for more efficient estimation in data-rich settings. In addition, it can be seen as a way to argue more convincingly that the control is in fact a valid counterfactual for the treated unit. All these algorithms have ample usage in other sciences, with some being increasingly popular also in economic research and practice.

1.1 Literature

The machine controls technique pieces together different types of machine learning techniques. This is in line with an increasing trend of applying machine learning methods in economic research to better explore the wealth of available data, as highlighted by Athey and Imbens (2019), Athey (2018) and other papers. In addition to the canonical SC applications mentioned above, this paper draws close inspiration from Doudchenko and Imbens (2016) and Ferman, Pinto, and Possebom (2020), who explore the implications of SC in more detail, and Viviano and Bradic (2022), Athey et al. (2019) and Quistorff, Goldman, and Thorpe (2020), who study non-parametric methods for SC estimation, including by combining predictors in ensembles. This paper proposes a more extensive use of non-parametric techniques through a combination of different classes of machine learning algorithms to make estimation of SC both fully data-driven and offer more flexibility in the estimation of the donor pool weights.

Consistent with the popularity of the original SC methods, various improve-

ments or more specialised adjustments have been proposed in the literature. Xu (2017) unifies SC methods with differences-in-differences (DiD) and Doudchenko and Imbens (2016) generalises SCs by discussing the implications of its restrictions, while Ferman (2021) analyse the asymptotic properties of SC estimators under imperfect fit. Abadie and L'hour (2021) posit the penalised SC method, geared towards settings with multiple treated individuals and Gobillon and Magnac (2016) discuss the effects of dimensionality reduction techniques or data transformations on the estimation. This work adds to this rich literature by proposing an empirical method for settings where machine learning algorithms can thrive, namely when ample data is available or when a non-linear combination might provide a better fit to the pre-treatment target individual.

1.2 Empirical application

I illustrate the combination of these methods with a brief analysis of the response of worker productivity in Brazil to an extensive labour market deregulation reform implemented in 2017. The reform was an attempt to cut wage costs by reducing bureaucracies related to hiring and dismissals, promoting work council collective wage negotiations and facilitating compliance with social security payments. Policymakers expected the reform would unlock between 1.5% and 2.0% of annual productivity gains. In this application, the elements of $\mathcal{J}$ are other countries. Since the number of countries is not large compared with computing capabilities, all of them could in practice be in the donor pool. However, it might be advantageous to restrict the set of countries acting as controls to those that resemble more Brazil (Abadie, Diamond, and Hainmueller (2010)). Still, there are no particular reasons to impose any specific *number* of countries to be analysed. This calls for a flexible clustering algorithm like affinity propagation (Frey and Dueck (2007)) to find $\hat{\mathbf{J}}$. The control itself, $\hat{\omega}$, is estimated using a random forest (Breiman (2001)) with fine-tuned hyperparameters. Finally, the overall fit of the control was assessed by means of a manifold learning technique, t-Distributed Stochastic Neighbour Embedding (t-SNE) (Van der Maaten and Hinton (2008), Van Der Maaten (2009)).

In this application, the causal policy effect τ_t was not different from zero, staying well within the range of placebo results for each $t \geq T0$. This indicates that the reform did not achieve one of its stated goals of enhancing productivity, even years after it was implemented. There might be several reasons to explain this (Vergeer and Kleinknecht (2010) lists arguments relating wage growth positively and negatively to productivity growth). Further exploring this finding, and its implications for labour reforms in other jurisdictions, is left to future work.

Interestingly, the other countries in $\hat{\mathbf{J}}$ are substantially different from Brazil at first glance: they include small islands, former (and current) communist economies, and other countries which do not seem like choices that would be made by most human analysts. Another noteworthy fact is that the manifold learning algorithm resulted in a very clear way to visualise the SCs, especially in the context of all of the other controls individually. In addition to plain

visualisation, one could check that the haversine distance between the treated unit and its control is the smallest of all other bilateral distances (ie, between the treated individual and the other controls) as a measure of the SC's quality.

2 Brief review of synthetic controls

This section briefly reviews the traditional SC methodology, based on the potential outcomes framework (Imbens and Rubin 2015). For ease of reference, the terms such as "treated", "intervention" and "event" mean the event of interest that is being studied. This intervention does not need to be anticipated or endogenous to the units; in fact, it could also be a completely exogenous treatment assignment.

Consider a panel dataset of J entities recorded over $t = (1, ..., T0, ..., T)$ time periods, with the treatment or intervention occuring in $t \geq T0$. Entity $j = 1$ is normalised as the target entity. A researcher wants to find the effect of the treatment on the outcome variable,

$$\tau_{1,t} = Y^I_{1,t} - Y^N_{1,t}, \tag{2}$$

where the superscript I represents the potential outcome under the intervention and N the potential outcome with no intervention. In other words, Equation 2 represents the effects of the idiosyncratic event on the affected individual. Since $Y^I_{1,t}$ is observed for $t \geq T0$ but $Y^N_{1,t}$ is not, estimation of this equation is accomplished by calculating the counterfactual

$$\hat{Y}^N_{1t} = \hat{\omega}_2 Y_{2t} + \cdots + \hat{\omega}_{J+1} Y_{J+1,t}. \tag{3}$$

Abadie and Gardeazabal (2003) show that SCs, rather than simple donor pool averages, are better able to approximate ω^*_j for $t < T0$ in the following equation:

$$\hat{H}_{1,t} = \omega^*_2 H_{2,t} + \cdots + \omega^*_{J+1} H_{J+1,t} \tag{4}$$

where $H_{j,t}$ can be a set of covariates $X_{j,t}$ as traditionally used, with or without some information from the outcome variable $Y_{j,t}$; or just the outcome variable itself (as in Doudchenko and Imbens (2016) and other papers in the SC literature). As noted by Kaul et al. (2015), either the full history of the outcome variable or the covariates can be used to calculate the weights. In the original application, the weights are restricted to be $0 \leq \omega^*_j \leq 1$, but other authors have discussed the advantages of a more flexible framework for weighting control individuals (eg, Doudchenko and Imbens (2016)).

In any case, the weighting step, which is the core of the SC methodology, is data-driven. However, the intermediate steps to get to this estimation involve the manual selection of $\hat{\mathbf{J}}$ (which involves deciding both the elements and implicitly

the numer of comparable units), and of the variables on which to measure the fit of the SC. The better the pre-treatment fit of the SC with this version made up from other entities, then the best the estimates of Equation 2 are, especially if the selected variables have explanatory power to $Y_{j,t}$, as argued by Abadie (2021). Even though this methodology might in many cases yield well-functioning SCs, the existing degree of subjectivity leads to a non-zero possibility that the best controls were not always achieved in practice: for example, the analyst might not have selected the donor pool in a data-driven way but instead only focused on a few characteristics.

For example, in the application below studying the impact of labour deregulation on worker productivity, in which the broader population of potential controls are the other countries, a baseline machine control estimation finds that countries like Albania and Mauritius are in the main donor pool for Brazil, even though they would hardly be selected by a human analyst. This is important because there are in many instances a variety of dimensions under which entities could be aggregated, with subjective judgement still playing a larger role in the selection of which ones are valid in each application. And this selection is important as it can lead to changes in the final result, as shown by Klößner et al. (2018).

And similarly, as extensively studied by Ferman, Pinto, and Possebom (2020), the specific features of the entity which are being matched by the synthetic control are also chosen subjectively in many cases. (Note that these variables are either used just as a comparison to further claim that the SCs are a good match to the pre-intervention target, and or as covariates themselves, in $H_{j,t}$.) While Ferman, Pinto, and Possebom (2020) recommend reporting a range of specifications and putting in place clear criteria for selecting the covariate space, a more straightforward and broadly applicable way would be to also here leverage machine learning techniques to automate this step. In particular, this can be achieved with manifold learning, which summarises the data available for each entity into substantially fewer data points. Manifold learning is posited as the best-suited group of techniques on account of their ability to find common patterns in data in a non-linear way, as opposed to, say, principal component analyses (Hotelling 1933). This way of comparing the original and synthetic outcomes is particularly important for settings where the analyst has access to a relevant number of variables for each entity.

3 Synthetic controls with machine learning techniques

Building on the preceding discussion, the current section elaborates on specific machine learning techniques that can help each step of the SC process.

3.1 Selection of control variables

While machine learning models can generally work well with high-dimensional input, including the case where the cardinality of $\mathcal{J}$ grows to a large number (Goodfellow, Bengio, and Courville (2016)), identifying a more restricted set of control individuals in $\mathbf{J}$ can help the estimation of $\widehat{\omega}$ by concentrating on a more informative set of covariates. This selection of the "donor pool" is also performed by the classic implementations of SC methods since Abadie and Gardeazabal (2003), many of which use a relatively small number of controls compared to the overall population. But the traditional selection process is a great deal subjective, although advances are occuring to alleviate that concern: for example, Abadie and L'hour (2021) use penalised linear regression to select control units. In contrast, machine learning algorithms used specifically for clustering are a completely data-driven way to estimate $\widehat{\mathbf{J}}$.

The diverse group of clustering algorithms take as input covariates of a population and map them to a set $\{c_{i,j} = (0,1) \, \forall \, i,j \in \mathcal{J}, j \neq i\}$, where 1 indicates that the two individuals i,j are in the same cluster. Equivalently, the result can be reformulated as the cluster mapping $C : \mathbf{N}^{\mathcal{J}} \rightarrow \mathbf{N}^{\mathcal{C}}$, with the first term standing in for the index of each individual in $\mathcal{J}$ and the last term representing the index of each cluster in the set of clusters $\mathcal{C}$. Important in the case of SC estimation is the estimation of the cluster of the treated individual $j = 1$: $C_1 = \{i | c_{1,i} = 1\} = \mathbf{J}$.

