On "Imputation of Counterfactual Outcomes when the Errors are Predictable": Discussions on Misspecification and Suggestions of Sensitivity Analyses

Luis A. F. Alvarez* Bruno Ferman†

March 2024

Abstract

Gonçalves and Ng (2024) propose an interesting and simple way to improve counterfactual imputation methods when errors are predictable. For unconditional analyses, this approach yields smaller mean-squared error and tighter prediction intervals in large samples, even if the dependence of the errors is misspecified. For conditional analyses, this approach corrects the bias of standard methods, and provides valid asymptotic inference, if the dependence of the errors is correctly specified. In this comment, we first discuss how the assumptions imposed on the errors depend on the model and estimator adopted. This enables researchers to assess the validity of the assumptions imposed on the structure of the errors, and the relevant information set for conditional analyses. We then propose a simple sensitivity analysis in order to quantify the amount of misspecification on the dependence structure of the errors required for the conclusions of conditional analyses to be changed.

1 Introduction

There is a large literature on the estimation of causal effects in panel data settings (see Arkhangelsky and Imbens (2023) for a recent survey). In such settings, the time series and cross-section of the errors may provide useful information to construct counterfactuals. However, the use of such information has been largely overlooked in this literature. We

*Department of Economics, University of São Paulo. E-mail address: luis.alvarez@usp.br

†São Paulo School of Economics, FGV. E-mail address: bruno.ferman@fgv.br

congratulate Silvia Gonçalves and Serena Ng for proposing an interesting and simple way to include this information into the analysis.

The correction proposed by Gonçalves and Ng (2024) can be seen as an add-on to standard methods in this literature, and is simple to implement. When we consider unconditional analysis, this correction (asymptotically) provides gains in terms of lower mean squared errors and more powerful tests at virtually no cost. This is true even if the dependence of the errors is misspecified. When we analyze the properties of the estimators conditionally on the relevant information (such as the realization of the errors of the pre-treatment periods), then standard estimators will generally be biased, and we may have inference distortions. The add-on proposed by Gonçalves and Ng (2024) corrects these problems, if we use the correct specification for the dependence of the errors. However, we cannot guarantee conditionally unbiased estimators and tests with correct sizes if the dependence of the errors is misspecified.

Overall, considering the possible combinations of scenarios, depending on whether the analysis is unconditional vs conditional, and whether the structure of the errors is correctly vs incorrectly specified, the add-on proposed by Gonçalves and Ng (2024) provides improvements in three out of four scenarios. In one of them, when we consider conditional analysis with misspecified dependence of the errors, it is not possible to guarantee that the add-on proposed by Gonçalves and Ng (2024) provides valid asymptotic conditional inference or, more generally, improvements relative to standard methods that do not use this correction. Importantly, standard methods without the correction proposed by Gonçalves and Ng (2024) would also fail to provide unbiased estimators and valid asymptotic inference in this scenario.

In this note, we first discuss how the definition of the error term in the setting proposed by Gonçalves and Ng (2024) depends on the specific model and estimator that is analyzed. This may be relevant in evaluating the information set we should condition on, and the assumptions we impose on the structure of the errors. For example, we show that there are some settings in which it is reasonable to consider the properties of the estimator conditional on only pre-treatment errors, and some settings in which it would be more reasonable to also condition on other variables, such as pre-treatment outcomes, and post-treatment outcomes of the controls. In such settings, if we consider a conditional analysis on all relevant information that may help predict counterfactual outcomes, then a correction using only information on pre-treatment errors would be misspecified. Moreover, we also discuss the possibility of misspecification even when we are conditioning on the correct variables. Given the possibility of misspecification, we then propose a simple sensitivity analysis to quantify the amount of misspecification on the dependence structure required for the conclusions of conditional analyses to be changed.

