Working Paper in Economics No. 788

Serial-correlation-robust power calculation for the analysis-of-covariance estimator

Claes Ek

Department of Economics, June 2020 & rev November 2021

ISSN 1403-2473 (Print)
ISSN 1403-2465 (Online)

UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

Serial-correlation-robust power calculation for the analysis-of-covariance estimator

Claes Ek[a,1]

[a] *Department of Economics, University of Gothenburg, P.O. Box 640, SE-405 30 Gothenburg, Sweden*

Abstract

Ex-ante power calculation is an essential part of the toolkit of experimental economics. In panel experiments, analysis of covariance (ANCOVA) is more efficient than difference-in-differences and is often preferred. The present paper derives a general serial-correlation-robust variance formula and provides the first analytical ANCOVA power-calculation framework to work with real data. A related earlier procedure for ANCOVA was found by Burlig et al. (2020) to yield incorrect power when used to calibrate a minimum detectable effect on actual data, and the authors caution against using it in practice. I show that these errors arose because time shocks were not properly accounted for in an intermediate procedure estimating residual-based variance parameters from pre-existing data. My procedure resolves such issues, thus providing a framework for accurate power calculation with ANCOVA.

Keywords: Power calculation, Experimental design, Panel data, ANCOVA
JEL classification: B4, C9, C23

1 Introduction

Economists are increasingly turning to randomized controlled trials to obtain causal estimates of treatment effects. Within experimental economics, *ex-ante* power analysis to determine an appropriate sample size forms an essential part of the toolkit of applied research; see e.g. Duflo et al. (2007) and Athey and Imbens (2017). In an influential paper, McKenzie (2012) provides the variance of common panel estimators for use in such calculations, noting that repeatedly measuring an experimental outcome may substantially improve precision; see also Frison and Pocock (1992). In particular, he considers the analysis of covariance

[☆] First version: 30 June 2020. This version: 18 November 2021. Funding from the Jan Wallander and Tom Hedelius Foundation is gratefully acknowledged. Funders were not involved in any part of the research. I am grateful to Andreas Dzemski, Peter Jochumzen, Mikael Lindahl, Eskil Rydhe, and Måns Söderbom for useful feedback and suggestions. Any remaining errors are my own.

[1] Corresponding author. *Tel.:* 0046 31 786 52 49. *Email address:* `claes.ek@economics.gu.se`

(ANCOVA) estimator, which replaces the unit fixed effects (FE) included in difference-in-differences (DD) estimation with a covariate for the pre-treatment outcome average of each experimental unit. ANCOVA is more efficient than both DD and an *ex-post* comparison of means, and thus tends to be preferred in randomized panel settings where unit FE are not needed for identification.[2]

However, no existing analytical power-calculation procedure for ANCOVA appears to work properly with actual data. In particular, Burlig et al. (2020) have recently demonstrated that the McKenzie (2012) power formula for ANCOVA generally implies incorrect power in the presence of arbitrarily serially correlated errors. Accounting for serial correlation is important, since it is likely to occur in most real-world settings (Bertrand et al., 2004); non-constant autocorrelation will arise e.g. whenever outcomes that occur close in time are more highly correlated than more distant ones. Being based on an assumption of constant serial correlation, the McKenzie (2012) ANCOVA variance formula thus performs poorly when used to calibrate a uniform minimum detectable effect (MDE) on real-world data.

The purpose of the present paper, therefore, is to provide all steps necessary for accurate and serial-correlation-robust ANCOVA power calculation. Importantly, I am able to confirm that my procedure is the first to successfully predict power for treatment effects that are added to real data.

In previous research, the closest approach to mine is that developed by Burlig et al. (2020). However, while the authors do consider arbitrary serial correlation, they are unable to solve for the variance of ANCOVA in the realistic situation where time shocks are present in the data generating process (DGP). Doing so requires the analyst to e.g. invert matrices of arbitrary dimension; observing such difficulties, Burlig et al. (2020) instead consider the special case of a DGP without time shocks. For that restricted model, they derive the small-sample ANCOVA variance formula

$$\begin{aligned}\mathrm{Var}(\hat{\tau}|\mathbf{X}) = &\left(\frac{1}{P(1-P)J} + \frac{(\bar{Y}_T^B - \bar{Y}_C^B)^2}{Z}\right) \\ &\times \left((1-\theta)^2\sigma_v^2 + \left(\frac{\theta^2}{m} + \frac{1}{r}\right)\sigma_\omega^2 + \frac{\theta^2(m-1)}{m}\psi^B + \frac{r-1}{r}\psi^A - 2\theta\psi^X\right) \qquad (1.1)\end{aligned}$$

[2] The efficiency gains from ANCOVA compared to DD increase as serial correlation approaches zero, so that fixed effects increasingly overcorrect for noisy pre-treatment information. Recent articles in top economics journals that use the ANCOVA estimator include, among others, Allcott and Rogers (2014); Attanasio et al. (2015); Haushofer and Shapiro (2016); Bruhn et al. (2018); Cunha et al. (2019); Armand et al. (2020); and Gerber et al. (2020).

approximated in large samples by equation 10 of Burlig et al. (2020). In this equation, ψ^B, ψ^A, and ψ^X are parameters governing serial correlation in the error term. The authors show that, while formula (1.1) is accurate for simulated panel data that does not include time shocks, it fails to predict realized power in an actual data set of Bloom et al. (2015). Burlig et al. (2020) attribute this outcome to the likely presence of time shocks in real data and thus, in conclusion, "do not recommend that researchers use this formula in real-world applications" (p. 8).[3]

In Section 2 below, I show that the reason that equation (1.1) yields incorrect power in Burlig et al. (2020) is in fact *not* because this formula does not apply in the presence of time shocks. Indeed, by solving a model that includes such effects, I am able to prove that they leave precision unchanged, with equation (1.1) remaining the correct ANCOVA variance. For intuition, note that ANCOVA forms a convex combination of an *ex-post* means comparison and DD, both of which involve comparing means across treatment arms affected identically by any time shocks. As a result, my power-calculation procedure effectively retains equation (1.1).

Then, in Section 3, I demonstrate that the errors encountered by Burlig et al. (2020) arise for a different reason, related to a step where the authors calculate formula (1.1) by first estimating residual-based parameters from a realized data set. I develop an approach that essentially corrects the error and, as a result, allows accurate and serial-correlation-robust ANCOVA power calculation with real data (Bloom et al., 2015). Again, these findings should prove useful, given that ANCOVA is arguably the estimator of choice in panel experimental settings. Finally, Section 4 concludes the paper.

2 ANCOVA regression variance with serial correlation and time shocks

As noted above, the main technical difference between my procedure and the previous research of Burlig et al. (2020) is that I consider the general case where time shocks are included in the DGP. Then, ANCOVA variance again turns out to be (1.1): time shocks do not impact ANCOVA precision. As preliminary evidence that this is so, consider Figure 1, which depicts simulated rejection rates for the regression ANCOVA estimator. The DGP

[3] For DD, by contrast, Burlig et al. (2020) develop a novel serial-correlation-robust power formula that accurately predicts statistical power in actual as well as simulated data. Their procedures are likely to prove highly useful to researchers planning experiments with that estimator.

underlying panel (i) of the figure is[4]

$$Y_{it} = \delta + \tau D_{it} + v_i + \omega_{it} \tag{2.2}$$

where δ is a constant intercept term, D_{it} is a treatment indicator, v_i is a unit intercept, and ω_{it} is a serially correlated error following an AR(1) process, with autoregressive parameter γ varying between zero and 0.9 across simulation sets.

In each of the 10,000 simulation draws underlying the figure, a data set is constructed from (2.2), and a constant treatment effect is calibrated to imply nominal 80% power according to either the McKenzie (2012) ANCOVA variance formula (left figure), or equation (1.1) (right figure). The treatment effect is then added to all units within a randomly drawn treatment group, and an ANCOVA regression is run *ex post*, with robust standard errors clustered by unit. The figure reports rejection rates for the regression treatment coefficient and for varying panel lengths, with treatment always occurring throughout the latter half of the data period. Clearly, the McKenzie (2012) formula is accurate only when $\gamma = 0$, while the Burlig et al. (2020) 'serial-correlation-robust' formula implies very nearly nominal rejection rates in all cases.

In panel (ii) of Figure 1, I examine whether formula (1.1) is appropriate in the presence of time shocks by making two very simple alterations to these procedures. First, panel (ii) is based on the model

$$Y_{it} = \delta_t + \tau D_{it} + v_i + \omega_{it} \tag{2.3}$$

which replaces the constant term with a set of time shocks δ_t, distributed i.i.d. $\mathcal{N}(\mu_\delta, \sigma_\delta^2)$, and drawn once per simulated data set. (Equivalently, we may view model (2.2) as imposing $\sigma_\delta^2 = 0$.) Specifically, in panel (ii), the δ_t have $\sigma_\delta^2 = 10$, which may be compared with $\sigma_v^2 = 80$ and $\sigma_\omega^2 = 10$.

Second, for each simulated data set, I then estimate the ANCOVA regression

$$Y_{it} = \alpha_t + \tilde{\tau} D_i + \theta \bar{Y}_i^B + \epsilon_{it} \tag{2.4}$$

where $\bar{Y}_i^B = (1/m) \sum_{t=-m+1}^{0} Y_{it}$ is the average of the outcome variable for unit i across all m pre-experimental periods, and ϵ_{it} is the regression residual error term. This equation is

[4]This panel also replicates Figure 4 in Burlig et al. (2020), where the authors confirm that formula (1.1) accurately predicts power in simulated data that do not include time shocks. Thus, it is based on the same set of assumptions and parameter values as the original figure; for more information on all simulation procedures and results, see Supplementary Appendix C.