There is a wide varierty of clustering algorithms from which to choose; one of them, affinity propagation (Frey and Dueck (2007)), concentrates all desirable characteristics for this particular type of application. The first dimension in which algorithms differ is in the choice of the number of clusters, $|\mathcal{C}|$. Intuitively, the data has all necessary information to guide the choice of the number of different clusters, unless there is a theoretical reason that individuals should be divided into a specific number of clusters. For example, if identification of a structural model requires that individuals be selected into one of two clusters (which can be interpreted as domestic vs foreign), then the number of clusters is implicitly defined from theory. But absent that, the number of clusters should in most cases reflect the similarity between individuals more than any assumption or hard-coding by the analyst. A second characteristic is the number of individuals in each cluster. It is hard to conceive good reasons why $\mathbf{J}$ must have 10, 20, 30 or any other specific number of different units. Ideally the data should be the only source of information about the cardinality of each cluster.

Other machine learning techniques can also select amongst a potentially large set of individuals. Penalised regressions of the target variable's characeristics on potential control units' data using techniques such as lasso (Tibshirani 1996) or elastic net (Zou and Hastie 2005) can result in sparse coefficients that identify the most relevant control units. Doudchenko and Imbens (2016), for example, use elastic net (Zou and Hastie (2005)) as a regularisation procedure to calculate linear weights amongst a larger set of donor individuals. They also present the

advantage that $\hat{\mathbf{J}}$ and $\hat{\omega}_Y$ are jointly optimised. But there the linearity of the model is an important downside. So while these methods result in a sparse selection of controls, I advocate for other techniques for this stage: clustering techniques. These methods can be used to first select the controls, and then the estimation of how these units can be combined would be done separately (and more flexibly) in a dedicated stage (see Section 3.2).

Another advantage of separating the steps of control selection and SC estimation is that the placebo estimation can be done more consistenly with a donor pool that is selected to be *jointly* more similar to one another, than a donor pool that optimises only for $j = 1$. Intuitively, this makes inference with the placebo test more precise.

Affinity propagation is an algorithm that assigns individuals to clusters based on a series of iterations, in which each unit exchanges information on their pair-wise similarity, with higher values denoting pairs that are more likely to be connected. For example, if the goal is to find other units that minimise the squared error, then $s(f, g) = -\|\mathbf{Y_f}^\star - \mathbf{Y_g}^\star\|^2$ where $f, g \in \mathcal{J}$ and the star superscript * denotes that the information is only based on time periods until $T0$. Based on these similarities, individuals gradually coalesce around spontaneously-forming cluster centres (also known as "exemplars"), until the cluster assignment stabilises. The mechanism behind this result is an interaction between all data points to exchange two types of "messages", denoted "responsibility", $r(\centerdot)$, and "availability", $a(\centerdot)$.[5]

Note that in this application I use the similarity criterion above, so the resulting clusters minimise the squared error between units; but other objectives could also be used depending on the case at hand. For example, the similarity measures might be asymmetric with $s(f, g) \neq s(g, f)$ or violate triangle inequality, $s(f, h) > s(f, g) + s(g, h)$. In fact, the similarity can also accomodate empirical or theoretical reasons why two given units (countries in this case) should be clustered together: if $s(\cdot)$ is considered the log-likelihood of f conditional on its cluster examplar being g, or as a log probability of a connection existing between f and g. More generally, this measure $s(\cdot)$ could be conditional on external covariates that are not influenced by the policy at hand. Another practical alternative is to use economic distance measures (eg, Fisher et al. (2015)) of unit embeddings. The similarity "adjacency matrix" might also be sparse, with only a subset of the possible f, g pairs having a connection.

Responsibility is set to the accumulated perception of individual f on the suitability of individual g to be the exemplar of its cluster, compared to all other examplar candidates, as described in Equation 5:

$$r(f, g) \leftarrow s(f, g) - max_{g' s.t. g' \neq g}\{a(f, g') + s(a, g')\} \tag{5}$$

[5]This is similar in spirit to message-passing algorithms, which perform well across a wide range of algorithmically hard problems (Mézard (2003)).

where the update rule of $a(\cdot)$ is described in Equation 6. Values of $r(g,g)$ represent the evidence that g is itself an exemplar. The other metric, availability, is the message sent by g to f describing how appropriate it would be for f to choose g as its exemplar, considering the support received from other units for having g as an examplar.

$$a(f,g) \leftarrow \begin{cases} min\{0, r(g,g) + \Sigma_{f' s.t. f' \neq f} max(0, r(f',g))\} & for f \neq g \\ \Sigma_{f' s.t. f' \neq f} max(0, r(f',g)) & for f = g \end{cases} \quad (6)$$

These cluster centres arise from a purely data-driven process. However, for machine controls the designation as exemplars is not particularly important: what matters more is the list of individuals in the same cluster as the target individual, since the calculation of the controls will be made with their data. Whether or not the targeted individual is a cluster exemplar does not make any difference.[6] Therefore, while users can influence the chance each unit f is chosen as exemplar by setting $s(f,f)$ for each entity differently, I recommend setting all $s(f,f)$ equal, minimising the priors on the cluster formation, unless there are clear theoretical reasons to set some units as being more likely than others to serve as cluster exemplars. Note also that even when $s(f,f) = s(f',f')$ for each f, f', the *number* of clusters depends on its level. A common practice is to set this self-preference to the median of the distances.

Because all individuals are considered to be potential exemplars, and those are set gradually by the algorithm, affinity propagation is not subject to eventually unlucky initialisations. Given that clustering is proposed to be used in the seleciton of the donor pool entities, this stability is a valuable characteristic. Another advantage is that the targeted individual does not need to be the centre of a cluster. This opens up the possibility that it is a peripheral member of a cluster, which helps find other units that resemble each other as a group. Yet another upshot from using affinity propagation is the possibility of selecting data from a potential sparsely related populaton (Frey and Dueck 2007). Further, this algorithm affords the researcher the possibility to establish functional forms for the similarity between units, which are flexible enough to accommodate cases where the metrics are not symmetric (ie, are not distances strictly speaking). For example, similarity in second moments can be used if it would make sense in terms of achieving a consistent fit of the SC.

Feature selection algorithms can also be used in this stage as well, in conjunction with a clustering algorithm or in its place. While these methods usually serve to select covariates on which models should be trained, in the case of machine controls they can be effectively used to pre-select controls from the larger donor pool. This could be especially useful when this population is significantly large. As before, it is important to reiterate that such a control selection (in lieu of

[6] As seen in the application below, an area where exemplars can be helpful is by using those from *other* clusters to act as "representatives" of non-neighbour data points if the researcher wants to compare them with the target individual, eg as a robustness exercise. See Figure 2.

feature selection) should be performed using only data up to the treatment date at the latest.

In spite of its advantages, in some cases a degenerate clustering might occur where the treated unit is its own cluster, with no other unit. In these cases, the first thing to do is to increase the number of iterations until stabilisation; it might just be that the algorithm takes a while longer to converge than the default number of iterations. If that still does not work, the analyst might reconsider whether the similarity function or the data format used are indeed the most adequate one, which depends of course on the specific application. For example, finding a cluster of publicly traded companies is likely to be more relevant if the squared error of share returns, rather than actual prices, are used. In situations where this also does not work, another possibility is to use other techniques such as DBSCAN. With this technique, clusters are not necessarily constrained to be convex in the low-dimensional representation, and DBSCAN is its deterministic nature, whereby the same data provided in the same order results in the same clusters. However, this method is sensitive to the order in which data appear and thus might not be suitable in practice for a wide range of applications as it seems an unreasonable assumption to justify why one unit would be considered a donor in one order, but not in the other. If still in this case the cluster is degenerate, the analyst should probably reconsider whether there are in fact units that could serve as suitable controls in the first place.

3.2 Estimation of counterfactuals with supervised machine learning algorithms

Statistical guarantees of SCs are traditionally derived from low-rank, linear data generating process, namely factor models or vector autoregressions. But as J. Chen (2023) shows, SCs actually benefit from theoretical guarantees irrespective of the outcome model, at least relative to other plausible alternative models: if there is a weighted matching or weighted differences-in-differences model that performs well, SCs would tend to have a similarly good perform.

More generally, the use of machine learning algorithms for the estimation of SCs can be cast under the light of a regression-on-residuals model (Robinson 1988), similar to the semiparametric models literature. This insight is inspired by the applications in Nie and Wager (2021) and Athey and Wager (2021). In effect, considering SC estimation through this prism is also an illustration of its flexibility: there are no specific requirements on the functional form (let alone that it be linear), other than that they must converege at a reasonable but not necessarily efficient pace to the true answer. This can be motivated under a potential outcomes framework (Rosenbaum and Rubin 1983).

In the subsequent analyses I follow Doudchenko and Imbens (2016) and J. Chen (2023) and estimate the controls directly on the outcome variable.

3.2.1 Choice of algorithm

There is a wide range of supervised machine learning algorithms from which an analyst can choose. In this paper, I follow Araujo (2023) and use random forests (Breiman (2001)) due to their robustness to various different use cases, ease of use (with limited required manipulation of the data, such as scaling), flexibility to tune widely different models with only a few parameters, and wide application.[7] Wager and Athey (2018) find that random forests are more powerful than k-nearest neighbours methods, especially in settings with noisy covariates. In essence, random forests are a collection of B trees θ_b, each creating a nonparametric expectation function $\mathbf{E}[Y_{j,t}] = f_{\theta_b}(X_t)$ based on the concept of "leafs" L_b, which are the final steps of covariate partitions that group together similar observations. These trees are trained with data points drawn with resample from the training dataset and, in many specifications, only using a certain subset of covariates (which may be randomly chosen) within a user-defined limit in the number of covariates. Because of the limited data seen by each θ_b, each one of them tends to individually have low bias but a high variance. Once they are aggregated, such as by averaging their predictions, their variance is reduced (Bühlmann and Yu (2002)). Empirically, gradient boosting algorithms such as XGBoost (T. Chen and Guestrin (2016)), LightGBM (Ke et al. (2017)) and CatBoost (Prokhorenkova et al. (2018)) tend to perform quite well also.