2 Setting

In line with Gonçalves and Ng (2024), we consider a setting with T periods, where treatment starts after period $T_0 < T$. We have N units. For ease of exposition, we consider the case in which only unit $i = 1$ is treated, so the number of control units is $N_0 = N - 1$. We let $Y_{i,t}(0)$ and $Y_{i,t}(1)$ be the potential outcomes for unit i at time t. Therefore, we observe $Y_{i,t} = Y_{i,t}(1)$ for unit $i = 1$ at $t > T_0$, and $Y_{i,t} = Y_{i,t}(0)$ for all other cases. The goal is to estimate $\delta_{1,T_0+h} = Y_{1,T_0+h}(1) - Y_{1,T_0+h}(0)$, for some $h \in [1, T - T_0] \cap \mathbb{N}$. The main idea is to estimate a counterfactual $\hat{Y}_{1,T_0+h}(0)$, which will in turn lead to an estimator for the treatment effect $\hat{\delta}_{1,T_0+h} = Y_{1,T_0+h}(1) - \hat{Y}_{1,T_0+h}(0)$.

Gonçalves and Ng (2024) consider a model for the potential outcome $Y_{1,t}(0) = m_{1,t} + e_{1,t}$, where it is assumed that $\mathbb{E}[e_{1,t}] = 0$, and $m_{1,t}$ is a pseudo-true conditional mean. Standard methods in this literature generally construct the counterfactual for $\hat{Y}_{1,T_0+h}(0)$ as an estimator for m_{1,T_0+h}, yielding the estimator for the treatment effects $\hat{\delta}_{1,T_0+h} = Y_{1,T_0+h} - \hat{m}_{1,T_0+h}$. Gonçalves and Ng (2024) propose a correction to take the predictability of the errors into account, by including a correction term $\hat{\delta}^{+}_{1,T_0+h} = \hat{\delta}_{1,T_0+h} - \hat{\rho}_h \hat{e}_{1,T_0}$, where $\hat{e}_{1,t}$ is the residual $Y_{1,t} - \hat{m}_{1,t}$, while $\hat{\rho}_h$ is the least squares slope of a regression of $\hat{e}_{1,t}$ on $\hat{e}_{1,t-h}$ in the pre-treatment periods (they also consider alternative forms of corrections).

3 Defining the error term

Importantly, $m_{1,t}$ is implicitly defined as the probability limit of the counterfactual estimator $\hat{m}_{1,t}$, and this in turn defines the error term $e_{1,t}$, which will depend on the assumptions on the potential outcomes and on the estimator used to construct the counterfactual. This modeling is very interesting by itself, because, for a given setting and estimator, we can derive the associated error term $e_{1,t}$, and analyze the restrictions we need to impose so that the assumption $\mathbb{E}[e_{1,t}] = 0$ is satisfied. This would guarantee that, for unconditional analysis, the original estimator (without the correction) is asymptotically unbiased, and that, under stationarity and weak dependence assumptions, it would be possible to derive asymptotically valid inference methods (for examples, see Chernozhukov et al. (2021) and Section 4). Moreover, an analysis of the associated error term $e_{1,t}$ is crucial to understand the assumptions for the conditional analysis considered by Gonçalves and Ng (2024).

For example, suppose potential outcomes $Y_{i,t}(0)$ follow a linear factor model, as considered by Ferman and Pinto (2021) and Ferman (2021),

$$Y_{i,t}(0) = c_i + \gamma_t + \lambda_t \mu_i + \varepsilon_{i,t}, \tag{1}$$

where c_i and δ_t are unit- and time-invariant fixed effects, λ_t is an $1 \times F$ vector of unobserved common factors, μ_i is an $F \times 1$ vector of unknown factor loadings, and $\{\epsilon_{i,t}\}$ are unobserved idiosyncratic shocks (which are independent across i, and independent of the factor structure $\{\lambda_t \mu_i\}$).

In this case, depending on the proposed estimator and the characteristics of the setting, we would have a different pseudo-true conditional mean and, consequently, we would have a different associated $e_{1,t}$. If we consider the standard SC estimator in a setting with $N_0 \to \infty$ and $T_0 \to \infty$, then under the assumptions considered by Ferman (2021), we have $\hat{m}_{1,t} \overset{p}{\to} c_1 + \gamma_t + \lambda_t \mu_1$. Therefore, we have that $e_{1,t} = \varepsilon_{1,t}$ in this setting. Interestingly, note that, in this case, the condition $\mathbb{E}[e_{1,t}] = 0$ does not impose restrictions on the factor model structure, so that we may have selection into treatment based on the factor model structure, as discussed by Ferman (2021). However, this restriction implies that we cannot have selection into treatment based on the idiosyncratic shocks $\varepsilon_{i,t}$. Moreover, for the purposes of Gonçalves and Ng (2024), we have that, conditional on the past errors, information on past outcomes would not provide any additional information on the prediction error of $\hat{\delta}_{1,T_0+h}$. Likewise, under the assumption that $\{\varepsilon_{i,t}\}$ is independent in the cross-section, post-treatment errors of the controls would also not provide relevant information to construct the counterfactual of Y_{1,T_0+h}. Therefore, it would be reasonable to condition the analysis on only the pre-treatment errors. This would also be the case in other settings/estimators, such as those considered by Arkhangelsky et al. (2021).