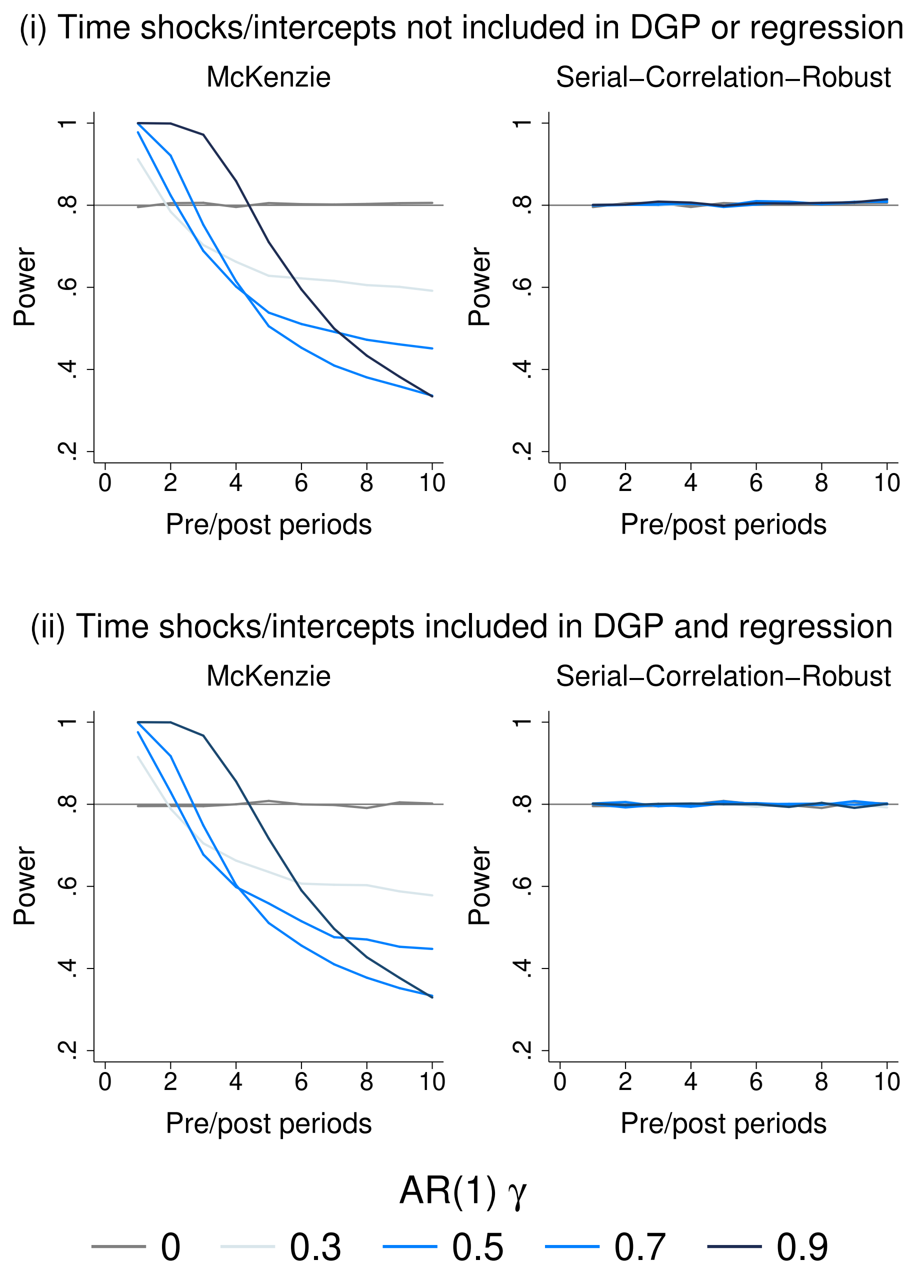

Figure 1: The power of regression ANCOVA is not affected by the presence of time shocks

estimated only on post-treatment observations, allowing the t subscript of D_{it} to be dropped. Regression (2.4) differs from the ANCOVA regressions run in panel (i) only in that time FE α_t are included in place of a constant term. Again, robust standard errors are clustered *ex post* at the unit level.

Clearly, despite the addition of time shocks in panel (ii), rejection rates are practically identical to panel (i). In particular, rejection rates corresponding to serial-correlation-robust formula (1.1) remain approximately nominal. Using other values of σ_δ^2 (including very large ones, such as $\sigma_\delta^2 = 1000$) does not alter these results, strongly suggesting that equation (1.1) can be used with data sets that include time shocks.

Indeed, in Supplementary Appendix A, I prove that this is the case: for DGP (2.3), ANCOVA variance is exactly equal to equation (1.1). In fact, equation (1.1) applies both when time fixed effects are included in the ANCOVA regression (as shown in Appendix A.1 of this note) and when they are not (Appendix A.2), with the added implication that including such terms in an ANCOVA regression does not improve precision. Running the *ex-post* regressions underlying panel (ii) of Figure 1 with only a constant term confirms this point.

While somewhat lengthy and relying heavily on matrix partitioning, the argument has an overall structure similar to the constant-only case. Calculating the variance of the ANCOVA estimator involves evaluating the expression

$$\begin{aligned}\mathrm{Var}(\hat{\tau}|\mathbf{X}) = &\sum_{i=1}^{PJ}\sum_{j=1}^{PJ}\left(M_{ij}^T\sum_{t=1}^{r}\sum_{s=1}^{r}E[\epsilon_{it}\epsilon_{js}|\mathbf{X}]\right) + \sum_{i=1}^{PJ}\sum_{j=PJ+1}^{J}\left(M_{ij}^X\sum_{t=1}^{r}\sum_{s=1}^{r}E[\epsilon_{it}\epsilon_{js}|\mathbf{X}]\right)\\ &+ \sum_{i=PJ+1}^{J}\sum_{j=PJ+1}^{J}\left(M_{ij}^C\sum_{t=1}^{r}\sum_{s=1}^{r}E[\epsilon_{it}\epsilon_{js}|\mathbf{X}]\right)\end{aligned}$$

which derives from a standard coefficient-variance sandwich formula. Here, J is the number of units in the experiment, P proportion of which are treated; r is the number of post-experimental periods in the data; factors $M_{ij}^T, M_{ij}^X, M_{ij}^C$ are all specific to each i and j; $\mathbf{X}$ is the ANCOVA regressor matrix; and ϵ_{it} is again the regression residual for unit i and period t.

The main difficulty in evaluating this expression concerns conditional means $E[\epsilon_{it}\epsilon_{js}|\mathbf{X}]$. In a constant-only model, conditioning on $\mathbf{X}$ amounts to conditioning only on the baseline averages of units i and j, included as controls in the ANCOVA regression. No other baseline averages need be considered, because they are uninformative regarding $\epsilon_{it}\epsilon_{js}$, being composed of average unit intercepts and idiosyncratic errors that are assumed independent across units. Under such conditions, $\sum_{t=1}^{r}\sum_{s=1}^{r} E[\epsilon_{it}\epsilon_{js}|\mathbf{X}] = 0$ whenever $i \neq j$; hence, the variance of ANCOVA is composed solely of those terms where $i = j$. Combining the value of $\sum_{t=1}^{r}\sum_{s=1}^{r} E[\epsilon_{it}\epsilon_{js}|\mathbf{X}]$ when $i = j$ with the variance expression given above then produces equation (1.1).

By contrast, when time shocks are included in the DGP, not only must time fixed effects $\alpha_1, ..., \alpha_r$ be added as conditioning variables in $E[\epsilon_{it}\epsilon_{js}|\mathbf{X}] = 0$, but so must the baseline averages of *all* other units in the experiment. The reason is that these variables now provide additional information about the average pre-treatment time shock; and conditional on $\bar{Y}_i^B$, those pre-treatment shocks are themselves informative regarding the components of ϵ_{it}. However, it turns out that, despite such differences, it remains the case that $\sum_{t=1}^{r}\sum_{s=1}^{r} E[\epsilon_{it}\epsilon_{js}|\mathbf{X}] = 0$ when $i \neq j$. Furthermore, the value taken by $\sum_{t=1}^{r}\sum_{s=1}^{r} E[\epsilon_{it}\epsilon_{js}|\mathbf{X}]$ whenever $i = j$ is exactly equal to the quantity summed across $i = j$ in Burlig et al. (2020). It follows that ANCOVA variance is again (1.1), concluding the proof.

3 An adjusted serial-correlation-robust power calculation for ANCOVA

An obvious question now arises: if it is not the case that the ANCOVA variance formula derived by Burlig et al. (2020) is inappropriate, what might account for the inaccurate rejection rates they obtain using real data? The answer is the following.

With real data, the parameters of the DGP are unknown, and Burlig et al. (2020) construct a useful procedure for calculating MDEs by first estimating a set of residual-based variance parameters from pre-existing data. Formula (1.1) can then be accurately re-written in terms of the estimands of those parameters, rather than the parameters of the DGP. In a reasonable attempt to remain consistent with their assumed time-shock free model (2.2), they ignore the possibility of time shocks throughout the parameter estimation step as well.

Unfortunately however, when time shocks are ignored in estimation, the variation that these cause in the data (which, as noted, does not affect ANCOVA precision) will instead be attributed to idiosyncratic residual-based parameter $\sigma_{\hat{\omega}}^2$, which *does* affect predicted power. As a result, the ANCOVA variance calculated using residual-based parameter estimates will be biased upward; and the implied MDE, as well as rejection rates, will likewise be too large. Fortunately, the problem has a simple solution: one simply takes the presence of time shocks into account during the estimation step as well. My modified procedure is described in detail in Appendix B of this paper.[5]

[5] In principle, the ANCOVA variance calculated from residual-based parameters may be negative due to sampling error in the parameters, in which case no MDE will be calculated. However, in simulations with