Similar to classical implementations of SC methods, tree-based methods are also able to select a small number of control units from the donor pool to construct the weights. But there are three crucial innovations. First, the selection done by the trees is nonlinear and depends on the specific datapoints. Second, the actual weights depend on the combination of trees. Another reason to use supervised learning algorithms is their flexibility to include more than one result. A third innovation of the machine control method is that it enables the study of more than one outcome variable at the same time (ie, $Y_{j,t}$ could actually be a vector), providing further flexibility to the estimation of treatment effects. Of note, during the actual estimation covariates can be included either directly as input features to train the model, or by estimating a residualised version of the outcome variables as suggested by Doudchenko and Imbens (2016).

3.2.2 Cross-validation

Cross-validation helps ensure that the estimator $\hat{\tau}$ is $N^{-1/2}$ consistent, although the literature offers mainly examples in cross-sectional problems (Chernozhukov et al. 2018). In the case of time series as considered in SC settings, cross-validation fulfills a different function: by avoiding the use of contemporaneous

[7] Other supervised machine learning methods have gained wider adoption in the literature recently. For example, Chernozhukov, Wüthrich, and Zhu (2021) use lasso (Tibshirani 1996) and elastic net (Zou and Hastie 2005) methods as regularisation procedures for the weights, as opposed to the original restrictions that they should not be negative and should sum to one. Farrell, Liang, and Misra (2021) show that causal inference with neural networks is also statistically valid as these algorithms converge sufficiently fast.

or future data to train models that are evaluated on held-out data, this step regularises the estimates of $\hat{\omega}_Z(\hat{\mathbf{J}})$ by identifying the best out-of-sample performance amongst different model parameterisations p, where p are algorithm-specific parameters from hyperparameter space $p \in \mathcal{P}$ defined by the analyst. Note that during cross-validation, all held-out data at each validation step always comes from the pre-treatment period.

More formally, consider $\mathcal{K}$ equally-sized divisions of the training period $t = (1, \dots, T0)$. The cutoff dates between folds are represented by the set $t^{\mathcal{K}} = \{t^k : k = (1, \dots, \mathcal{K}-1)\}$. Then for each $t^k \in t^{\mathcal{K}}$, candidate models $M(p) \forall p \in \mathcal{P}$ are sequentially trained on $t < t^k$ and evaluated on data from periods $t^k \leq t < t^{k+1}$ (where $t^{k+1} \equiv t^*$ at the last fold). For a loss function $L(\cdot)$ and a function $T(\cdot)$ mapping each cutoff date t^k to a corresponding duple $\langle [1, \dots, t^k), [t^k, t^{k+1}) \rangle$, the chosen parameter is:

$$\arg\min_{p} \frac{1}{\mathcal{K}} \sum_{k \in \mathcal{K}} L(\hat{\omega}_{Z_{t \in T_2(k)}}(\hat{\mathbf{J}}, M(p)_{t \in T_1(k)})). \tag{7}$$

In other words, Equation 7 chooses the parameter set that minimises the average loss computed in held-out data in sequential folds, all contained in the pre-treatment periods.

A related procedure is sample-splitting, dividing the pre-treatment period in two parts (usually halves). Viviano and Bradic (2022) uses the first part to train candidate models $g_q(\cdot)$ where $q \in \mathcal{Q}$ is an element in the space of data-driven, machine learning models. The second part of the pre-treatment sample is used to assess the different weights that each model should have in an ensemble that minimises loss.

3.2.3 Limitations

Beyond just random forests, one of the drawbacks of using supervised machine learning models in general as weighting algorithms is that depending on the algorithm, it is not straightforward to guarantee non-negative weights, or that the weights will be sparse as originally proposed by Abadie and Gardeazabal (2003). However, this is not a major impediment: retaining the ability to weight individuals negatively better allows for cases where the treated observation is an outlier, ie outside the normal range of donor pool outcomes, and it also avoids spuriously assigning non-zero weights to individuals with low correlation (Doudchenko and Imbens 2016). Another practical disadvantage is that it might be challenging to re-use ω to weight

3.3 Assessment of fit with manifold learning algorithms

Economically-relevant settings increasingly count with a large number of variables. Still, this higher dimensionality of datasets is not necessarily meaningful, in that the underlying data manifold is probably low-dimensional in practice.

For example, in finance, the Fama-French factors (Fama and French 1993) summarise stock and bond returns in a few dimensions, and factors similarly summarise hedge funds returns (Fung and Hsieh (2004)). At the macroeconomic level, Rey (2015) documents that capital flows, credit growth and asset prices exhibit a large degree of comovement in a global financial cycle that seems to be driven by US monetary policy. Those are only some of the ample evidence that economic and financial phenomena tends to be amenable to lower-dimensional representations.[8]

For this reason, I propose the use of manifold learning techniques to accommodate the greater data availability in empirical settings. With more data, a subjective choice of the set of matching covariates might be more strenous, as well as hard to justify in some cases: why was this variable chosen, and not that? Manifold learning completely bypasses this problem by summarising all the available data in a low dimensional way that is still comparable by humans. Manifold learning algorithms resemble principal component analyses (PCA), but in a way that considers potential non-linear combinations of components. These methods are mainly used to reduce dimensionality of a variable space such as $\mathbf{R}^n$ for a large n, while retaining a substantial chunk of the original signal, which is hypothesised to come originally from the manifold.

But even when the dataset itself is not particularly large and there is already certainty as to which are the specific variables that should be compared, such techniques are useful because they confirm that the treated entity and the calculated control are indeed close even in this more fundamental way. For example, if all an analyst has is the time series of the outcome variable, each unit's time series can be used to estimate the manifold. In factor model language, it would be analogous to finding $j \neq 1$ with similar factor loadings β_j as β_1 in $y_{jt} = \beta_j \lambda_t + \nu_t$, with the last component a nuisance factor. Note that, in this case the treated unit will be compared to other units' cycles and developments over time, which itself might have interesting empirical applications. For simplicity of exposition, this is the approach used in the empirical simulation.

In empirical applications, the use of SC methods entail the researcher examining the distance between the treated unit and the synthetic control with respect to specific variables. The more they are close together, intuitively the more the control matches the data. Consider a setting where a large number k of distinct covariates is associated with each individual j, collected in the $m-$ dimensional vector x_j. Out of this data, only x_j^i, $i = 1, ..., n$ for $n \ll m$ carry distinct signals for a comparison with x_1^i to be meaningful (and accepted as such by other analysts). However, n is unknown and therefore the analyst never really fully knows whether the match is indeed good. Conversely, relying on the underlying manifold directs, by design, the comparison to a signal-rich set of variables. Algorithms that "learn" the underlying manifold tend to represent it in a way that makes distances between points comparable. Of course, the

[8] This is also seen in the physical world, as illustrated by examples mentioned in Goodfellow, Bengio, and Courville (2016), sections 5.11.3 and figure 20.6.

"manual" comparison will still be interesting if those particular variables of interest tend to not follow the same underlying manifold as the other ones.

Importantly, a broad range of manifold learning algorithms are sensitive to occasions when the underlying manifold is not linear. Specifically, the empirical application in this paper uses t-SNE (itself based on normally-distributed prior work by Hinton and Roweis (2002)). t-SNE enables the visualisation of multi-dimensional data in two or three dimensions by estimating a map to the low-dimensional manifold that underlies the structure of the dataset. This map minimises the Kullback-Leibner divergence between the probability distributions of the distances between points in the original and the embedding dimensions. The distribution of distances in the original data are modelled as a normal distribution but in the embedding, they are modelled as a heavy-tailed t-distribution. As a consequence, t-SNE achieves a low-dimensional representation that keeps similar data points close together in a two- (or three-)dimensional representation, while giving points far from each other liberty to position themselves far away. This means that if the treated individual and the control are close to each other, compared to other units, they are estimated to have a similar probability of being drawn from an underlying distribution. This is not usually attainable with linear mappings such as PCA. t-SNE' representation reflects both local and global underlying structure of the data, ie it can also reveal the presence of clusters of individuals (albeit not "clustering" them in the sense of establishing well-defined groups). This feature is something which most other nonlinear manifold learning algorithms are also not able to deliver at the same time in a broad range of naturally occuring datasets. t-SNE is the algorithm of choice in across various domains, and has been proved theoretically (Arora, Hu, and Kothari (2018)).

Two disadvantages of this particular manifold learning technique are that different settings for the perplexity parameter may yield considerably different results, and that if the feature space (ie data dimensionality) is very high, then calculation can be slow and other methods might be more adequate. This drawback is well illustrated by Wattenberg, Viégas, and Johnson (2016). Another disadvantage is that t-SNE might not perform well on datasets with a high-dimensional underlying manifold, ie, where the data points are really different across each other across multiple fundamentally different dimensions. However, this is unlikely to be the case in empirical research in economics and finance.

Note that the objective function optimised by t-SNE is not convex; the objective is minimised by gradient descent, randomly initialised every time the algorithm is run. For this reason, although results tend to be fairly close across each random state run, a reasonable practice is to run the simulations a few dozen times and average them out or pick the one that performs better. Another important thing to note on the interpretation of these results: the divergence metric between two points might not necessarily represent distance, in that it is not symmetrical. The dimensions resulting from t-SNE or other algorithms might carry some interpretative meaning, which might not always have a clear

interpretation.