In contrast, if we consider a setting in which $T_0 \to \infty$, but N_0 is fixed, as considered by Ferman and Pinto (2021), then $e_{1,t}$ would depend on how we construct the estimator for the counterfactual of $Y_{1,T_0+h}(0)$. For example, if we consider a demeaned SC estimator (Ferman and Pinto, 2021; Doudchenko and Imbens, 2017), then $e_{1,t} = \varepsilon_{1,t} - \sum_{i=2}^{N} \bar{w}_i \varepsilon_{i,t} + \lambda_t \left(\mu_1 - \sum_{i=2}^{N} \bar{w}_i \mu_i \right)$, where $(\bar{w}_2, ..., \bar{w}_N)$ will generally be such that $\mu_1 \neq \sum_{i=2}^{N} \bar{w}_i \mu_i$. In this case, the assumption $\mathbb{E}[e_{1,t}] = 0$ has a different economic meaning, as it would essentially require not only $\mathbb{E}[\varepsilon_{i,t}] = 0$, but also $\mathbb{E}[\lambda_t] = 0$. This essentially means that we cannot have selection on time-varying unobservables (see Ferman and Pinto (2021) for details). For the purposes of Gonçalves and Ng (2024), we note that, in this case, pre-treatment outcomes would be informative about the post-treatment errors, even after we condition on pre-treatment errors. In this case, a simple correction that relies on, for example, an AR(1) model for the errors (if this is correctly specified) would provide asymptotically unbiased estimators and valid inference, conditional on *only* pre-treatment errors. However, this would not be the case once we condition on all available pre-treatment information. Likewise, this structure of the errors also means that post-treatment errors of the treated would be correlated with the post-treatment errors and outcomes of the controls. Therefore, if we consider an analysis conditional also on post-treatment information of the controls, then we should also take that into account. If we consider the original SC estimator (instead of its demeaned version), we would have a different associated error, but similar conclusions would apply.

Overall, we note that the properties of the error term $e_{1,t}$ depend crucially on the setting we analyze and on the estimator we use. Understanding the properties of this error term is crucial when we want to consider conditional analysis, as this would inform us what are the relevant information set we should condition on. For example, if we are considering the setting and assumptions from Ferman (2021), then we would only have to condition on pre-treatment errors. In contrast, if we are considering the setting and assumptions from Ferman and Pinto (2021), then we might also have to condition on pre-treatment outcomes, and post-treatment outcomes of the controls (in addition to the pre-treatment errors). Moreover, we also note that the characteristics of $e_{1,t}$ would also determine which kind of assumptions are reasonable to assume on the cross-section and serial dependence of $e_{1,t}$.

4 Sensitivity Analysis

The discussion from Section 2 shows that, in some settings, it is reasonable to condition only on past errors $e_{1,t}$. This would be the case, for example, when the estimator for the counterfactual, $\hat{Y}_{1,T_0+h}(0)$ asymptotically recovers all the systematic part of $Y_{1,T_0+h}(0)$, so the estimation error reflects only idiosyncratic shocks of the treated unit. In several settings, however, this would not be the case, and $\mathbb{E}[e_{1,T_0+h}|\mathcal{H}]$, where $\mathcal{H}$ is the set of relevant information we are conditioning on, should not be a simple function $\rho_h e_{1,T_0}$. In other cases, we may even consider a setting in which the relevant information set $\mathcal{H}$ is actually only the pre-treatment errors, but the conditional expectation $\mathbb{E}[e_{1,T_0+h}|\mathcal{H}]$ does not follow a simple linear function of only e_{1,T_0} (as would be the case if $e_{1,t}$ follows an AR(1) process). Yet another possibility is a setting with conditional heterokesdaticity, in which correctly accounting for the conditional mean yields a conditionally unbiased estimator, but this would not be sufficient to ensure valid conditional tests. We do not focus on this last case in the main text, but we discuss the possibility of conducting sensitivity analysis in such settings in Appendix B.