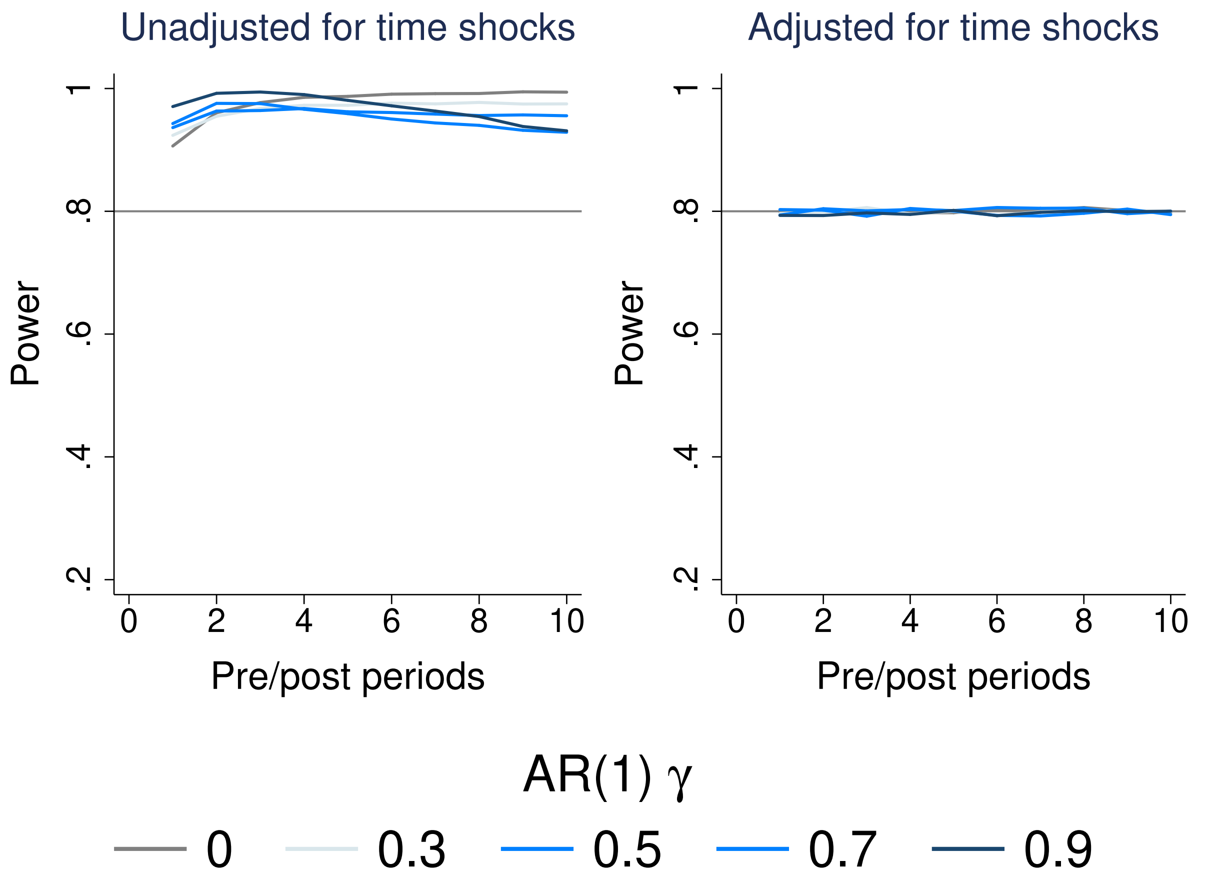

Figure 2: Allowing for time shocks in residual-based parameter estimation: simulated data

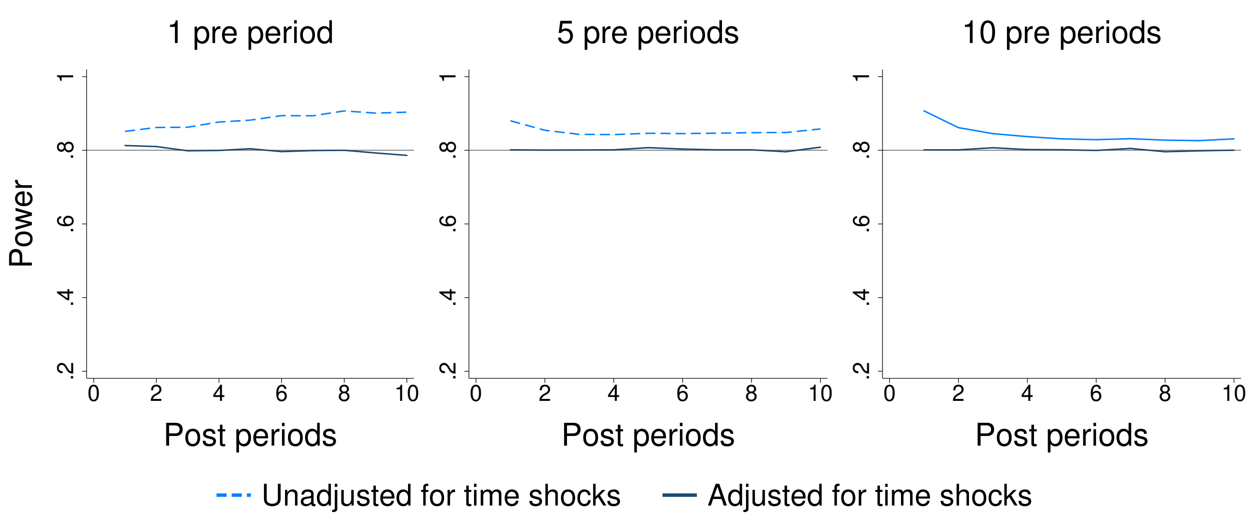

Figure 3: Allowing for time shocks in residual-based parameter estimation: real data

In Figure 2, I compare the two approaches for simulated data. The figure is based on the same model, parameters, and overall simulation procedure as panel (ii) of Figure 1; but instead of computing an MDE directly from formula (1.1) and the parameters of the DGP, I use a set of residual-based parameters estimated from each simulated data set. In the left-hand panel of Figure 2, I follow the estimation procedure described for ANCOVA in Appendix E.3 of Burlig et al. (2020); as expected, this procedure ignores the presence of time shocks and consequently yields excessively high rejection rates.[6] In the right-hand panel, I use the modified approach that correctly accounts for time shocks, and attain nominal power. Note that both panels estimate ANCOVA *ex post*, clustering robust standard errors by unit, but the regressions include time FE only in the right-hand panel.

Then, in Figure 3, the exercise is repeated for real data, specifically the Bloom et al. (2015) data set used for Figure 7 of Burlig et al. (2020). Here, as in the original figure, each panel simulates experiments with a certain number of pre-treatment periods $m \in \{1, 5, 10\}$, while horizontal axes vary the number of post-treatment periods ($1 \leq r \leq 10$). I retain all procedures and steps used in Figure 2 except those for generating the data: in particular, I calibrate an MDE by combining power formula (1.1) with a set of residual-based parameters estimated from the Bloom et al. (2015) data. Then, in each simulation run, the effect is added to, and estimated from, a randomly drawn treatment group (again, with regressions including time FE only in the adjusted procedure).

When not accounting for time shocks in the parameter estimation step (dashed lines), I am able to closely replicate the original figure, where rejection rates deviate from nominal levels. When instead I account for time shocks in the proper way (solid lines), appropriate rejection rates are again achieved. Thus, the modified estimation procedure also appears to work well with actual data.

4 Conclusion

This short paper has provided the first demonstration that accurate *ex-ante* power calculation may feasibly be performed for ANCOVA. To summarize, I compute a minimum

AR(1) processes, this problem seems extremely rare and seems to require a combination of low sample size, many periods, extreme variance structures, and weak serial correlation. For example, when $J = 80$, $m = 40$, $r = 20$, $\sigma_v^2 = 2$, $\sigma_\omega^2 = 400$, and $\gamma = 0.1$, the problem arose in 2.3% of simulation runs.

[6] For each simulated data set, I estimate $\tilde{\sigma}^2_{\hat{v},S}$, $\tilde{\sigma}^2_{\hat{\omega},S}$, $\tilde{\psi}^A_{\hat{\omega},S}$, and $\tilde{\psi}^B_{\hat{\omega},S}$ only once, with estimation range S and sample size I given by all periods and all units in the data, respectively. $\tilde{\sigma}^2_{\hat{v},S}$ is estimated as the sample variance of the fitted unit fixed effects, $\hat{v}_i$. To obtain unbiased estimates of the residual-based parameters, I then deflate $\tilde{\sigma}^2_{\hat{\omega},S}$ by $(IT-1)/(IT)$ for panel length T, and $\tilde{\sigma}^2_{\hat{v},S}$ by $(I-1)/I$. I leave the (sample) $\tilde{\psi}$ estimates unadjusted, in accordance with the discussion of e.g. $E[\tilde{\psi}^B_{\hat{\omega}}]$ in Appendix E.3 of Burlig et al. (2020).

detectable effect by combining formula (1.1) with the steps for residual-based parameter estimation outlined in Appendix B. In the end, the procedure turns out to involve only relatively minor modifications to the previous serial-correlation-robust approach of Burlig et al. (2020). It is hoped that these new methods will usefully expand the power-calculation toolkit available to experimenters even further.

References

H. Allcott and T. Rogers. The short-run and long-run effects of behavioral interventions: experimental evidence from energy conservation. *American Economic Review*, 104(10), 2014.

A. Armand, A. Coutts, P.C. Vicente, and I. Vilela. Does information break the political resource curse? Experimental evidence from Mozambique. *American Economic Review*, 110(11):3431–3453, 2020.

S. Athey and G.W. Imbens. The econometrics of randomized experiments. In A.V. Banerjee and E. Duflo, editors, *Handbook of Economic Field Experiments*, volume 1. 2017.

O. Attanasio, B. Augsburg, R. De Haas, E. Fitzsimons, and H. Harmgart. The impacts of microfinance: Evidence from joint-liability lending in Mongolia. *American Economic Journal: Applied Economics*, 7(1):90–122, 2015.

M. Bertrand, E. Duflo, and S. Mullainathan. How Much Should We Trust Differences-in-Differences Estimates? *The Quarterly Journal of Economics*, 119(1), 2004.

N. Bloom, J. Liang, J. Roberts, and ZJ Ying. Does working from home work? Evidence from a Chinese experiment. *The Quarterly Journal of Economics*, 130(1):165–218, 2015.

M. Bruhn, D. Karlan, and A. Schoar. The impact of consulting services on small and medium enterprises: Evidence from a randomized trial in Mexico. *Journal of Political Economy*, 126(2):635–687, 2018.

F. Burlig, L. Preonas, and M. Woerman. Panel data and experimental design. *Journal of Development Economics*, 144:102458, 2020.

J.M. Cunha, G. De Giorgi, and S. Jayachandran. The price effects of cash versus in-kind transfers. *Review of Economic Studies*, 86(1):240–281, 2019.

E. Duflo, R. Glennerster, and M. Kremer. Using randomization in development economics research: a toolkit. *Handbook of Development Economics*, 4:3895–3962, 2007.

L. Frison and S.J. Pocock. Repeated measures in clinical trials: analysis using mean summary statistics and its implications for design. *Statistics in Medicine*, 11(13), 1992.

A. Gerber, M. Hoffman, J. Morgan, and C. Raymond. One in a million: Field experiments on perceived closeness of the election and voter turnout. *American Economic Journal: Applied Economics*, 12(3):287–325, 2020.

J. Haushofer and J. Shapiro. The short-term impact of unconditional cash transfers to the poor: Experimental evidence from Kenya. *The Quarterly Journal of Economics*, 131(4): 1973–2042, 2016.

D. McKenzie. Beyond baseline and follow-up: The case for more T in experiments. *Journal of Development Economics*, 99(2):210–221, 2012.