In addition to a visual check, the estimated lower-dimensional manifold can provide quantitative information on the appropriateness of the machine control. A simple but intuitive test looks at the Euclidean distance between the target unit and all others, including the SC, across m runs, for m a potentially large number. An ideal control would always have the lowest distance, ie be the first in the rank of negatively ordered distances. If $w(j)$ is a function that returns 1 when j is the closest unit to $j = 1$ and 0 otherwise, then the percentage of runs in which the control is the closests, $w(j = \text{SC})/m$, can be interpreted as a form of p-value.

3.4 Modularity

These techniques can all be applied separately as modules, or used together to estimate a fully-data-driven SC, the machine control[9] The module approach might be of particular value in settings where there is not enough data. More specifically:

- when the number of pre-treatment periods is small compared to the number of controls ($T0 \ll J$), then clustering techniques could be used to pre-select the control individuals from the larger donor pool, and if the number of individuals is still relatively large, even they might be weighted using regularised methods such as lasso or elastic net, as suggested by Doudchenko and Imbens (2016);
- in cases with much more pre-treatment periods than the number of control units ($T0 \gg J$), the analyst might consider supervised machine learning techniques to estimate the weights because of their greater flexibility in reflecting nonlinear relationships in the data, such as the ones that might occur in tail events observed in economic or financial time series;
- when the covariate dimensionality is potentially large and the analyst wants a completely data-driven way to prove that the SC and the treated individual closely resemble each other in the pre-treatment period, the use of manifold learning techniques can offer a simple and intuitive way to accomplish that.

4 Empirical estimation

The machine controls method described above is illustrated with an estimation of the impact of labour deregulation on output per worker, measured in constant

[9] An extension of this work is being explored whereby the machine control is estimated by a joint search over the hyperparameter space using the three techniques, ie the supervised learning cross-validation includes a search over controls from the treated unit's own cluster or from the larger donor pool, using an objective function that includes, or consists exclusively of, minimising the distance of the estimated low-dimensional manifolds between the control and the treated individual.

2017 international US dollars PPP.[10] Because of the method's flexibility, the analyses can be significantly more complex if the use case so requires. But the point of this admittedly simple exercise is to illustrate the thought process and a basic workflow of machine controls, while informing about the effects of structural reforms that many countries are encouraged to undertake to increase growth potential.

More specifically, the use case focuses on Brazil's 2017 labour reforms, which comprised four new legislations. The primary statute reforming labour relations is Law No 13,467/2017, but other contemporaneous reforms also changed the face of labour relations in Brazil: Law No 13,429/2017 introduced the concept of outsourcing in Brazilian legislation, Law No 13,446/2017 about worker severance fund reform, Law No 13,456 about unemployment insurance (Biljanovska and Sandri (2018), Silva (2018), Carvalho (2017)). The reform deregulated labour markets, while introducing specific regulations in modern themes that were previously absent in Brazilian legislation (such as outsourcing or home office work). One of the main stated purposes of the reform is to increase productivity by facilitating a more efficient labour allocation, thereby unlocking growth (Silva (2018)). The labour reform was enacted in July 2017 and went into effect in November of the same year.[11]

The reform was wide-reaching; the main points for productivity and wage costs relate to active employer-employee relationship and to dismissal costs.[12] Changes influencing active relationships include the provisions that consider as "work hours" only the time actually spent working, with time on-call no longer part of "work hours"; allow work contracts by period, with remuneration proportional only to time worked; establish the prominence of collective bargaining between firms and employees over "blanket" statutory provisions; and convert union fees paid by employees from mandatory to voluntary.[13] The labour reform also introduced fines to companies that discriminate wages for workers in the same function based on gender or ethnicity.[14] And as for labour termination, the reform lowers costs for employers of termination without just cause. The new Law considerably curtails the margin for intervention by courts, discouraging labour litigation by employees. Further, dismissals without cause were previously subject to compensation to the worker proportional to their amounts deposited in the worker severage fund (FGTS in the Brazilian Portuguese acronym); after the reform dismissal costs are much lower to employers if both parties reach an agreement.

[10] The empirical calculations use gingado's `MachineControl` estimator.

[11] Brazil also implemented rules for home office in response to the pandemic and a more recent law to promote gender pay equality.

[12] The reform is described in more detail in Brazilian Portuguese sources (eg, Carvalho (2017)).

[13] A. G. Campos and Silva (2023) examines the impact on unions, and found that in 2020 unions had lost a third of pre-reform workers

[14] The Brazilian Constitution of 1988 forbids wage, function or admission criteria differences based on gender, age, skin colour or civil status. Krawczun et al. (2020) discusses this aspect of the 2017 labour reform in more detail.

This empirical exercise illustrates the machine controls method but also fills a gap in the literature on labour reforms: their effects are estimated usually in terms of unemployment outcomes, which is definitely important but does not tell the whole story about long-term prospects for growth and welfare. Serra, Bottega, and Sanches (2022) use traditional SC methods to analyse the effect of the 2017 Brazilian labour reforms on unemployment, finding no discernible effect. Adascalitei and Pignatti Morano (2016) reviews a large number of reforms at different levels implemented by various countries and find that deregulation tends to lead to higher unemployment in the short term. Brancaccio, De Cristofaro, and Giammetti (2020) conducts a meta-analyses of 53 academic analyses of the employment effects of labour deregulations and find a slight majority that show deregulation to increase unemployment and another large portion to have ambiguous results.[15] Data reported by Lima, Wilbert, and Silva (2021) suggest that there was an increase in informality levels after the Brazilian labour reform.

The Brazilian reforms were expected to increase productivity between 1.5% and 2.0% annually.[16] In general, this was expected to occur through lower wage and non-wage costs, as well as by facilitating termination decisions, ultimately making hiring decisions more easily. One of the channels by which productivity could increase is the greater efficiency of worker councils for collective negotiations on wages. Lima Júnior, Cavalcante, and Pinto (2016) argue that while work councils were permitted by the Brazilian Constitution, they were rarely used, in part because of counts striking down agreements that did not conform with a number of other worker protection rules. These authors claim that greater use of worker councils would increase productivity, in line for example with the experience in Germany (Addison, Schnabel, and Wagner (2001)). Similarly, Biljanovska and Sandri (2018) use a sample of advanced and emerging economy countries in regressions that suggest that labour market reforms (in the sense of deregulating labour relationships) could lead to statistically higher total factor productivity. Still, Biljanovska and Sandri (2018) take a cautiously optimistic tone that reforms might generate large productivity gains but it would pay off to wait and see the full impact from the 2017 labour reform before advancing further changes. Other widely known positive results of modernising reforms such as in Colombia (Eslava et al. (2004)) could have played a role.[17]

Other countries have enacted different levels of labour market reform in the years leading up to the Brazilian reforms. Serra, Bottega, and Sanches (2022) removed Argentina, Costa Rica, Paraguay, and Uruguay from the donor pool due to recent reforms these countries implmented, but their synthetic control donor pool consisted exclusively of geographically close countries. But doing

[15] Theoretical work by Saint-Paul (2002) examines when countries decide to implement labour reforms. Galí and Rens (2020) finds in an RBC model with labour frictions that lower frictions cause employment levels rather than worker effort to be the margin of adjustment.

[16] https://economia.uol.com.br/noticias/estadao-conteudo/2017/06/03/reforma-trabalhista-elevara-produtividade-diz-governo.htm

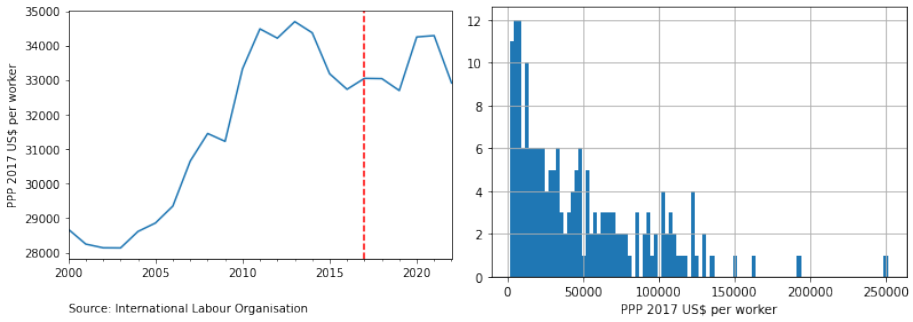
[17] Kouamé and Tapsoba (2019) also documents positive effects on firm-level productivity in developing economies from structural reforms, including in the labour market.

this when looking at a broader set of potential donors is impractical. Many countries have also enacted labour reforms in that period (as seen by Adascalitei and Pignatti Morano (2016) in ILO databases, European Commission's LABREF and IZA's fRDB data). Countries typically implement various levels of adjustment to their labour laws on a rolling basis. In addition, even if it were easy to identify countries with similarly structural levels of reforms as in Brazil in a straightforward way, I don't see a problem in keeping them in the main estimation database for two reasons. First, they might not make it to $\hat{\mathbf{J}}$ in the first place. And even if they did, their presence would make inference based on a placebo test harder, not easier, if the effects were (a) in the same direction and (b) significant.

Figure 1 shows what the series of annual output per worker in Brazil looks like over time, and a distribution of the productivity across countries in 2016. In the left panel, a vertical line mark November 2017, when the labour reforms entered into force.