For these settings in which we consider a simple parametric correction $\rho_h e_{1,T_0}$, but we have misspecification $\mathbb{E}[e_{1,T_0+h}|\mathcal{H}] \neq \rho_h e_{1,T_0}$ (for example, because other variables are relevant for this conditional mean, or because we need a more complex parametrization of past errors), then we would be in a scenario in which Gonçalves and Ng (2024) do not offer guarantees in terms of asymptotic conditional unbiasedness and conditionally valid inference. One alternative in this case would be to consider a more flexible specification for $\mathbb{E}[e_{1,T_0+h}|\mathcal{H}]$. For example, by letting it be a function of more variables (such as the pre-treatment outcomes, in addition to the pre-treatment errors), and/or by allowing for more flexible functional forms. The addition of more variables in the correction is mentioned in Section 4.1 of Gonçalves and Ng (2024). However, we note that both alternatives would make the correction substantially more complex, and would require more of the data.

We consider instead a sensitivity analysis with respect to the approximation error of the conditional expectation function, $|\mathbb{E}[e_{1,T_0+h}|\mathcal{H}] - \rho_h e_{1,T_0}|$. Suppose the researcher believes there exists a constant $\Delta \geq 0$ that bounds the misspecification error almost-surely, i.e. $|\mathbb{E}[e_{1,T_0+h}|\mathcal{H}] - \rho_h e_{1,T_0}| \leq \Delta$ almost surely. We seek to find lower bounds for the misspecification degree Δ that would revert the conclusions of a conditional analysis.

In order to present the sensitivity analysis in more generality, we assume the researcher uses a resampling procedure to approximate the unconditional distribution of $e_{1,T_0+h} - \rho_h e_{1,T_0}$, which we denote by F_+. While the results derived by Gonçalves and Ng (2024) assume Gaussianity for simplicity, they also mention in their Section 5 the possibility of relaxing this assumption using resampling procedures. Specifically, we assume the researcher estimates F_+ by:

$$\hat{F}_+(c) = \frac{1}{T_0 - h} \sum_{t=1}^{T_0-h} \mathbf{1}\{\hat{e}_{1,T_0+h} - \hat{\rho}_h \hat{e}_{1,T_0} \leq c\}, \quad c \in \mathbb{R}\,. \tag{2}$$

This estimator will be consistent provided that $\{e_{1,t}\}_t$ is stationary and weakly dependent (as assumed by Gonçalves and Ng (2024)), and under mild conditions on the estimated pseudo-mean (e.g. Chernozhukov et al. (2021) or Alvarez and Ferman (2023)). If the distribution of the pair $(e_{1,t}, e_{1,t+h})$ is continuous and the estimator $\hat{F}_+$ is pointwise consistent to F_+, then the prediction interval

$$[\hat{\delta}^+_{1,T_0+h} - Q_{\hat{F}_+}(1-\alpha/2), \hat{\delta}^+_{1,T_0+h} - Q_{\hat{F}_+}(\alpha/2)]\,, \tag{3}$$

will asymptotically cover δ_{1,T_0+h} with unconditional probability $(1-\alpha)$, where Q_H denotes the quantile function of a distribution function H (see Appendix A). Therefore, these prediction intervals would be valid for unconditional analysis, even under misspecification of $\mathbb{E}[e_{1,T_0+h}|\mathcal{H}]$. However, with misspecification of the correction term, such prediction intervals may not be valid for conditional analysis.