Online Appendices for article "Serial-correlation-robust power calculation for the analysis-of-covariance estimator"

Appendix A. Analysis of covariance (ANCOVA) variance formulas

This appendix derives the variance of the ANCOVA treatment estimator under the assumption that time shocks are present in the data generating process and possibly in the ANCOVA regression equation as well. I retain all model assumptions in Burlig et al. (2020) and also repeat them below for convenience, with the exception of the parts of Assumption 1 and 2 related to time shocks, which have been updated accordingly.

There are J experimental units, P proportion of which are randomized into treatment. The researcher collects outcome data Y_{it} for each unit i, across m pre-treatment time periods and r post-treatment time periods. For treated units, $D_{it} = 0$ in pre-treatment periods and $D_{it} = 1$ in post-treatment periods; for control units, $D_{it} = 0$ in all periods.

Assumption 1.1. *(Data generating process). The data are generated according to the following model:*

$$Y_{it} = \delta_t + \tau D_{it} + v_i + \omega_{it} \tag{A.1}$$

where Y_{it} is the outcome of interest for unit i at time t; τ is the treatment effect that is homogenous across all units and all time periods; D_{it} is a time-varying treatment indicator; v_i is a time-invariant unit effect distributed i.i.d. $\mathcal{N}(0, \sigma_v^2)$; ω_{it} is an idiosyncratic error term distributed (not necessarily i.i.d.) $\mathcal{N}(0, \sigma_\omega^2)$. Finally, δ_t is a time shock specific to time t that is homogenous across all units and distributed i.i.d. $\mathcal{N}(\mu_\delta, \sigma_\delta^2)$.

Assumption 1.2. *(Strict exogeneity). $E[\omega_{it}|\mathbf{X}_r] = 0$, where $\mathbf{X}_r$ is a full rank matrix of regressors, including a constant, the treatment indicator $\mathbf{D}$, $J - 1$ unit dummies, and $(m + r) - 1$ time dummies. This follows from random assignment of D_{it}.*

Assumption 1.3. *(Balanced panel). The number of pre-treatment observations, m, and post-treatment observations, r, is the same for each unit, and all units are observed in every time period.*

Assumption 1.4. *(Independence across units). $E[\omega_{it}\omega_{js}|\mathbf{X}_r] = 0, \forall i \neq j, \forall t, s$.*

Assumption 1.5. *(Uniform covariance structures).* Define:

$$\psi_i^B \equiv \frac{2}{m(m-1)} \sum_{t=-m+1}^{-1} \sum_{s=t+1}^{0} Cov(\omega_{it}, \omega_{is} | \mathbf{X}_r)$$

$$\psi_i^A \equiv \frac{2}{r(r-1)} \sum_{t=1}^{r-1} \sum_{s=t+1}^{r} Cov(\omega_{it}, \omega_{is} | \mathbf{X}_r)$$

$$\psi_i^X \equiv \frac{1}{mr} \sum_{t=-m+1}^{0} \sum_{s=1}^{r} Cov(\omega_{it}, \omega_{is} | \mathbf{X}_r)$$

to be the average pre-treatment, post-treatment, and across-period covariance between different error terms of unit i, respectively. Using these definitions, assume that $\psi^B = \psi_i^B$, $\psi^A = \psi_i^A$, and $\psi^X = \psi_i^X$ $\forall i$.

We will derive the variance of the ANCOVA treatment-effect estimator for two different regression specifications. First, in Appendix S1.1, we consider the case when time shocks are included in the regression equation; then, in Appendix S1.2, we consider the case when they are not. In both cases, the result will be equal to variance equation (1.1) in the main text.

A.1 Time shocks included in ANCOVA regression

Consider the following ANCOVA regression model

$$Y_{it} = \alpha_t + \tilde{\tau} D_i + \theta \bar{Y}_i^B + \epsilon_{it} \tag{A.2}$$

where Y_{it} and D_i are defined as above; $\bar{Y}_i^B = (1/m) \sum_{t=-m+1}^{0} Y_{it}$ is the pre-period average of the outcome variable for unit i; and ϵ_{it} is the regression residual error term. Finally, α_t is one of r time fixed effects replacing the constant term in Burlig et al. (2020). As usual for ANCOVA regressions, equation (A.2) is estimated only on post-treatment observations, allowing the t subscript of D_{it} to be dropped.

Regression (A.2) consistently estimates the coefficients of the linear projection of the outcome as given by (A.1) on the set of regressors. As usual, the resulting projection error will satisfy $E[\mathbf{X}'\boldsymbol{\epsilon}] = 0$, where $\mathbf{X}$ is the regressor matrix in (A.2), and $\boldsymbol{\epsilon}$ is the full vector of residuals. This equation system includes the conditions $\epsilon_{it} = 0$ for all i and t.[1] Moreover, we may use it to solve for the set of projection parameters to which the ANCOVA (i.e.

[1]Under the alternative assumption that D_i is a random variable with expected value P, we need consider only the equations associated with a single unit i: $E[\mathbf{X}_i'\boldsymbol{\epsilon}_i] = 0$.

OLS) regression estimator will converge.[2] These coefficients are $\alpha_t = \delta_t - \theta\bar{\delta}^B$, where $\bar{\delta}^B = \frac{1}{m}\sum_{p=-m+1}^{0}\delta_p$; $\tilde{\tau} = \tau$, i.e. the treatment effect of the DGP; and

$$\theta = \frac{m(\sigma_v^2 + \psi^X)}{m\sigma_v^2 + \sigma_\omega^2 + (m-1)\psi^B} \tag{A.3}$$

As a result, we also have

$$\epsilon_{it} = v_i + \omega_{it} - \theta\left(v_i + \frac{1}{m}\sum_{p=-m+1}^{0}\omega_{it}\right) = v_i + \omega_{it} - \theta(v_i + \bar{\omega}_i^B)$$

where $\bar{\omega}_i^B = (1/m)\sum_{p=-m+1}^{0}\omega_{ip}$.

Our goal is now to derive the variance of the $\hat{\tau}$ coefficient estimate implied by the combination of DGP (A.1) and regression (A.2). Denoting as $\hat{\boldsymbol{\beta}}$ the set of regression coefficients from OLS estimation of (A.2) given (A.1), the coefficient covariance matrix is given by the sandwich formula

$$\text{Var}(\hat{\boldsymbol{\beta}}|\mathbf{X}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E[\boldsymbol{\epsilon}\boldsymbol{\epsilon}'|\mathbf{X}]\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \tag{A.4}$$

where, since $\hat{\boldsymbol{\beta}}$ contains r time fixed effects, $\text{Var}(\hat{\tau}|\mathbf{X})$ forms element $(r+1, r+1)$.

[2] OLS estimation occurs within data set, i.e. for a given draw of time shocks. Thus, in computing the projection parameters, we need to treat the set of time shocks δ_t as nonstochastic; otherwise, for instance, every α_t will equal the same value. All other computations throughout this proof do treat the time shocks as random.

As a first step in calculating this quantity, matrix multiplication yields

$$
\mathbf{X}'E[\boldsymbol{\epsilon}\boldsymbol{\epsilon}'|\mathbf{X}]\mathbf{X} = \begin{pmatrix}
\sum_{i=1}^{J}\sum_{j=1}^{J} E[\epsilon_{i1}\epsilon_{j1}|\mathbf{X}] & \cdots & \sum_{i=1}^{J}\sum_{j=1}^{J} E[\epsilon_{i1}\epsilon_{jr}|\mathbf{X}] & \sum_{i=1}^{PJ}\sum_{j=1}^{J}\sum_{t=1}^{r} E[\epsilon_{it}\epsilon_{j1}|\mathbf{X}] & \sum_{i=1}^{J}\sum_{j=1}^{J}\sum_{t=1}^{r} \bar{Y}_i^B E[\epsilon_{it}\epsilon_{j1}|\mathbf{X}] \\
\vdots & \ddots & \vdots & \vdots & \vdots \\
\sum_{i=1}^{J}\sum_{j=1}^{J} E[\epsilon_{ir}\epsilon_{j1}|\mathbf{X}] & \cdots & \sum_{i=1}^{J}\sum_{j=1}^{J} E[\epsilon_{ir}\epsilon_{jr}|\mathbf{X}] & \sum_{i=1}^{PJ}\sum_{j=1}^{J}\sum_{t=1}^{r} E[\epsilon_{it}\epsilon_{jr}|\mathbf{X}] & \sum_{i=1}^{J}\sum_{j=1}^{J}\sum_{t=1}^{r} \bar{Y}_i^B E[\epsilon_{it}\epsilon_{jr}|\mathbf{X}] \\
\sum_{i=1}^{PJ}\sum_{j=1}^{J}\sum_{t=1}^{r} E[\epsilon_{it}\epsilon_{j1}|\mathbf{X}] & \cdots & \sum_{i=1}^{PJ}\sum_{j=1}^{J}\sum_{t=1}^{r} E[\epsilon_{it}\epsilon_{jr}|\mathbf{X}] & \sum_{i=1}^{PJ}\sum_{j=1}^{PJ}\sum_{t=1}^{r}\sum_{s=1}^{r} E[\epsilon_{it}\epsilon_{js}|\mathbf{X}] & \sum_{i=1}^{PJ}\sum_{j=1}^{J}\sum_{t=1}^{r}\sum_{s=1}^{r} \bar{Y}_j^B E[\epsilon_{it}\epsilon_{js}|\mathbf{X}] \\
\sum_{i=1}^{J}\sum_{j=1}^{J}\sum_{t=1}^{r} \bar{Y}_i^B E[\epsilon_{it}\epsilon_{j1}|\mathbf{X}] & \cdots & \sum_{i=1}^{J}\sum_{j=1}^{J}\sum_{t=1}^{r} \bar{Y}_i^B E[\epsilon_{it}\epsilon_{jr}|\mathbf{X}] & \sum_{i=1}^{PJ}\sum_{j=1}^{J}\sum_{t=1}^{r}\sum_{s=1}^{r} \bar{Y}_j^B E[\epsilon_{it}\epsilon_{js}|\mathbf{X}] & \sum_{i=1}^{J}\sum_{j=1}^{J}\sum_{t=1}^{r}\sum_{s=1}^{r} \bar{Y}_i^B \bar{Y}_j^B E[\epsilon_{it}\epsilon_{js}|\mathbf{X}]
\end{pmatrix} \quad \text{(A.5)}
$$