Figure 1: Worker productivity

(a) Brazil (b) Cross-section as of 2016



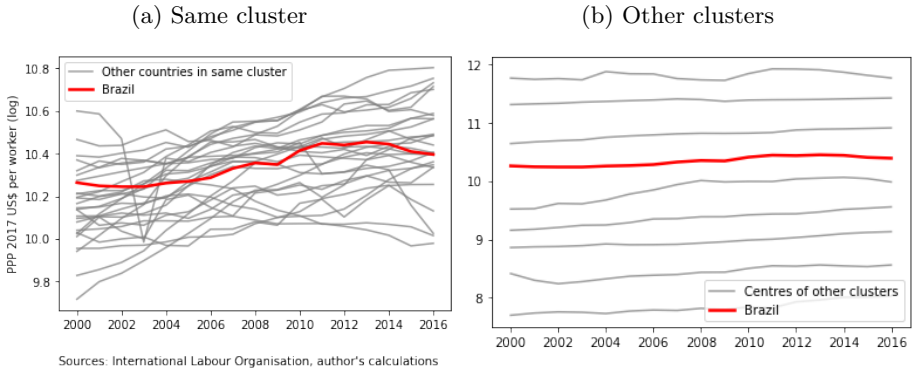
As discussed above, the machine control estimator comprises: a clustering algorithm that selects the donor pool from the larger population (affinity propagation), a supervised learning algorithm that will use the donor pool to estimate contemporanous values for Brazil (random forest) and a manifold learning algorithm that summarises the different time series in a 2-dimensional embedding space, enabling easier comparison of Brazil with each other country and with the synthetic control (t-SNE). Before creating the machine control, a final comment on the intervention date. Since the reforms were enacted and entered into force in the same year of 2017, we can be conservative and consider up to end-2016 as the pre-intervention period.

Choosing the donor pool subjectively is challenging. Which countries would best represent a control for Brazil? One possibility is to explore geographical proximity and select the other countries in South America. Or should other Latin countries in Central America and Mexico be included as well? How about

the United States, an important trading partner? But then that opens up for China, an even more important trading partner of Brazil. Perhaps then the controls should be the BRICS countries, ie, include also other large emerging economies such as India, Russia and South Africa. But why not other emerging economies with similar population such as Pakistan and Indonesia? The questions would be endless. Even if the choice is made based on data - for example, choosing a few neighbours to the right and left of Brazil in the distribution of worker productivity, the next question naturally is, "how many?"

The clustering algorithm removes subjectivity from this step, but the results can be counterintuitive as to almost require a leap of faith. In this case, the list of countries that will be used in the control based on the t-SNE method are: Albania, Bosnia and Herzegovina, Belarus, Belize, Barbados, Colombia, Costa Rica, Cuba, Dominican Republic, Ecuador, Egypt, Fiji, Guyana, Iraq, Jamaica, Kazakhstan, Saint Lucia, Mauritius, Namibia, Paraguay, Palestine, Serbia, Eswatini, Tunisia, Saint Vincent and the Grenadines, Yemen. Figure 2 compares the outcome variable for these countries (grey) compares with Brazil's (red). A quick glance already indicates a major difference between the countries in Brazil's cluster and other countries, exemplified here by the cluster centres: the scale of their worker productivity. A natural question is whether scaling the worker productivity could have taken away these differences and let only the *dynamics* of worker productivity over time be reflected in cluster compositions. There is no general right answer here - in this case, I judge the scale of worker productivity to also be informative because it reflects deeper socio-economic issues that make these countries more relatable, even if at first the cluster would not be anyone's first subjective choice.[18] Plus, we see separately that the SC created from these countries creates an excellent match with actual Brazil and thus offers a valid counterfactual.

Figure 2: Pre-reform worker productivity in Brazil and other countries

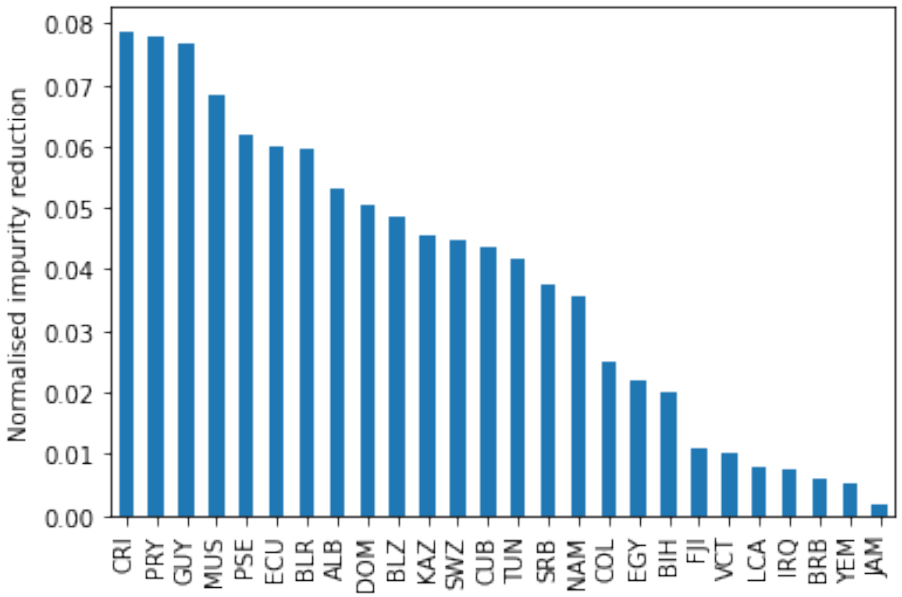


This list does not imply all countries contribute equally to estimating the syn-

[18]This has a paralell with the Harrod-Balassa-Samuelson hypothesis.

thetic version of Brazil. In fact, some might even end up not even contributing in the estimating equation. Figure 3 shows the empirical importance of each country to the estimation of pre-intervention synthetic Brazil - in terms of how much they contribute to creating informating "leafs" in the trees. The selected list also does not imply a causal explanation from any of those countries into the Brazilian dynamics. They are merely closest to Brazil in this clustering exercise, and as such, likely to be a good predictor at the same time period based on the interpretation that their pre-intervention *level* of productivity reflects the organisation of their economy.

Figure 3: Importance of control jurisdiction for estimation of synthetic Brazil



What to make of Figure 3? A first observation is that the jurisdictions are all selected based on their data, which provides comfort that the elements of $\hat{\mathbf{J}}$ are indeed the closest to Brazil as a group. A few controls that could be seen as "traditional" appear in the list, but others are more puzzling at first. To be sure, higher importance values do not mean that Brazil's labour productivity are *explained* by these countries, but simply that they tend to correlate better, including out-of-sample. And it is advisable to bear in mind that even when the importance attributed to an unit is low, they can still influence the estimated value for Brazil in meaningful ways, eg if they are used to determine values in either extreme. All in all, the best way to interpret this figure is as an attempt to order the contributing individuals for average values of the outcome variable.

As an example of the decision trees, Figure 4 presents a selected tree from the random forest $\hat{\omega}_Y$. Note from the top node that the tree considers only a subset

of time periods (ie, "samples" is lower than $T0 = 17$). Another regularisation mechanism of the random forests is that the tree tries to optimise prediction based only on a limited number of countries in $\hat{\mathbf{J}}$. Both sampling mechanisms (across the time and cross-section dimension) are standard in the random forest methodology and in fact contribute to its high adaptability across use cases. It is also worth highlighting that the the trees in the standard random forest algorithm do not have any causal interpretation, only the goal of minimising prediction error.

Figure 4: Selected tree (No 13) from synthetic control estimator

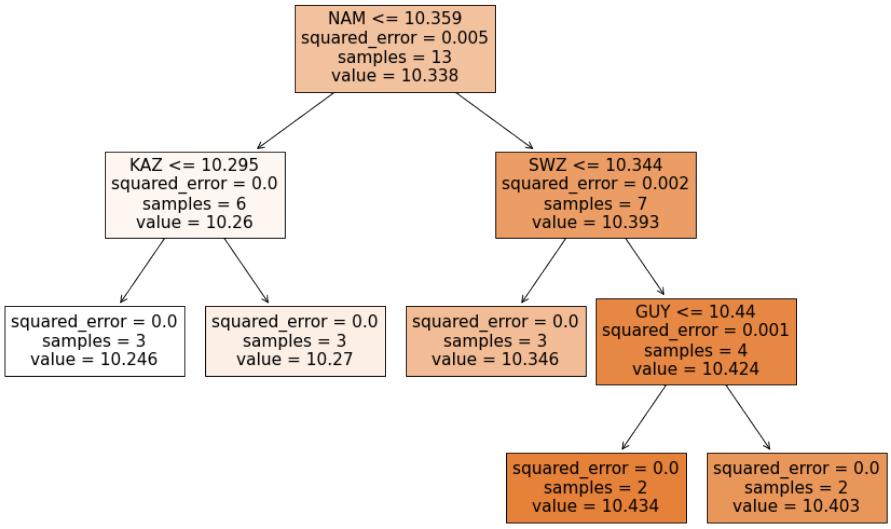


Figure 5 shows the evolution of worker productivity in Brazil before and after the labour reform law enters into force, comparing with synthetic Brazil. In order to check if the observed difference is meaningful, below I proceed with a placebo test on the other countries in the same cluster. Figure 6 shows the difference between the actual and the synthetic controls for Brazil (red lines) and the other countres (grey). The left panel has all countries in the cluster, and the right panel facilitates visualisation by removing two outliers Guyana and Yemen.

The result of the manifold learning underscores that the synthetic control seems indeed to come from a similar space in the data distribution as Brazil (Figure 7).