For the sensitivity analysis, we assume a location model for the conditional distribution, i.e. that $e_{1,T_0+h} - \mathbb{E}[e_{1,T_0+h}|\mathcal{H}]$ is independent of $\mathcal{H}$. This assumption implies that it would suffice to correctly account for the conditional mean $\mathbb{E}[e_{1,T_0+h}|\mathcal{H}]$ in order to construct a conditionally valid test. This is the case in (nonlinear) autorregressive models for $e_{1,t}$ with additive innovations that are independent from $\mathcal{H}$, an example of which is the linear AR(1) model assumed in Section 5.1 of Gonçalves and Ng (2024) with the choice $\mathcal{H} = \sigma(e_{1,T_0}, e_{1,T_0-1}, \ldots)$. We note however that this assumption precludes conditional heteroskedasticity. In Appendix B, we show how sensitivity analysis can be performed in a model that allows for both misspecification of the conditional mean as well as conditional heterokedasticity. This extension can also be considered for sensitivity analysis in unconditional inference when one believes that the unconditional variance of $e_{1,t}$ changes after T_0. In such case, the resampling procedure we outline (as well as other procedures, such as those proposed by Chernozhukov et al. (2021)) would be invalid even for unconditional analysis. Therefore, the sensitivity analyses we propose would be an alternative for these settings as well. These analyses can be conducted using either the estimator with the correction proposed by Gonçalves and Ng (2024) or without it.

Under the assumption that $e_{1,T_0+H} - \mathbb{E}[e_{1,T_0+h}|\mathcal{H}]$ is independent of $\mathcal{H}$, we are able to show that, for any $s \in \mathbb{R}$:

$$P[e_{1,T_0+h} - \rho_h e_{1,T_0} \leq s|\mathcal{H}] \leq P[e_{1,T_0+h} - \mathbb{E}[e_{1,T_0+h}|\mathcal{H}] \leq s + \Delta] \leq F_+(s + 2\Delta)\,,$$

and, similarly

$$P[e_{1,T_0+h} - \rho_h e_{1,T_0} \leq s|\mathcal{H}] \geq F_+(s - 2\Delta)$$

As a consequence, we have that, for any $u \in (0,1)$:

$$Q_{F_+}(u) - 2\Delta \leq Q_{e_{1,T_0+h}-\rho_h e_{1,T_0}|\mathcal{H}}(u) \leq Q_{F_+}(u) + 2\Delta. \tag{4}$$

Suppose that, upon computation of prediction interval (3), the researcher finds an interval $[a,b]$, with $a > 0$. The researcher would like to quantify the bound Δ on the misspecification that would lead, in a worst-case scenario, the conditional prediction interval based on the correct conditional quantile $Q_{e_{1,T_0+h}-\rho_h e_{1,T_0}|\mathcal{H}}(u)$ to contain zero. From (4), a lower bound for this quantity can be estimated as:

$$\Delta_* = \frac{a}{2}\,.$$

Notice that Δ_* can be easily computed. It also has a very intuitive interpretation. The larger is a, meaning the farther is a from zero, one would require a larger degree of misspecification to change the conclusions of the analysis. A similar sensitivity analysis holds in the case where $[a,b]$, with $b < 0$. In this case, a lower bound for the minimal degree of misspecification that would revert the conclusions of the analysis in a worst-case scenario is $\Delta^* = -\frac{b}{2}$.

An alternative to the sensitivity analysis is to start with an upper bound Δ for the misspecification, and to use (4) to compute prediction intervals that would have valid conditional coverage, if the misspecification were at most Δ. In this case, a misspecification-robust conditional $(1-\alpha)$-prediction interval for δ_{1,T_0+h} would be given by:

$$[\hat{\delta}^{+}_{1,T_0+h} - Q_{\hat{F}_+}(1-\alpha/2) - 2\Delta, \hat{\delta}^{+}_{1,T_0+h} - Q_{\hat{F}_+}(\alpha/2) + 2\Delta]\,. \tag{5}$$