Next, consider inverting $(1/J)\mathbf{X}'\mathbf{X}$, i.e. the following symmetric square matrix of dimension $r+2$:

$$
\frac{1}{J}\mathbf{X}'\mathbf{X} = \left(\begin{array}{ccc:cc}
1 & \cdots & 0 & P & \bar{Y}^B \\
\vdots & \ddots & \vdots & \vdots & \vdots \\
0 & \cdots & 1 & P & \bar{Y}^B \\
\hdashline
P & \cdots & P & rP & rP\bar{Y}_T^B \\
\bar{Y}^B & \cdots & \bar{Y}^B & rP\bar{Y}_T^B & \frac{r}{J}\sum_{j=1}^{J}(\bar{Y}_i^B)^2
\end{array}\right)
$$

which we have partitioned into four submatrices according to the dashed lines, and where

$$\begin{aligned}
\bar{Y}^B &= \frac{1}{mJ}\sum_{i=1}^{J}\sum_{t=-m+1}^{0} Y_{it} \\
\bar{Y}_T^B &= \frac{1}{mPJ}\sum_{i=1}^{PJ}\sum_{t=-m+1}^{0} Y_{it} \\
\sum_{i=1}^{J}(\bar{Y}_i^B)^2 &= \sum_{i=1}^{J}\left(\frac{1}{m}\sum_{t=-m+1}^{0} Y_{it}\right)^2 \\
&= Z + PJ(\bar{Y}_T^B)^2 + (1-P)J(\bar{Y}_C^B)^2
\end{aligned}$$

for $Z = \sum_{k=1}^{PJ}(\bar{Y}_k^B - \bar{Y}_T^B)^2 + \sum_{k=PJ+1}^{J}(\bar{Y}_k^B - \bar{Y}_C^B)^2$. In general, for any partitioned matrix $\mathbf{G}$,

$$\begin{aligned}
\mathbf{G}^{-1} &= \left(\begin{array}{c|c} \mathbf{G}_{11} & \mathbf{G}_{12} \\ \hline \mathbf{G}_{21} & \mathbf{G}_{22} \end{array}\right)^{-1} \\
&= \begin{pmatrix} (\mathbf{G}_{11} - \mathbf{G}_{12}\mathbf{G}_{22}^{-1}\mathbf{G}_{21})^{-1} & -\mathbf{G}_{11}^{-1}\mathbf{G}_{12}(\mathbf{G}_{22} - \mathbf{G}_{21}\mathbf{G}_{11}^{-1}\mathbf{G}_{12})^{-1} \\ -(\mathbf{G}_{22} - \mathbf{G}_{21}\mathbf{G}_{11}^{-1}\mathbf{G}_{12})^{-1}\mathbf{G}_{21}\mathbf{G}_{11}^{-1} & (\mathbf{G}_{22} - \mathbf{G}_{21}\mathbf{G}_{11}^{-1}\mathbf{G}_{12})^{-1} \end{pmatrix} \qquad \text{(A.6)}
\end{aligned}$$

where the top-right submatrix of $\mathbf{G}^{-1}$ also equals $-(\mathbf{G}_{11} - \mathbf{G}_{12}\mathbf{G}_{22}^{-1}\mathbf{G}_{21})^{-1}\mathbf{G}_{12}(\mathbf{G}_{22})^{-1}$.

For $(1/J)\mathbf{X}'\mathbf{X}$, due to the inclusion of time fixed effects in the regression, $\mathbf{G}_{11}$ is an $r \times r$ identity matrix, simplifying the calculations. Indeed, all submatrices of $((1/J)\mathbf{X}'\mathbf{X})^{-1}$ except the top-left one are straightforward to calculate. That final, top-left submatrix is the inverse of

$$\begin{pmatrix} 1 - \frac{\frac{PZ}{J} + (1-P)(\bar{Y}_C^B)^2}{r\left(\frac{Z}{J} + (1-P)(\bar{Y}_C^B)^2\right)} & \cdots & -\frac{\frac{PZ}{J} + (1-P)(\bar{Y}_C^B)^2}{r\left(\frac{Z}{J} + (1-P)(\bar{Y}_C^B)^2\right)} \\ \vdots & \ddots & \vdots \\ -\frac{\frac{PZ}{J} + (1-P)(\bar{Y}_C^B)^2}{r\left(\frac{Z}{J} + (1-P)(\bar{Y}_C^B)^2\right)} & \cdots & 1 - \frac{\frac{PZ}{J} + (1-P)(\bar{Y}_C^B)^2}{r\left(\frac{Z}{J} + (1-P)(\bar{Y}_C^B)^2\right)} \end{pmatrix} \qquad \text{(A.7)}$$

which is a symmetric matrix where all (off)diagonal elements are equal to the same value; note that $\bar{Y}_C^B = \frac{1}{m(1-P)J}\sum_{i=PJ+1}^{J}\sum_{t=-m+1}^{0} Y_{it}$.[3] To invert this matrix, we use the following lemma.

Lemma 1.1. *Any n-dimensional square matrix of the form*

$$\mathbf{Y}_1 = \begin{pmatrix} x_1 & x_2 & \cdots & x_2 \\ x_2 & x_1 & \cdots & x_2 \\ \vdots & \vdots & \ddots & \vdots \\ x_2 & x_2 & \cdots & x_1 \end{pmatrix}$$

[3]Because we are interested in element $(r+1, r+1)$ of $\mathrm{Var}(\hat{\tau}|\mathbf{X})$, calculating this final submatrix is, strictly speaking, unnecessary and is done for completeness only.

has $|\mathbf{Y}_1| = (x_1 - x_2)^{n-1}(x_1 + (n-1)x_2)$*, and any* n*-dimensional square matrix of the form*

$$\mathbf{Y}_2 = \begin{pmatrix} x_2 & x_2 & \cdots & x_2 \\ x_2 & x_1 & \cdots & x_2 \\ \vdots & \vdots & \ddots & \vdots \\ x_2 & x_2 & \cdots & x_1 \end{pmatrix}$$

has $|\mathbf{Y}_2| = x_2(x_1 - x_2)^{n-1}$.

Proof: Assuming the lemma holds for matrices of dimension $n-1$, we have (note that the second term is based on interchanging columns or rows to produce a submatrix of type $\mathbf{Y}_2$):

$$\begin{aligned} |\mathbf{Y}_1| &= x_1((x_1 - x_2)^{n-2}(x_1 + (n-2)x_2)) - (n-1)x_2^2(x_1 - x_2)^{n-2} \\ &= (x_1 - x_2)^{n-1}(x_1 + (n-1)x_2) \end{aligned}$$

and

$$\begin{aligned} |\mathbf{Y}_2| &= x_2(x_1 - x_2)^{n-2}(x_1 + (n-2)x_2) - (n-1)x_2^2(x_1 - x_2)^{n-2} \\ &= x_2(x_1 - x_2)^{n-1} \end{aligned}$$

Finally, it is simple to confirm that these expressions also hold for $n = 2$. $\square$

Lemma 1 may be immediately applied to invert submatrix (A.7), interchanging cofactor columns and/or rows as needed for the result to apply. In summary, since $(\mathbf{X}'\mathbf{X})^{-1} = (1/J)((1/J)\mathbf{X}'\mathbf{X})^{-1}$, we have

$$(\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{r(1-P)Z} \times \left(\begin{array}{ccccc} \frac{(P+(1-P)r)Z}{J} + (1-P)(\bar{Y}_C^B)^2 & \cdots & \frac{PZ}{J} + (1-P)(\bar{Y}_C^B)^2 & -\frac{Z}{J} + (1-P)\bar{Y}_C^B(\bar{Y}_T^B - \bar{Y}_C^B) & -(1-P)\bar{Y}_C^B \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ \frac{PZ}{J} + (1-P)(\bar{Y}_C^B)^2 & \cdots & \frac{(P+(1-P)r)Z}{J} + (1-P)(\bar{Y}_C^B)^2 & -\frac{Z}{J} + (1-P)\bar{Y}_C^B(\bar{Y}_T^B - \bar{Y}_C^B) & -(1-P)\bar{Y}_C^B \\ -\frac{Z}{J} + (1-P)\bar{Y}_C^B(\bar{Y}_T^B - \bar{Y}_C^B) & \cdots & -\frac{Z}{J} + (1-P)\bar{Y}_C^B(\bar{Y}_T^B - \bar{Y}_C^B) & \frac{Z}{PJ} + (1-P)(\bar{Y}_T^B - \bar{Y}_C^B)^2 & -(1-P)(\bar{Y}_T^B - \bar{Y}_C^B) \\ -(1-P)\bar{Y}_C^B & \cdots & -(1-P)\bar{Y}_C^B & -(1-P)(\bar{Y}_T^B - \bar{Y}_C^B) & 1-P \end{array} \right) \quad \text{(A.8)}$$

and may combine (A.8) with (A.5) to calculate element $(r+1, r+1)$ of (A.4) as

$$
\begin{aligned}
\mathrm{Var}(\hat{\tau}|\mathbf{X}) = \frac{1}{J^2 r^2 Z^2}\Bigg\{ & \frac{1}{P^2}\sum_{i=1}^{PJ}\sum_{j=1}^{PJ}\Bigg((Z - PJ(\bar{Y}_T^B - \bar{Y}_C^B)(\bar{Y}_i^B - \bar{Y}_T^B)) \\
& \times (Z - PJ(\bar{Y}_T^B - \bar{Y}_C^B)(\bar{Y}_j^B - \bar{Y}_T^B)) \times \left(\sum_{t=1}^{r}\sum_{s=1}^{r} E[\epsilon_{it}\epsilon_{js}|\mathbf{X}]\right)\Bigg) \\
& + \frac{2}{P(1-P)}\sum_{i=1}^{PJ}\sum_{j=PJ+1}^{J}\Bigg((Z - PJ(\bar{Y}_T^B - \bar{Y}_C^B)(\bar{Y}_i^B - \bar{Y}_T^B)) \\
& \times (-Z - (1-P)J(\bar{Y}_T^B - \bar{Y}_C^B)(\bar{Y}_j^B - \bar{Y}_C^B)) \times \left(\sum_{t=1}^{r}\sum_{s=1}^{r} E[\epsilon_{it}\epsilon_{js}|\mathbf{X}]\right)\Bigg) \\
& + \frac{1}{(1-P)^2}\sum_{i=PJ+1}^{J}\sum_{j=PJ+1}^{J}\Bigg((-Z - (1-P)J(\bar{Y}_T^B - \bar{Y}_C^B)(\bar{Y}_i^B - \bar{Y}_C^B)) \\
& \times (-Z - (1-P)J(\bar{Y}_T^B - \bar{Y}_C^B)(\bar{Y}_j^B - \bar{Y}_C^B)) \times \left(\sum_{t=1}^{r}\sum_{s=1}^{r} E[\epsilon_{it}\epsilon_{js}|\mathbf{X}]\right)\Bigg)\Bigg\}
\end{aligned} \tag{A.9}
$$

which, despite the inclusion of time FE, is identical to the corresponding expression (A51) in Burlig et al. (2020). For the remainder of the derivation, we will be concerned with evaluating this expression. To do so, we first need to compute the summed conditional means included in each of the three terms in (A.9).