As mentioned above, the Euclidean distances from the embedding can be used to test whether countries occuring from the same underlying "distribution of countries" would be closer to Brazil than the machine control. In other words, it can be interpreted as a time of p-value: considering the distibution underlying

Figure 5: Labour productivity in Brazil

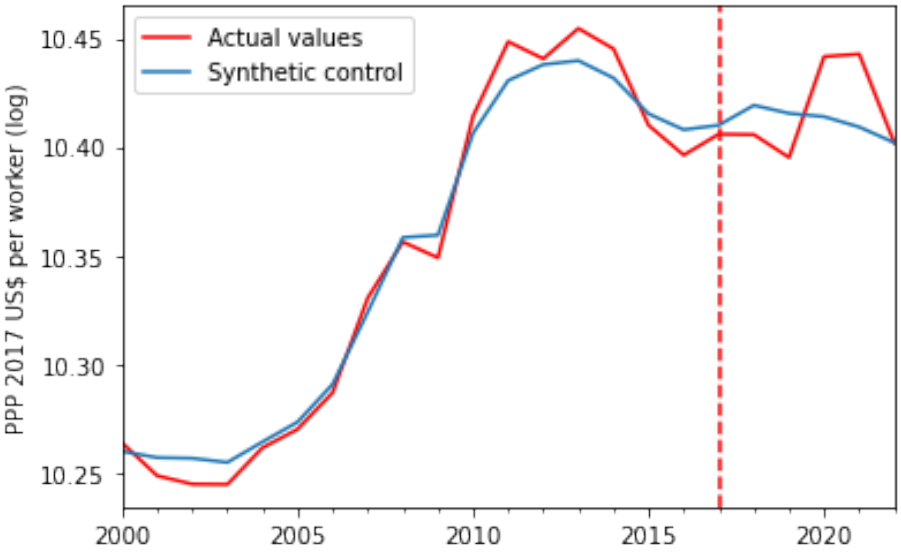


Figure 6: Effect of deregulation on labour productivity in Brazil

(a) All countries in same cluster (b) Except Guyana and Yemen (outliers)

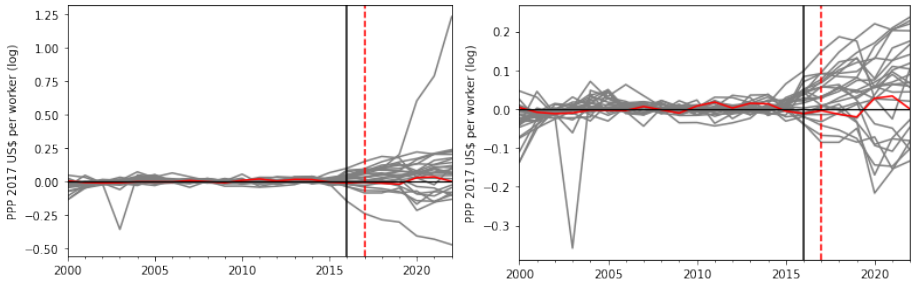
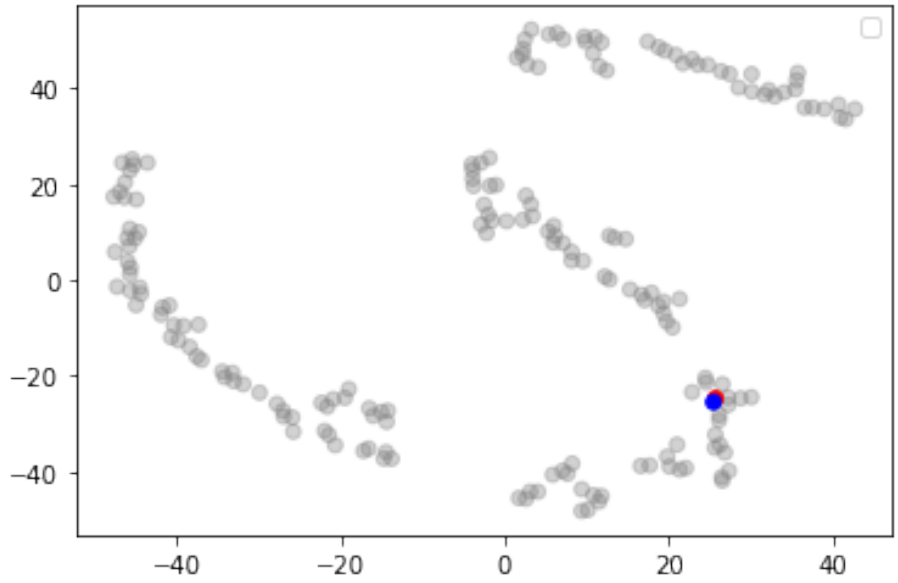


Figure 7: 2-dimensional representation of the pre-reform countries and control



the t-SNE test, it would be rare to spontaneously have a country that is closer to Brazil than the SC. For this application, an exercise with $m = 100$ yielded $w(j = \text{SC})/m = 1$.

The above results point to a null effect of the reforms on labour productivity, frustrating policymakers' expectations. This is not completely surprising considering the voluminous literature pointing to *negative* effects of labour market deregulation and reduction in unit wage costs on worker productivity. Reasons include the boon to less productive firms from exogenously lower unit wage costs, less incentives for human capital accumulation on the job (eg, workers might prefer to acquire more general, marketable skills rather than firm-specific knowledge), and lower trust or engagement levels between workers and firms. Vergeer and Kleinknecht (2010) and Vergeer et al. (2015) present results and more extensive references on these arguments. Relatedly, international comparisons find no effect from labour deregulations in GDP growth (Brancaccio, Garbellini, and Giammetti (2018)). The results in this paper and in these papers contrast with the more positive findings with respect to *other* structural reforms on labour productivity, such as from trade and financial sectors (Konte, Kouamé, and Mensah (2022)).

Further studying why the labour market deregulation failed to increase worker productivity is beyond the scope of this paper not the least because of limited availability of post-reform data that is not meaningfully disrupted by the pandemic. Still, I describe some potential channels as possible suggestions for

follow-up work. First, while the reform explicitly lowered wage-related costs, this might not have lowered labour market frictions. For example, information asymmetries on both sides of the employment relations might account for a persistence of frictions if employers think the marginal employee might be someone that is now in the market because it is more affordable for the previous employer to dismiss; and conversely for the employees the marginal employer is hiring after the reform because they previously were not accommodating to workers' rights. A second possibility is that the bottlenecks to worker productivity were not related to labour frictions or wage costs, but rather to deeper issues such as human capital, both from education gaps and from on-the-job learning when staying in the job is now less likely.

One way to test the first hypothesis is to leverage insights from the literature that discusses labour inputs, which can be seen as a combination of observed employment n_t and unobserved effort e_t (see Galí and Rens (2020) and references). In particular, Gali and van Rens' (2020) model of endogenous worker effort might offer a practical way to test whether labour market frictions are indeed smaller after the reform. Using log variables, note that GDP in a constant returns to scale model is the sum of labour inputs and total productivity, $y_t = (1-\alpha)(n_t + \psi e_t) + a_t$, where y_t, n_t are output and number of workers respectively, α measures the diminishing returns to labour, and a_t is the total factor productivity. In this case, worker productivity is:

$$y_t - n_t = -\alpha n_t + (1-\alpha)\psi e_t + a_t, \tag{8}$$

which in a real business cycle model with labour frictions and z_t a non-technological preference shock, results in the extreme cases of, respectively, no friction or infinite friction:

$$n_t = \beta a_t + z_t e_t = \beta a_t + z_t, \tag{9}$$

assigning $\beta = (1-\eta)$. In other words, using $0 \leq \phi \leq 1$ to denote labour frictions, $(1-\phi)n_t + \phi e_t = \beta a_t + z_t$. Therefore the predictions from Equations 8 and 9 are that labour productivity becomes less procyclical as labour market frictions become smaller.[19] Also, with less frictions it is more optimal to adjust labour through employment levels rather than effort.

So in future years, when more data is available outside the Covid-19 pandemic period, one way to test whether labour market frictions are actually lower after the reforms is to check if the (a) correlations between employment and GDP increased, while (b) the correlation of proxies for effort (such as labour accidents) with GDP decreased. If these results are confirmed, then it will be up for future research to explore why productivity did not respond to lower frictions. If not, then the challenge will be to understand what other labour market frictions continue to hold back productivity growth.

[19] See Equations (31) and (32) of Galí and Rens (2020).

5 Conclusions

This paper describes a more data-driven way to estimate SCs for causal analyses, leveraging machine learning techniques that are ideally suited to each step of the estimation process. All these methods benefit from time-tested existing implementations; I offer another one that proceeds with end-to-end estimation via the package gingado (Araujo (2023)). They benefit from considerable forecasting performance, matching their ability to analyse large datasets, flexible functional forms and widely available software implementations with the needs of a SC method for a wider range of empirical applications.

Of course, these individual techniques could also be applied in a "modular way", when a researcher is best served by only one or two of these steps. While this paper puts forward the combined use of clustering methods, supervised learning and manifold learning as a group, they can also be used individually if required by the application at hand. For example, if the population of control individuals is not large but there are various covariates for each unit, there is no reason to use clustering algorithms but manifold learning can still be useful to assess the SC quality. For illustration purposes, this paper shows their joint use in a relatively simple example. Machine controls might also be used "wholesale", seeking the identification of outlier movements or to flag entities that might require further scrutiny by a human analyst. For example, financial supervisors, who receive multiple data points from a given universe of supervised firms, might apply this method periodically to stake out any particular supervised entity as going through some idiosyncratic process (which might or might not be adverse). Ultimately, it can also be used for pure outlier detection because unlike other traditional outlier detection algorithms, machine controls can identify entities that would normally have reported values outside the typical range of the other entities in the same population, but who are actually abnormally reporting data that seems normal at a first glance.