This construction is closely related to the construction of confidence regions for identified sets in the partial identification literature. To see this, consider for simplicity the case in which treatment effects are homogeneous (nonstochastic) over repeated samples, so we may view δ_{1,T_0+h} as a fixed parameter. Notice that, under the misspecification bound $|\mathbb{E}[e_{1,T_0+h}|\mathcal{H}] - \rho_h e_{1,T_0}| \leq \Delta$ almost surely, we can ensure that $\delta_{1,T_0+h} \in \mathcal{S}$, where $\mathcal{S} = [\tilde{\delta} - \Delta, \tilde{\delta} + \Delta]$ and $\tilde{\delta} = \mathbb{E}[\text{plim}_{T\to\infty} \hat{\delta}^{+}_{1,T_0+h}|\mathcal{H}]$. In this setting, an asymptotically conditionally valid confidence region for $\mathcal{S}$ would be given by $[\hat{\delta}^{+}_{1,T_0+h} - Q^*(1-\alpha/2) - \Delta, \hat{\delta}^{+}_{1,T_0+h} - Q^*(\alpha/2) + \Delta]$, where Q^* is the quantile of the distribution of $e_{1,T_0+h} - \mathbb{E}[e_{1,T_0+h}|\mathcal{H}]$. However, this object depends on $\mathbb{E}[e_{1,T_0+h}|\mathcal{H}]$, about whose specification we wish to remain agnostic. We note, however, that it follows from the misspecification bound that $Q_{F^+}(u) - \Delta \leq Q^*(u) \leq Q_{F^+}(u) + \Delta$, for every $u \in (0,1)$. Therefore, the region (5) may be seen as relying on these inequalities to construct a conditionally valid confidence region for $\mathcal{S}$. Since $\delta_{1,T_0+h} \in \mathcal{S}$, this ensures conditional coverage of δ_{1,T_0+h}. Following the insight of Imbens and Manski (2004), smaller prediction intervals could also be obtained by directly targeting coverage of δ_{1,T_0+h}. That is, one could consider intervals $[\hat{\delta}^{+}_{1,T_0+h} - b, \hat{\delta}^{+}_{1,T_0+h} - a]$ that ensure conditional coverage of δ_{1,T_0+h} uniformly over misspecification $|\mathbb{E}[e_{1,T_0+h}|\mathcal{H}] - \rho_h e_{1,T_0}| \leq \Delta$. In this case, a and b must satisfy:

$$\begin{aligned}\mathbb{P}[\hat{\delta}^{+}_{1,T_0+h} - \delta_{1,T_0+h} \in [a,b]] \approx F^*(b - (\mathbb{E}[e_{1,T_0+h}|\mathcal{H}] - \rho_h e_{1,T_0})) - F^*(a - (\mathbb{E}[e_{1,T_0+h}|\mathcal{H}] - \rho_h e_{1,T_0}))\\ \geq 1-\alpha\,,\end{aligned}$$

where F^* is the distribution function of $e_{1,T_0+h} - \mathbb{E}[e_{1,T_0+h}|\mathcal{H}]$. This object depends on the unspecified $\mathbb{E}[e_{1,T_0+h}|\mathcal{H}]$; however, it follows from the misspecification bound that $F^*(x) - F^*(y) \geq F_+(x-\Delta) - F_+(y+\Delta)$. Therefore a feasible and smaller conditional prediction interval for δ_{1,T_0+h} could be found by choosing $a \geq b$ that minimize the interval length $b-a$, subject to:

$$\inf_{|d|\leq\Delta} \hat{F}_+(b-d-\Delta) - \hat{F}_+(a-d+\Delta) = 1-\alpha\,.$$

Finally, a distinction should be made between partially identified and misspecified settings. Note that we are operating in an environment where δ_{1,T_0+h} is effectively point-identified from the conditional on $\mathcal{H}$ distribution of observables. However, if we want to adopt a parsimonious specification for $\mathbb{E}[e_{1,T_0+h}|\mathcal{H}]$, such as the linear correction $\rho_h e_{1,T_0}$, while still acknowledging the possibility of misspecification, then we effectively operate in a setting where δ_{1,T_0+h} can only be bounded as a function of the conditional distribution of observables (and the bound on the misspecification). In this case, tools from the partial identification literature can be applied to conduct misspecification-robust inference.

References

Alvarez, L. and Ferman, B. (2023). Inference in difference-in-differences with few treated units and spatial correlation.

Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., and Wager, S. (2021). Synthetic difference-in-differences. *American Economic Review*, 111(12):4088–4118.

Arkhangelsky, D. and Imbens, G. (2023). Causal Models for Longitudinal and Panel Data: A Survey. *arXiv e-prints*, page arXiv:2311.15458.

Chernozhukov, V., Wüthrich, K., and Zhu, Y. (2021). An exact and robust conformal inference method for counterfactual and synthetic controls. *Journal of the American Statistical Association*, 116(536):1849–1864.