For a given single conditional mean with $i \neq j$,

$$
\begin{aligned}
E[\epsilon_{it}\epsilon_{js}|\mathbf{X}] &= E[\epsilon_{js}E[\epsilon_{it}|\epsilon_{js}, \mathbf{X}]|\mathbf{X}] \\
&= E[\epsilon_{js}E[\epsilon_{it}|\epsilon_{js}, \bar{Y}_j^B, \bar{Y}_i^B, \bar{Y}_{-i,-j}^B, \alpha_1, ..., \alpha_r]|\bar{Y}_j^B, \bar{Y}_i^B, \bar{Y}_{-i,-j}^B, \alpha_1, ..., \alpha_r]
\end{aligned}
$$

where the first equality uses the law of iterated expectations, and $\bar{Y}_{-i,-j}^B$ is the set of all baseline averages associated with units other than i and j. Thus, although (as we will see below) ϵ_{it} is unconditionally uncorrelated with baseline averages other than $\bar{Y}_i^B$, evaluating the mean of ϵ_{it} conditional on $\mathbf{X}$ nevertheless implies conditioning on *all* baseline averages in the experiment. The reason is somewhat subtle: each baseline average (as well as each α_t) provides additional information regarding the average pre-period time shocks $\bar{\delta}^B$; but conditional on $\bar{Y}_i^B$, this is itself informative regarding v_i and $\bar{\omega}_i^B$, both of which are components of ϵ_{it}.

When instead $i = j$, we have

$$
E[\epsilon_{it}\epsilon_{is}|\mathbf{X}] = E[\epsilon_{is}E[\epsilon_{it}|\epsilon_{is}, \bar{Y}_i^B, \bar{Y}_{-i}^B, \alpha_1, ..., \alpha_r]|\bar{Y}_i^B, \bar{Y}_{-i}^B, \alpha_1, ..., \alpha_r]
$$

where $\bar{Y}_{-i}^{B}$ is the set of all baseline averages associated with units other than i.

Since the residuals as well as all conditioning variables are linear functions of normal variables and thus themselves normally distributed, we may evaluate either of the above conditional means using the following formula:

$$E[x|\mathbf{y}] = \mu_x + \mathbf{\Sigma}^{\mathbf{xy}}(\mathbf{\Sigma}^{\mathbf{yy}})^{-1}(\mathbf{y} - \boldsymbol{\mu}_{\mathbf{y}}) \tag{A.10}$$

where μ_x is the mean of the normal variable x; $\mathbf{\Sigma}^{\mathbf{xy}}$ is a row vector collecting the covariances between x and each element of the vector of normally distributed conditioning variables $\mathbf{y}$; $\mathbf{\Sigma}^{\mathbf{yy}}$ is the variance-covariance matrix of $\mathbf{y}$; and $\boldsymbol{\mu}_{\mathbf{y}}$ is the vector of means of $\mathbf{y}$.[4] In our case, $\mu_x = 0$, since all residuals have mean zero by the properties of linear projection. Also, $E(\bar{Y}_i^B) = \mu_\delta$ for all i, and $E(\alpha_t) = (1-\theta)\mu_\delta$ for all t.

Both when $i \neq j$ and when $i = j$, the $(1+J+r)$-dimensional covariance matrix corresponding to the inner conditional expectation $E[\epsilon_{it}|\epsilon_{js}, \mathbf{X}]$ is

$$\mathbf{\Sigma}^{\mathbf{yy}} = \left(\begin{array}{ccccc|cccc}
a_s & b_s & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\
b_s & c & d & \cdots & d & -\theta d & -\theta d & \cdots & -\theta d \\
0 & d & c & \cdots & d & -\theta d & -\theta d & \cdots & -\theta d \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & & \vdots \\
0 & d & d & \cdots & c & -\theta d & -\theta d & \cdots & -\theta d \\
\hline
0 & -\theta d & -\theta d & \cdots & -\theta d & (m+\theta^2)d & \theta^2 d & \cdots & \theta^2 d \\
0 & -\theta d & -\theta d & \cdots & -\theta d & \theta^2 d & (m+\theta^2)d & \cdots & \theta^2 d \\
\vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & -\theta d & -\theta d & \cdots & -\theta d & \theta^2 d & \theta^2 d & \cdots & (m+\theta^2)d
\end{array}\right)$$

$$\equiv \begin{pmatrix} \mathbf{\Sigma}_{11}^{\mathbf{yy}} & \mathbf{\Sigma}_{12}^{\mathbf{yy}} \\ \mathbf{\Sigma}_{21}^{\mathbf{yy}} & \mathbf{\Sigma}_{22}^{\mathbf{yy}} \end{pmatrix} \tag{A.11}$$

[4]If one assumes that all δ_t are fixed rather than random, then pre-treatment time shocks cease to be informative regarding the residuals. Thus, $E[\epsilon_{it}|\epsilon_{js}, \mathbf{X}] = E[\epsilon_{it}|\bar{Y}_i^B]$ for $i \neq j$, $E[\epsilon_{it}|\epsilon_{is}, \mathbf{X}] = E[\epsilon_{it}|\epsilon_{is}, \bar{Y}_i^B]$ for $i = j$, and moreover $E[\epsilon_{is}|\mathbf{X}] = E[\epsilon_{is}|\bar{Y}_i^B]$ while also $E[\epsilon_{is}^2|\mathbf{X}] = E[\epsilon_{is}^2|\bar{Y}_i^B]$. It is reasonably simple to confirm these statements using (A.10). Under such conditions, one may essentially follow the proof of Burlig et al. (2020), again leading to the same ANCOVA variance expression. Entirely analogous points apply to the model in Appendix S1.2.

For convenience, the matrix uses the following definitions.

$$\begin{aligned}
a_s &= \text{Var}(\epsilon_{js}) = (1-\theta)^2\sigma_v^2 + \left(1+\frac{\theta^2}{m}\right)\sigma_\omega^2 - \theta\text{Cov}(\omega_{js}, \bar{\omega}_j^B) + \frac{\theta^2(m-1)}{m}\psi^B \\
b_s &= \text{Cov}(\epsilon_{js}, \bar{Y}_j^B) = \text{Cov}(\omega_{js}, \bar{\omega}_j^B) - \psi^X \\
c &= \text{Var}(\bar{Y}_k^B) = \frac{1}{m}(\sigma_\delta^2 + \sigma_\omega^2 + m\sigma_v^2 + (m-1)\psi^B) \text{ for all } k \\
d &= \text{Cov}(\bar{Y}_k^B, \bar{Y}_l^B) = \frac{\sigma_\delta^2}{m} \text{ for all } k \neq l
\end{aligned}$$

Notice that $\sum_{s=1}^r b_s = 0$ for any s. Furthermore, when $i \neq j$,

$$\mathbf{\Sigma}^{\mathbf{xy}} = \begin{pmatrix} 0 & 0 & b_t & 0 & \cdots & 0 \end{pmatrix} \tag{A.12}$$

where $b_t = \text{Cov}(\epsilon_{it}, \bar{Y}_i^B) = \text{Cov}(\omega_{it}, \bar{\omega}_i^B) - \psi^X$, so $\sum_{t=1}^r b_t = 0$ for any t. For $i = j$, we instead have

$$\mathbf{\Sigma}^{\mathbf{xy}} = \begin{pmatrix} e_{ts} & b_t & 0 & \cdots & 0 \end{pmatrix} \tag{A.13}$$

where

$$\begin{aligned}
e_{ts} &= \text{Cov}(\epsilon_{it}, \epsilon_{is}) \\
&= (1-\theta)^2\sigma_v^2 + \frac{\theta^2}{m}\sigma_\omega^2 + \frac{\theta^2(m-1)}{m}\psi^B + \text{Cov}(\omega_{it}, \omega_{is}) - \theta\text{Cov}(\omega_{it}, \bar{\omega}_i^B) - \theta\text{Cov}(\omega_{is}, \bar{\omega}_i^B)
\end{aligned}$$

which we may also note implies

$$\begin{aligned}
\sum_{t=1}^r e_{ts} = r\Bigg(&(1-\theta)^2\sigma_v^2 + \left(\frac{\theta^2}{m} + \frac{1}{r}\right)\sigma_\omega^2 + \frac{\theta^2(m-1)}{m}\psi^B + \frac{\sum_{p\neq s}\text{Cov}(\omega_{ip}, \omega_{is})}{r} \\
&- \theta\psi^X - \theta\text{Cov}(\omega_{is}, \bar{\omega}_i^B)\Bigg) \\
\equiv\ & r\bar{e}_s
\end{aligned}$$

Our next objective is to invert the matrix (A.11), again using partitioning result (A.6). Note that because nearly all elements of covariance vectors (A.12) and (A.13) are zero, we need calculate only the topmost three rows of $(\mathbf{\Sigma}^{\mathbf{yy}})^{-1}$.