The "machine control" is applied to the evaluate how worker productivity responds to the extensive 2017 labour market deregulation. Interestingly, I find no noiceable impact on productivity, frustrating one of the stated original intentions of the reforms. The difficulties in permanently raising productivity across the board in Brazil are structural and therefore would not be easy for labout deregulation to change, even after many years trying (World Bank Group (2018)). According to the BCB (2023), almost 80% of the growth in value added between 2000 and 2022 are attributable to the increase in labour supply rather than efficiency gains. The remainder 20% of growth (25% in the estimation of Veloso (2024)) are attributable to efficiencies generated by improving allocation of labour between different economic sectors: the agricultural sector contributed to allocative efficiency gains as less people worked moved to other sectors, leaving just the most effiencient workers in that area; conversely many activities in the services sector saw considerable employment growth and lower productivity, leading both to lower direct productivity and allocative efficiency. This is consistent with longer-term challenges in Latin American and African countries

in improving labour productivity through labour reallocation to more productive sectors (M. S. McMillan and Rodrik (2011), M. McMillan, Rodrik, and Verduzco-Gallo (2014)). Compounding the challenge, A. Campos et al. (2017) and references therein document that low productivity in Brazil is pervasive across sectors.

It is worrisome that even a substantial structural reform in the labour market was not sufficient to "move the needle" enough to increase efficiency, which calls into question whether the road ahead should be further deregulation of labour markets. For example, weaker unions (A. G. Campos and Silva (2023)) might not be good for worker protection, although there are questions about the pre-reform representativeness of unions (A. Campos et al. (2017)). Another possibility is that a lower level of labour protections might disincentivise marginal workers from remaining on the job, especially as they grow older or undergo other life events such as pregnancy, therefore putting further pressure on labour supply as an engine of growth. Conversely, the reform's effects might be latent and only accrue over time and with further reforms. Say, a more flexible labour market might increase the extent to which future structural changes in other sectors, such as trade, positively impact the contribution of intersectoral allocation of labour to productivity (Konte, Kouamé, and Mensah (2022)). Or, the labour reforms might be an example of the paradoxical results found after the many microeconomic reforms Brazil and other countries underwent in the last decades: higher firm- and sector-level productivity offset by worse allocation of workers, including into unemployment (M. McMillan, Rodrik, and Verduzco-Gallo (2014)). Another channel through which the reform can still impact worker productivity is by mitigating some of the strong incentives for workers to be dismissed as soon as they are eligible for benefits (A. Campos et al. (2017)).

I offer no precise answers here on whether the 2017 Brazil labour reform will ultimately prove a success, only a wake-up call that increasing average worker productivity, a key component of durable economic prosperity, will require considerably more work to reformulate the economy. This is especially the case when informality creeps up, as documented in the literature. Looking forward, the complexity of policymaking in this area is further increased by the still considerable uncertainty about the potential effects on labour markets of advances in artificial intelligence.

The Brazilian experience discussed in this paper might help countries with similar low productivity issues formulate and prioritise their own reforms. And I hope the methods herein introduced will support more analysis of policy impacts in a more flexible and data-driven way.

References

Abadie, Alberto. 2021. "Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects." *Journal of Economic Literature* 59 (2):

391–425. https://doi.org/10.1257/jel.20191450.

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program." *Journal of the American Statistical Association* 105 (490): 493–505.

———. 2015. "Comparative Politics and the Synthetic Control Method." *American Journal of Political Science* 59 (2): 495–510. https://doi.org/https://doi.org/10.1111/ajps.12116.

Abadie, Alberto, and Javier Gardeazabal. 2003. "The Economic Costs of Conflict: A Case Study of the Basque Country." *American Economic Review* 93 (1): 113–32. https://doi.org/10.1257/000282803321455188.

Abadie, Alberto, and Jérémy L'hour. 2021. "A Penalized Synthetic Control Estimator for Disaggregated Data." *Journal of the American Statistical Association* 116 (536): 1817–34.

Adascalitei, Dragos, and Clemente Pignatti Morano. 2016. "Drivers and Effects of Labour Market Reforms: Evidence from a Novel Policy Compendium." *IZA Journal of Labor Policy* 5: 1–32.

Addison, John T., Claus Schnabel, and Joachim Wagner. 2001. "Works Councils in Germany: Their Effects on Establishment Performance." *Oxford Economic Papers* 53 (4): 659–94. http://www.jstor.org/stable/3488608.

Araujo, Douglas KG. 2023. "Gingado: A Machine Learning Library Focused on Economics and Finance." Working Paper 1122. BIS Working Paper. Bank for International Settlements.

Arora, Sanjeev, Wei Hu, and Pravesh K Kothari. 2018. "An Analysis of the t-SNE Algorithm for Data Visualization." *Proceedings of Machine Learning Research Vol* 75: 1–8.

Athey, Susan. 2018. "The Impact of Machine Learning on Economics." In *The Economics of Artificial Intelligence: An Agenda*, 507–47. University of Chicago Press.

Athey, Susan, Mohsen Bayati, Guido Imbens, and Zhaonan Qu. 2019. "Ensemble Methods for Causal Effects in Panel Data Settings." *AEA Papers and Proceedings* 109 (May): 65–70. https://doi.org/10.1257/pandp.20191069.

Athey, Susan, and Guido W Imbens. 2019. "Machine Learning Methods That Economists Should Know About." *Annual Review of Economics* 11: 685–725.

Athey, Susan, and Stefan Wager. 2021. "Policy Learning with Observational Data." *Econometrica* 89 (1): 133–61. https://doi.org/https://doi.org/10.3982/ECTA15732.

Banco Central do Brasil. 2023. "Relatório de Inflação."

Bank for International Settlements. 2024. "Central Bank Speeches, 1996-2024." https://www.bis.org/cbspeeches/download.htm.

Biljanovska, Nina, and Damiano Sandri. 2018. "Structural Reform Priorities for Brazil." IMF Working Paper WP/18/224. International Monetary Fund.

Brancaccio, Emiliano, Fabiana De Cristofaro, and Raffaele Giammetti. 2020. "A Meta-Analysis on Labour Market Deregulations and Employment Performance: No Consensus Around the IMF-OECD Consensus." *Review of*

Political Economy 32 (1): 1–21.

Brancaccio, Emiliano, Nadia Garbellini, and Raffaele Giammetti. 2018. "Structural Labour Market Reforms, GDP Growth and the Functional Distribution of Income." *Structural Change and Economic Dynamics* 44: 34–45.

Breiman, Leo. 2001. "Random Forests." *Machine Learning* 45: 5–32.

Bühlmann, Peter, and Bin Yu. 2002. "Analyzing Bagging." *The Annals of Statistics* 30 (4): 927–61.

Campos, André Gambier, and Sandro Pereira Silva. 2023. "Impactos Estruturais Da Reforma Trabalhista de 2017 Sobre Sindicatos de Trabalhadores No Brasil."

Campos, André, Carlos Henrique L Corseuil, Miguel N Foguel, and Hélio Zylberstajn. 2017. "Instituições Trabalhistas e Produtividade Do Trabalho: Uma análise Do Caso Brasileiro." *Rio de Janeiro: Oficina de Trabalho Realizada Nas Dependências Do IPEA*.

Carvalho, Sandro Sacchet de. 2017. "Uma Visão Geral Sobre a Reforma Trabalhista."

Chen, Jiafeng. 2023. "Synthetic Control as Online Linear Regression." *Econometrica* 91 (2): 465–91.

Chen, Tianqi, and Carlos Guestrin. 2016. "Xgboost: A Scalable Tree Boosting System." In *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, 785–94.

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. 2018. "Double/Debiased Machine Learning for Treatment and Structural Parameters: Double/Debiased Machine Learning." *The Econometrics Journal* 21 (1).

Chernozhukov, Victor, Kaspar Wüthrich, and Yinchu Zhu. 2021. "An Exact and Robust Conformal Inference Method for Counterfactual and Synthetic Controls." *Journal of the American Statistical Association* 116 (536): 1849–64.

Doudchenko, Nikolay, and Guido W Imbens. 2016. "Balancing, Regression, Difference-in-Differences and Synthetic Control Methods: A Synthesis." National Bureau of Economic Research.

Eslava, Marcela, John Haltiwanger, Adriana Kugler, and Maurice Kugler. 2004. "The Effects of Structural Reforms on Productivity and Profitability Enhancing Reallocation: Evidence from Colombia." *Journal of Development Economics* 75 (2): 333–71.

Evdokimova, Tatiana, Piroska Nagy-Mohacsi, Olga Ponomarenko, and Elina Ribakova. 2023. "Central Banks and Policy Communication: How Emerging Markets Have Outperformed the Fed and ECB." *Peterson Institute for International Economics Working Paper*, no. 23-10.

Fama, Eugene F, and Kenneth R French. 1993. "Common Risk Factors in the Returns on Stocks and Bonds." *Journal of Financial Economics* 33 (1): 3–56.

Farrell, Max H, Tengyuan Liang, and Sanjog Misra. 2021. "Deep Neural Networks for Estimation and Inference." *Econometrica* 89 (1): 181–213.

Ferman, Bruno. 2021. "On the Properties of the Synthetic Control Estimator

with Many Periods and Many Controls." *Journal of the American Statistical Association* 116 (536): 1764–72.

Ferman, Bruno, Cristine Pinto, and Vitor Possebom. 2020. "Cherry Picking with Synthetic Controls." *Journal of Policy Analysis and Management* 39 (2): 510–32. https://doi.org/https://doi.org/10.1002/pam.22206.

Fisher, Eric O'N, John Gilbert, Kathryn G Marshall, and Reza Oladi. 2015. "A New Measure of Economic Distance."

Frey, Brendan J, and Delbert Dueck. 2007. "Clustering by Passing Messages Between Data Points." *Science* 315 (5814): 972–76.

Fung, William, and David A. Hsieh. 2004. "Hedge Fund Benchmarks: A Risk-Based Approach." *Financial Analysts Journal* 60 (5): 65–80. http://www.jstor.org/stable/4480604.

Galí, Jordi, and Thijs van Rens. 2020. "The Vanishing Procyclicality of Labour Productivity." *The Economic Journal* 131 (633): 302–26. https://doi.org/10.1093/ej/ueaa065.