Doudchenko, N. and Imbens, G. W. (2017). Balancing, regression, difference-in-differences and synthetic control methods: A synthesis.

Ferman, B. (2021). On the properties of the synthetic control estimator with many periods and many controls. *Journal of the American Statistical Association*, 116(536):1764–1772.

Ferman, B. and Pinto, C. (2021). Synthetic controls with imperfect pretreatment fit. *Quantitative Economics*, 12(4):1197–1221.

Gonçalves, S. and Ng, S. (2024). Imputation of counterfactual outcomes when the errors are predictable. Working paper.

Imbens, G. W. and Manski, C. F. (2004). Confidence intervals for partially identified parameters. *Econometrica*, 72(6):1845–1857.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

A Asymptotic unconditional coverage of (2)

Suppose that the distribution of $(\epsilon_{1,T_0}, \epsilon_{1,T_0+h})$ is continuous, and that $\hat{F}_+$ is a pointwise consistent estimator of F_+. We show that the prediction interval (2) satisfies:

$$\lim_{T_0\to\infty} P[\delta_{1,T_0+h} \in [\hat{\delta}^+_{1,T_0+h} - Q_{\hat{F}_+}(1-\alpha/2), \hat{\delta}^+_{1,T_0+h} - Q_{\hat{F}_+}(\alpha/2)]] = 1-\alpha\,.$$

To see this, note that, since F_+ is continuous, pointwise convergence of $\hat{F}_+$ to F_+ implies uniform convergence (van der Vaart, 1998, p. 339). Lemma 21.2 of van der Vaart (1998) thus implies that $Q_{\hat{F}_+}(\epsilon) \xrightarrow{p} Q_{F_+}(\epsilon)$ for every $\epsilon \in (0,1)$. It then follows by the continuous mapping theorem that $\mathbf{1}\{\delta_{1,T_0+h} \in [\hat{\delta}^+_{1,T_0+h} - Q_{\hat{F}_+}(\alpha/2), \hat{\delta}^+_{1,T_0+h} - Q_{\hat{F}_+}(1-\alpha/2)]\} \xrightarrow{p} \mathbf{1}\{e_{1,T_0+h} - \rho_h e_{1,T_0} \in [Q_{F_+}(\alpha/2), Q_{F_+}(1-\alpha/2)]\}$. Asymptotic coverage is then a consequence of the bounded convergence theorem, since $\lim_{T_0\to\infty} P[\delta_{1,T_0+h} \in [\hat{\delta}^+_{1,T_0+h} - Q_{\hat{F}_+}(1-\alpha/2), \hat{\delta}^+_{1,T_0+h} - Q_{\hat{F}_+}(\alpha/2)]] = P[e_{1,T_0+h} - \rho_h e_{1,T_0} \in [Q_{F_+}(\alpha/2), Q_{F_+}(1-\alpha/2)]] = 1-\alpha$

B Sensitivity analysis in a location-scale model

We consider the case where $\frac{e_{1,T_0+h} - \mathbb{E}[e_{1,T_0+h}|\mathcal{H}]}{\mathrm{sd}(e_{1,T_0+h}|\mathcal{H})}$ is independent of $\mathcal{H}$. This is a setting in which correctly accounting for the conditional mean and variance would be sufficient to construct a conditionally valid test.

Let $\nu \geq 1$ be a nonstochastic upper-bound for $\frac{\mathrm{sd}(e_{1,T_0+h}|\mathcal{H})}{\mathrm{sd}(e_{1,T_0+h})}$ and $\frac{\mathrm{sd}(e_{1,T_0+h})}{\mathrm{sd}(e_{1,T_0+h}|\mathcal{H})}$, i.e. a constant such that $\frac{\mathrm{sd}(e_{1,T_0+h}|\mathcal{H})}{\mathrm{sd}(e_{1,T_0+h})} \vee \frac{\mathrm{sd}(e_{1,T_0+h})}{\mathrm{sd}(e_{1,T_0+h}|\mathcal{H})} \leq \nu$ almost-surely. Write σ_h for $\mathrm{sd}(e_{1,T_0+h})$. Proceeding as in the main text, we have that, for $s \geq -\Delta$:

$$P[e_{1,T_0+h} - \rho_h e_{1,T_0} \leq s|\mathcal{H}] \leq P\left[\frac{e_{1,T_0+h} - \mathbb{E}[e_{1,T_0+h}|\mathcal{H}]}{\mathrm{sd}(e_{1,T_0+h}|\mathcal{H})} \leq \frac{\nu(s+\Delta)}{\sigma_h}\right] \leq F_+(\Delta + \nu^2(s+\Delta))\,,$$

and, for $s < -\Delta$:

$$P[e_{1,T_0+h} - \rho_h e_{1,T_0} \leq s|\mathcal{H}] \leq P\left[\frac{e_{1,T_0+h} - \mathbb{E}[e_{1,T_0+h}|\mathcal{H}]}{\mathrm{sd}(e_{1,T_0+h}|\mathcal{H})} \leq \frac{(s+\Delta)}{\nu\sigma_h}\right] \leq F_+\left(\Delta + \frac{1}{\nu^2}(s+\Delta)\right).$$

Consequently, we obtain that, for any $u \in (0,1)$, $Q_{e_{1,T_0+h} - \rho_h e_{1,T_0}|\mathcal{H}}(u) \geq \Omega_{\Delta,\nu}(u)$, where

$$\Omega_{\Delta,\nu}(u) = \begin{cases} \nu^2 Q_{F_+}(u) - (\nu^2+1)\Delta, & \text{if } u < F_+(\Delta) \\ \frac{1}{\nu^2} Q_{F+}(u) - \left(\frac{\nu^2+1}{\nu^2}\right)\Delta, & \text{if } u \geq F_+(\Delta)\,. \end{cases} \tag{6}$$

This can be used to construct a *sensitivity region* as follows. Suppose that, upon computation of prediction interval (3), the researcher finds an interval $[a, b]$, with $b < 0$. She would like to compute a set that contains the pairs (Δ, ν) for which a correctly specified conditional prediction interval that uses the true $Q_{e_{1,T_0+h}-\rho_h e_{1,T_0}|\mathcal{H}}(u)$ would, in a worst-case scenario, contain 0. This set may be estimated by the pairs that satisfy:

$$b + (Q_{\hat{F}_+}(\alpha/2) - \hat{\Omega}_{\Delta,v}(\alpha/2)) = \hat{\delta}^+_{1,T_0+h} - \hat{\Omega}_{\Delta,v}(\alpha/2) \geq 0\,,$$

where $\hat{\Omega}_{\Delta,v}$ is an estimator of (6) that replaces Q_{F_+} with $Q_{\hat{F}_+}$. The lower contour for this set can be computed by calculating, for every $\Delta \geq 0$, the ν^2 that solves:

$$b + (Q_{\hat{F}_+}(\alpha/2) - \hat{\Omega}_{\Delta,v}(1-\alpha/2)) = 0\,, \tag{7}$$

and storing those pairs (ν, Δ) such that $\nu^2 \geq 1$. Notice that this calculation nests two particular cases. If we assume that $\nu = 1$, meaning that there is no conditional heteroskedasticity, we arrive from (7) at $\Delta^* = -b/2$, which is precisely the lower bound from the sensitivity analysis in the main text. In contrast, if we assume that $\Delta = 0$, meaning that the conditional mean is correctly specified, we arrive, for $\alpha/2 < F_+(0)$, at $\nu = \sqrt{\frac{b+Q_{\hat{F}_+}(\alpha/2)}{Q_{\hat{F}_+}(\alpha/2)}}$. This is the smallest amount of conditional heteroskedasticity, measured relatively to the unconditional standard deviation, required to change the conclusions of a conditional analysis in a worst-case scenario. This quantity may also be interpreted as a sensitivity analysis in an unconditional setting, where one believes the variance of $e_{1,t}$ may change after T_0. In this situation, ν quantifies the smallest change in the standard deviation, relatively to its pre-treatment value, required to change the conclusions of the analysis in a worst-case scenario.

Finally, we note the sensitivity analsyses discussed in the main text and in this section could be extended to accomodate the uncorrected estimator $\hat{\delta}_{1,T_0+h}$, as this may be seen as a "corrected" estimator with correction equal to 0. In this case, sensitivity with respect to the location quantifies how much the conditional mean should deviate from 0 to revert the conclusions of tests.