First, we use Lemma 1 to calculate

$$(\mathbf{\Sigma}_{22}^{\mathbf{yy}})^{-1} = \frac{1}{md(m+r\theta^2)} \begin{pmatrix} m+(r-1)\theta^2 & -\theta^2 & \cdots & -\theta^2 \\ -\theta^2 & m+(r-1)\theta^2 & \cdots & -\theta^2 \\ \vdots & \vdots & \ddots & \vdots \\ -\theta^2 & -\theta^2 & \cdots & m+(r-1)\theta^2 \end{pmatrix}$$

implying that the top-left partition of $(\boldsymbol{\Sigma}^{\mathbf{yy}})^{-1}$ is the inverse of

$$\boldsymbol{\Sigma}_{11}^{\mathbf{yy}} - \boldsymbol{\Sigma}_{12}^{\mathbf{yy}}(\boldsymbol{\Sigma}_{22}^{\mathbf{yy}})^{-1}\boldsymbol{\Sigma}_{21}^{\mathbf{yy}} = \begin{pmatrix} a_s & b_s & 0 & \cdots & 0 \\ b_s & c - \frac{r\theta^2 d}{m+r\theta^2} & \frac{md}{m+r\theta^2} & \cdots & \frac{md}{m+r\theta^2} \\ 0 & \frac{md}{m+r\theta^2} & c - \frac{r\theta^2 d}{m+r\theta^2} & \cdots & \frac{md}{m+r\theta^2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \frac{md}{m+r\theta^2} & \frac{md}{m+r+\theta^2} & \cdots & c - \frac{r\theta^2 d}{m+r\theta^2} \end{pmatrix} \tag{A.14}$$

This matrix has determinant (use Lemma 1 to perform cofactor expansion e.g. along the first row)

$$(c-d)^{J-2}\left[a_s(c-d)\left(c-d+\frac{Jmd}{m+r\theta^2}\right) - b_s^2\left(c-d+\frac{(J-1)md}{m+r\theta^2}\right)\right] \equiv |\boldsymbol{\Sigma}| \tag{A.15}$$

Notice that this determinant does not depend on t, a fact which will prove useful below. Applying Lemma 1 once more, the inverse of $\boldsymbol{\Sigma}_{11}^{\mathbf{yy}} - \boldsymbol{\Sigma}_{12}^{\mathbf{yy}}(\boldsymbol{\Sigma}_{22}^{\mathbf{yy}})^{-1}\boldsymbol{\Sigma}_{21}^{\mathbf{yy}}$, i.e. the top-left partition of $(\boldsymbol{\Sigma}^{\mathbf{yy}})^{-1}$, is

$$(\boldsymbol{\Sigma}_{11}^{\mathbf{yy}} - \boldsymbol{\Sigma}_{12}^{\mathbf{yy}}(\boldsymbol{\Sigma}_{22}^{\mathbf{yy}})^{-1}\boldsymbol{\Sigma}_{21}^{\mathbf{yy}})^{-1} = \frac{(c-d)^{J-2}}{|\boldsymbol{\Sigma}|} \left(\begin{matrix} (c-d)\left(c-d+\frac{Jmd}{m+r\theta^2}\right) & -b_s\left(c-d+\frac{(J-1)md}{m+r\theta^2}\right) \\ -b_s\left(c-d+\frac{(J-1)md}{m+r\theta^2}\right) & a_s\left(c-d+\frac{(J-1)md}{m+r\theta^2}\right) \\ \frac{md}{m+r\theta^2}b_s & -\frac{md}{m+r\theta^2}a_s \\ \vdots & \vdots \\ \frac{md}{m+r\theta^2}b_s & -\frac{md}{m+r\theta^2}a_s \end{matrix}\right.$$

$$\left.\begin{matrix} \frac{md}{m+r\theta^2}b_s & \cdots & \frac{md}{m+r\theta^2}b_s \\ -\frac{md}{m+r\theta^2}a_s & \cdots & -\frac{md}{m+r\theta^2}a_s \\ a_s\left(c-d+\frac{(J-1)md}{m+r\theta^2}\right) - b_s^2\left(1+\frac{(J-2)md}{(c-d)(m+r\theta^2)}\right) & \cdots & \frac{md}{m+r\theta^2}\left(-a_s+\frac{b_s^2}{c-d}\right) \\ \vdots & \ddots & \vdots \\ \frac{md}{m+r\theta^2}\left(-a_s+\frac{b_s^2}{c-d}\right) & \cdots & a_s\left(c-d+\frac{(J-1)md}{m+r\theta^2}\right) - b_s^2\left(1+\frac{(J-2)md}{(c-d)(m+r\theta^2)}\right) \end{matrix}\right) \tag{A.16}$$

while the top-right partition of $(\boldsymbol{\Sigma}^{\mathbf{yy}})^{-1}$ may be calculated as

$$\begin{aligned} &-(\boldsymbol{\Sigma}_{11}^{\mathbf{yy}} - \boldsymbol{\Sigma}_{12}^{\mathbf{yy}}(\boldsymbol{\Sigma}_{22}^{\mathbf{yy}})^{-1}\boldsymbol{\Sigma}_{21}^{\mathbf{yy}})^{-1}\boldsymbol{\Sigma}_{12}^{\mathbf{yy}}(\boldsymbol{\Sigma}_{22}^{\mathbf{yy}})^{-1} \\ &= \frac{\theta(c-d)^{J-2}}{(m+r\theta^2)|\boldsymbol{\Sigma}|} \times \begin{pmatrix} -b_s(c-d) & -b_s(c-d) & \cdots & -b_s(c-d) \\ a_s(c-d) & a_s(c-d) & \cdots & a_s(c-d) \\ a_s(c-d)-b_s^2 & a_s(c-d)-b_s^2 & \cdots & a_s(c-d)-b_s^2 \\ \vdots & \vdots & & \vdots \\ a_s(c-d)-b_s^2 & a_s(c-d)-b_s^2 & \cdots & a_s(c-d)-b_s^2 \end{pmatrix} \end{aligned} \tag{A.17}$$

Combining expressions (A.16) and (A.17) with (A.12) in formula (A.10) now yields the inner expectation as

$$\begin{aligned}
E[\epsilon_{it}|\epsilon_{js},\mathbf{X}] &= \frac{b_t(c-d)^{J-2}}{|\mathbf{\Sigma}|}\Bigg(\frac{mb_s d}{m+r\theta^2}\epsilon_{js} - \frac{ma_s d}{m+r\theta^2}(\bar{Y}_j^B-\mu_\delta) \\
&+ \left(a_s\left(c-d+\frac{(J-1)md}{m+r\theta^2}\right) - b_s^2\left(1+\frac{(J-2)md}{(c-d)(m+r\theta^2)}\right)\right)(\bar{Y}_i^B-\mu_\delta) \\
&+ \frac{md}{m+r\theta^2}\left(-a_s+\frac{b_s^2}{c-d}\right)\sum_{k\neq i,j}(\bar{Y}_k^B-\mu_\delta) + \frac{\theta(a_s(c-d)-b_s^2)}{m+r\theta^2}\sum_{p=1}^{r}(\alpha_p-(1-\theta)\mu_\delta)\Bigg] \\
&= A_1^{i\neq j}\epsilon_{js} + A_2^{i\neq j}(\bar{Y}_j^B-\mu_\delta) + A_3^{i\neq j}(\bar{Y}_i^B-\mu_\delta) + A_4^{i\neq j}\sum_{k\neq i,j}(\bar{Y}_k^B-\mu_\delta) \\
&+ A_5^{i\neq j}\sum_{p=1}^{r}(\alpha_p-(1-\theta)\mu_\delta)
\end{aligned}$$

with $A_1^{i\neq j},...,A_5^{i\neq j}$ defined accordingly. Since these factors are all functions only of model parameters, it follows that the full expectation is

$$\begin{aligned}
E[\epsilon_{it}\epsilon_{js}|\mathbf{X}] = A_1^{i\neq j}E[\epsilon_{js}^2|\mathbf{X}] &+ \Big(A_2^{i\neq j}(\bar{Y}_j^B-\mu_\delta) + A_3^{i\neq j}(\bar{Y}_i^B-\mu_\delta) + A_4^{i\neq j}\sum_{k\neq i,j}(\bar{Y}_k^B-\mu_\delta) \\
&+ A_5^{i\neq j}\sum_{p=1}^{r}(\alpha_p-(1-\theta)\mu_\delta)\Big) \times E[\epsilon_{js}|\mathbf{X}]
\end{aligned} \tag{A.18}$$

However, as will become clear below, neither $E[\epsilon_{js}|\mathbf{X}] = E[\epsilon_{js}|\bar{Y}_j^B,\bar{Y}_i^B,\bar{Y}_{-i,-j}^B,\alpha_1,...,\alpha_r]$ nor $E[\epsilon_{js}^2|\mathbf{X}] = \mathrm{Var}(\epsilon_{js}|\mathbf{X}) + (E[\epsilon_{js}|\mathbf{X}])^2$ depend on t; in (A.18), only b_t does. Thus, because $\sum_{t=1}^r b_t = 0$,

$$\sum_{t=1}^{r}\sum_{s=1}^{r}E[\epsilon_{it}\epsilon_{js}|\mathbf{X}] = \sum_{s=1}^{r}\Big(\sum_{t=1}^{r}E[\epsilon_{it}\epsilon_{js}|\mathbf{X}]\Big) = 0$$

and the case of $i\neq j$ will not contribute to $\mathrm{Var}(\hat{\tau}|\mathbf{X})$ in (A.9). Moving on to the case of $i=j$, we combine (A.16) and (A.17) with (A.13) in formula (A.10). This produces a different expression for the inner expectation, namely

$$
\begin{aligned}
E[\epsilon_{it}|\epsilon_{is},\mathbf{X}] = \frac{(c-d)^{J-2}}{|\mathbf{\Sigma}|}\Bigg(&\Bigg(e_{ts}(c-d)\bigg(c-d+\frac{Jmd}{m+r\theta^2}\bigg)-b_tb_s\bigg(c-d+\frac{(J-1)md}{m+r\theta^2}\bigg)\Bigg)\epsilon_{is}\\
&+\Bigg(-e_{ts}b_s\bigg(c-d+\frac{(J-1)md}{m+r\theta^2}\bigg)+a_sb_t\bigg(c-d+\frac{(J-1)md}{m+r\theta^2}\bigg)\Bigg)(\bar{Y}_i^B-\mu_\delta)\\
&+\frac{md}{m+r\theta^2}(e_{ts}b_s-a_sb_t)\sum_{k\neq i}(\bar{Y}_k^B-\mu_\delta)\\
&+\frac{\theta(c-d)}{m+r\theta^2}(-e_{ts}b_s+a_sb_t)\sum_{p=1}^{r}(\alpha_p-(1-\theta)\mu_\delta)\Bigg)
\end{aligned}
$$