Gobillon, Laurent, and Thierry Magnac. 2016. "Regional Policy Evaluation: Interactive Fixed Effects and Synthetic Controls." *Review of Economics and Statistics* 98 (3): 535–51.

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning*. MIT press.

Hinton, Geoffrey E, and Sam Roweis. 2002. "Stochastic Neighbor Embedding." *Advances in Neural Information Processing Systems* 15.

Hotelling, Harold. 1933. "Analysis of a Complex of Statistical Variables into Principal Components." *Journal of Educational Psychology* 24 (6): 417.

Imbens, Guido W, and Donald B Rubin. 2015. *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press.

Kaul, Ashok, Stefan Klößner, Gregor Pfeifer, and Manuel Schieler. 2015. "Synthetic Control Methods: Never Use All Pre-Intervention Outcomes Together with Covariates."

Ke, Guolin, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. 2017. "Lightgbm: A Highly Efficient Gradient Boosting Decision Tree." *Advances in Neural Information Processing Systems* 30.

Klößner, Stefan, Ashok Kaul, Gregor Pfeifer, and Manuel Schieler. 2018. "Comparative Politics and the Synthetic Control Method Revisited: A Note on Abadie Et Al.(2015)." *Swiss Journal of Economics and Statistics* 154: 1–11.

Konte, Maty, Wilfried A Kouamé, and Emmanuel B Mensah. 2022. "Structural Reforms and Labor Productivity Growth in Developing Countries: Intra or Inter-Reallocation Channel?" *The World Bank Economic Review* 36 (3): 646–69.

Kouamé, Wilfried AK, and Sampawende J-A Tapsoba. 2019. "Structural Reforms and Firms' Productivity: Evidence from Developing Countries." *World Development* 113: 157–71.

Krawczun, Lopes, Natalia Branco, Magno Rogério Gomes, and Solange de Cassia Inforzato de Souza. 2020. "Reforma Trabalhista e Desigualdade de gênero No Brasil: Uma Perspectiva Jurídica e Econômica." *Revista*

Brasileira de Políticas Públicas 10 (2).

Lima, D., M. Wilbert, and A. Silva. 2021. "Impacto Da Reforma Trabalhista Na Informalidade e Nas Contas Previdenciárias Do Brasil." *RTPS - Revista Trabalho, Política e Sociedade* 6 (10): 291–316.

Lima Júnior, Arnaldo Barbosa, Daniel Cavalcante, and Igor Parente Pinto. 2016. "O Fortalecimento Das Negociações Coletivas é Indispensável Para o Crescimento Da Produtividade Do Trabalho No Brasil."

McMillan, Margaret S, and Dani Rodrik. 2011. "Globalization, Structural Change and Productivity Growth." National Bureau of Economic Research.

McMillan, Margaret, Dani Rodrik, and Íñigo Verduzco-Gallo. 2014. "Globalization, Structural Change, and Productivity Growth, with an Update on Africa."

Mézard, Marc. 2003. "Passing Messages Between Disciplines." *Science* 301 (5640): 1685–86.

Nie, Xinkun, and Stefan Wager. 2021. "Quasi-Oracle Estimation of Heterogeneous Treatment Effects." *Biometrika* 108 (2): 299–319.

Prokhorenkova, Liudmila, Gleb Gusev, Aleksandr Vorobev, Anna Veronika Dorogush, and Andrey Gulin. 2018. "CatBoost: Unbiased Boosting with Categorical Features." *Advances in Neural Information Processing Systems* 31.

Quistorff, Brian, Matthew Goldman, and Jason Thorpe. 2020. "Sparse Synthetic Controls: Unit-Level Counterfactuals from High-Dimensional Data." In *Microsoft Journal of Applied Research*, 14:155–70.

Rey, Hélène. 2015. "Dilemma Not Trilemma: The Global Financial Cycle and Monetary Policy Independence." Working Paper 21162. Working Paper Series. National Bureau of Economic Research. https://doi.org/10.3386/w21162.

Robinson, Peter M. 1988. "Root-n-Consistent Semiparametric Regression." *Econometrica: Journal of the Econometric Society*, 931–54.

Rosenbaum, Paul R., and Donald B. Rubin. 1983. "The central role of the propensity score in observational studies for causal effects." *Biometrika* 70 (1): 41–55. https://doi.org/10.1093/biomet/70.1.41.

Saint-Paul, Gilles. 2002. "The Political Economy of Employment Protection." *Journal of Political Economy* 110 (3): 672–704.

Serra, Gustavo Pereira, Ana Bottega, and Marina Da Silva Sanches. 2022. "A Reforma Trabalhista de 2017 Teve Efeito Sobre a Taxa de Desemprego No Brasil? Uma análise Dos Primeiros Anos de Vigência Da Lei 13.467/2017." *Nota de Política Econômica* 21.

Silva, Sandro Pereira. 2018. "A Estratégia Argumentativa Da Reforma Trabalhista No Brasil à Luz de Dados Internacionais."

Tibshirani, Robert. 1996. "Regression Shrinkage and Selection via the Lasso." *Journal of the Royal Statistical Society: Series B (Methodological)* 58 (1): 267–88.

Van Der Maaten, Laurens. 2009. "Learning a Parametric Embedding by Preserving Local Structure." In *Artificial Intelligence and Statistics*, 384–91. PMLR.

Van der Maaten, Laurens, and Geoffrey Hinton. 2008. "Visualizing Data Using t-SNE." *Journal of Machine Learning Research* 9 (11).

Veloso, Fernando. 2024. "Productivity and Growth in Brazil." *FGV IBRE Articles*. Online; Equipe FGV IBRE.

Vergeer, Robert, Steven Dhondt, Alfred Kleinknecht, and Karolus Kraan. 2015. "Will 'Structural Reforms' of Labour Markets Reduce Productivity Growth? A Firm-Level Investigation." *European Journal of Economics and Economic Policies* 12 (3): 300–317.

Vergeer, Robert, and Alfred Kleinknecht. 2010. "The Impact of Labor Market Deregulation on Productivity: A Panel Data Analysis of 19 OECD Countries (1960-2004)." *Journal of Post Keynesian Economics* 33 (2): 371–408.

Viviano, Davide, and Jelena Bradic. 2022. "Synthetic Learner: Model-Free Inference on Treatments over Time." *Journal of Econometrics*.

Wager, Stefan, and Susan Athey. 2018. "Estimation and Inference of Heterogeneous Treatment Effects Using Random Forests." *Journal of the American Statistical Association* 113 (523): 1228–42.

Wattenberg, Martin, Fernanda Viégas, and Ian Johnson. 2016. "How to Use t-SNE Effectively." *Distill*. https://doi.org/10.23915/distill.00002.

World Bank Group. 2018. "Emprego e Crescimento: A Agenda Da Produtividade." Washington, D.C.: World Bank Group.

Xu, Yiqing. 2017. "Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models." *Political Analysis* 25 (1): 57–76.

Yeh, Christopher, Anthony Perez, Anne Driscoll, George Azzari, Zhongyi Tang, David Lobell, Stefano Ermon, and Marshall Burke. 2020. "Using Publicly Available Satellite Imagery and Deep Learning to Understand Economic Well-Being in Africa." *Nature* 11. https://doi.org/https://doi.org/10.1038/s41467-020-16185-w.

Zou, Hui, and Trevor Hastie. 2005. "Regularization and Variable Selection via the Elastic Net." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67 (2): 301–20. https://doi.org/https://doi.org/10.1111/j.1467-9868.2005.00503.x.

Previous volumes in this series

1180 April 2024	Incentive-Compatible Unemployment Reinsurance for the Euro Area	Alexander Karaivanov, Benoit Mojon, Luiz Pereira da Silva, Albert Pierres Tejada and Robert M Townsend
1179 April 2024	The impact of artificial intelligence on output and inflation	Iñaki Aldasoro, Sebastian Doerr, Leonardo Gambacorta and Daniel Rees
1178 April 2024	Finternet: the financial system for the future	Agustín Carstens and Nandan Nilekani
1177 March 2024	Pre-publication revisions of bank financial statements: a novel way to monitor banks?	Andre Guettler, Mahvish Naeem, Lars Norden and Bernardus F Nazar Van Doornik
1176 March 2024	The effect of Covid pension withdrawals and the Universal Guaranteed Pension on the income of future retirees in Chile	Carlos Madeira
1175 March 2024	Unmitigated disasters? Risk- sharing and macroeconomic recovery in a large international panel	Goetz von Peter, Sebastian von Dahlen, and Sweta Saxena
1174 March 2024	The impact of information and communication technologies on banks, credit and savings: an examination of Brazil	Flavia Alves
1173 March 2024	The macroprudential role of central bank balance sheets	Egemen Eren, Timothy Jackson and Giovanni Lombardo
1172 March 2024	Navigating by falling stars: monetary policy with fiscally driven natural rates	Rodolfo G Campos, Jesús Fernández-Villaverde, Galo Nuño and Peter Paz
1171 March 2024	DeFi Leverage	Lioba Heimbach and Wenqian Huang
1170 March 2024	Monetary Policy Transmission in Emerging Makerts: Proverbial Concerns, Novel Evidence	Ariadne Checo, Francesco Grigoli, and Damiano Sandri
1169 February 2024	Risk-based pricing in competitive lending markets	Carola Müller, Ragnar E. Juelsrud, Henrik Andersen
1168 February 2024	Corporate payout policy: are financial firms different?	Emmanuel Caiazzo, Leonardo Gambacorta, Tommaso Oliviero and Hyun Song Shin
1167 February 2024	Monetary Policy with Profit-Driven Inflation	Enisse Kharroubi and Frank Smets

All volumes are available on our website www.bis.org.