which, similarly to the $i\neq j$ case, implies that

$$
\begin{aligned}
\sum_{t=1}^{r}E[\epsilon_{it}\epsilon_{is}|\mathbf{X}] = \frac{r(c-d)^{J-2}}{|\mathbf{\Sigma}|}\Bigg(&\bar{e}_s(c-d)\bigg(c-d+\frac{Jmd}{m+r\theta^2}\bigg)E[\epsilon_{is}^2|\mathbf{X}]\\
&+\Bigg(-\bar{e}_sb_s\bigg(c-d+\frac{(J-1)md}{m+r\theta^2}\bigg)(\bar{Y}_i^B-\mu_\delta)+\frac{md}{m+r\theta^2}\bar{e}_sb_s\sum_{k\neq i}(\bar{Y}_k^B-\mu_\delta)\\
&-\frac{\theta(c-d)}{m+r\theta^2}\bar{e}_sb_s\sum_{p=1}^{r}(\alpha_p-(1-\theta)\mu_\delta)\Bigg)E[\epsilon_{is}|\mathbf{X}]\Bigg)\\
&= r\Big(A_1E[\epsilon_{is}^2|\mathbf{X}]+\Big(A_2(\bar{Y}_i^B-\mu_\delta)+A_3\sum_{k\neq i}(\bar{Y}_k^B-\mu_\delta)\\
&+A_4\sum_{p=1}^{r}(\alpha_p-(1-\theta)\mu_\delta)\Big)E[\epsilon_{is}|\mathbf{X}]\Big) \qquad \text{(A.19)}
\end{aligned}
$$

with $A_1,...,A_4$ defined accordingly. The next step is to calculate the expectation $E[\epsilon_{is}|\mathbf{X}] = E[\epsilon_{is}|\bar{Y}_i^B,\bar{Y}_{-i}^B,\alpha_1,...,\alpha_r]$, and this may again be done using formula (A.10). Note first that the appropriate covariance matrix of conditioning variables, which has dimension $J+r$, is now

$$
\hat{\boldsymbol{\Sigma}}^{\mathbf{yy}} = \left(\begin{array}{cccc:cccc}
c & d & \cdots & d & -\theta d & -\theta d & \cdots & -\theta d \\
d & c & \cdots & d & -\theta d & -\theta d & \cdots & -\theta d \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & & \vdots \\
d & d & \cdots & c & -\theta d & -\theta d & \cdots & -\theta d \\
\hdashline
-\theta d & -\theta d & \cdots & -\theta d & (m+\theta^2)d & \theta^2 d & \cdots & \theta^2 d \\
-\theta d & -\theta d & \cdots & -\theta d & \theta^2 d & (m+\theta^2)d & \cdots & \theta^2 d \\
\vdots & \vdots & & \vdots & \vdots & \vdots & \ddots & \vdots \\
-\theta d & -\theta d & \cdots & -\theta d & \theta^2 d & \theta^2 d & \cdots & (m+\theta^2)d
\end{array}\right)
\equiv \begin{pmatrix} \hat{\boldsymbol{\Sigma}}^{\mathbf{yy}}_{11} & \hat{\boldsymbol{\Sigma}}^{\mathbf{yy}}_{12} \\ \hat{\boldsymbol{\Sigma}}^{\mathbf{yy}}_{21} & \hat{\boldsymbol{\Sigma}}^{\mathbf{yy}}_{22} \end{pmatrix}
$$

As was the case for $\boldsymbol{\Sigma}^{\mathbf{yy}}$, the dashed lines partition this matrix into four submatrices, and result (A.6), along with Lemma 1, may be used to invert it. Since $\hat{\boldsymbol{\Sigma}}^{\mathbf{xy}} = \begin{pmatrix} b_s & 0 & \cdots & 0 \end{pmatrix}$, we will require only the first line of $(\hat{\boldsymbol{\Sigma}}^{\mathbf{yy}})^{-1}$. Now, since $\hat{\boldsymbol{\Sigma}}^{\mathbf{yy}}_{22} = \boldsymbol{\Sigma}^{\mathbf{yy}}_{22}$, we may directly calculate

$$
\hat{\boldsymbol{\Sigma}}^{\mathbf{yy}}_{11} - \hat{\boldsymbol{\Sigma}}^{\mathbf{yy}}_{12}(\hat{\boldsymbol{\Sigma}}^{\mathbf{yy}}_{22})^{-1}\hat{\boldsymbol{\Sigma}}^{\mathbf{yy}}_{21} = \begin{pmatrix}
c - \frac{r\theta^2 d}{m+r\theta^2} & \frac{md}{m+r\theta^2} & \cdots & \frac{md}{m+r\theta^2} \\
\frac{md}{m+r\theta^2} & c - \frac{r\theta^2 d}{m+r\theta^2} & \cdots & \frac{md}{m+r\theta^2} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{md}{m+r\theta^2} & \frac{md}{m+r+\theta^2} & \cdots & c - \frac{r\theta^2 d}{m+r\theta^2}
\end{pmatrix} \tag{A.20}
$$

which is identical to (A.14), except that the first row and column are not present. By Lemma 1, the determinant of matrix (A.20) is

$$
(c-d)^{J-1}\Big(c - d + \frac{Jmd}{m+r\theta^2}\Big) \equiv |\hat{\boldsymbol{\Sigma}}| \tag{A.21}
$$

while its inverse, forming the top-left partition of $(\hat{\boldsymbol{\Sigma}}^{\mathbf{yy}})^{-1}$, is

$$
(\hat{\boldsymbol{\Sigma}}^{\mathbf{yy}}_{11} - \hat{\boldsymbol{\Sigma}}^{\mathbf{yy}}_{12}(\hat{\boldsymbol{\Sigma}}^{\mathbf{yy}}_{22})^{-1}\hat{\boldsymbol{\Sigma}}^{\mathbf{yy}}_{21})^{-1} = \frac{(c-d)^{J-2}}{|\hat{\boldsymbol{\Sigma}}|}\begin{pmatrix}
c - d + \frac{(J-1)md}{m+r\theta^2} & \cdots & -\frac{md}{m+r\theta^2} \\
\vdots & \ddots & \vdots \\
-\frac{md}{m+r\theta^2} & \cdots & c - d + \frac{(J-1)md}{m+r\theta^2}
\end{pmatrix} \tag{A.22}
$$

Here, as usual, all (off)diagonal elements equal the same value. Finally, the top-right partition of $(\hat{\boldsymbol{\Sigma}}_{\mathbf{yy}})^{-1}$ is

$$-(\hat{\boldsymbol{\Sigma}}_{11}^{\mathbf{yy}} - \hat{\boldsymbol{\Sigma}}_{12}^{\mathbf{yy}}(\hat{\boldsymbol{\Sigma}}_{22}^{\mathbf{yy}})^{-1}\hat{\boldsymbol{\Sigma}}_{21}^{\mathbf{yy}})^{-1}\hat{\boldsymbol{\Sigma}}_{12}^{\mathbf{yy}}(\hat{\boldsymbol{\Sigma}}_{22}^{\mathbf{yy}})^{-1} = \frac{\theta(c-d)^{J-1}}{(m+r\theta^2)|\hat{\boldsymbol{\Sigma}}|}\begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix} \tag{A.23}$$

Application of formula (A.10) now yields

$$\begin{aligned} E[\epsilon_{is}|\mathbf{X}] &= \frac{b_s(c-d)^{J-2}}{|\hat{\boldsymbol{\Sigma}}|}\Bigg(\left(c-d+\frac{(J-1)md}{m+r\theta^2}\right)(\bar{Y}_i^B-\mu_\delta) - \frac{md}{m+r\theta^2}\sum_{k\neq i}(\bar{Y}_k^B-\mu_\delta) \\ &\quad + \frac{\theta(c-d)}{m+r\theta^2}\sum_{p=1}^{r}(\alpha_p-(1-\theta)\mu_\delta)\Bigg) \\ &= B_1(\bar{Y}_i^B-\mu_\delta) + B_2\sum_{k\neq i}(\bar{Y}_k^B-\mu_\delta) + B_3\sum_{p=1}^{r}(\alpha_p-(1-\theta)\mu_\delta) \end{aligned} \tag{A.24}$$

where B_1, B_2 and B_3 are defined accordingly. Note that no part of (A.24) depends on t.

Next, to evaluate $E[\epsilon_{is}^2|\mathbf{X}]$ we need to calculate $\mathrm{Var}(\epsilon_{is}|\mathbf{X}) = \mathrm{Var}(\epsilon_{is}|\bar{Y}_i^B, \bar{Y}_{-i}^B, \alpha_1, .., \alpha_r)$ as well. Again, because all variables involved are normally distributed, this may be done by the following conditional-variance formula:

$$\mathrm{Var}(x|\mathbf{y}) = \sigma_x^2 - \boldsymbol{\Sigma}^{\mathbf{xy}}(\boldsymbol{\Sigma}^{\mathbf{yy}})^{-1}(\boldsymbol{\Sigma}^{\mathbf{xy}})' \tag{A.25}$$

where σ_x^2 is the unconditional variance of x and all other quantities are as defined in (A.10). Here, $\sigma_x^2 = a_s$; combining this fact with $\hat{\boldsymbol{\Sigma}}^{\mathbf{xy}}$, (A.22) and (A.23) in accordance with the above formula yields

$$\mathrm{Var}(\epsilon_{is}|\mathbf{X}) = \frac{|\boldsymbol{\Sigma}|}{|\hat{\boldsymbol{\Sigma}}|} \tag{A.26}$$

which also does not depend on t. Finally, inserting (A.26) and (A.24) into (A.19) and collecting terms, we find that the summed full expectation is

$$\begin{aligned}\sum_{t=1}^{r} E[\epsilon_{it}\epsilon_{is}|\mathbf{X}] = r\Bigg(& A_1\frac{|\mathbf{\Sigma}|}{|\hat{\mathbf{\Sigma}}|} + B_1(A_1B_1 + A_2)(\bar{Y}_i^B - \mu_\delta)^2 + B_2(A_1B_2 + A_3)\Big(\sum_{k\neq i}(\bar{Y}_k^B - \mu_\delta)\Big)^2 \\
& + B_3(A_1B_3 + A_4)\Big(\sum_{p=1}^{r}(\alpha_p - (1-\theta)\mu_\delta)\Big)^2 \\
& + (B_1(A_1B_2 + A_3) + B_2(A_1B_1 + A_2))(\bar{Y}_i^B - \mu_\delta)\sum_{k\neq i}(\bar{Y}_k^B - \mu_\delta) \\
& + (B_1(A_1B_3 + A_4) + B_3(A_1B_1 + A_2))(\bar{Y}_i^B - \mu_\delta)\sum_{p=1}^{r}(\alpha_p - (1-\theta)\mu_\delta) \\
& + (B_2(A_1B_3 + A_4) + B_3(A_1B_2 + A_3))\sum_{k\neq i}(\bar{Y}_k^B - \mu_\delta)\sum_{p=1}^{r}(\alpha_p - (1-\theta)\mu_\delta)\Bigg)\end{aligned}$$

The first term within the bracket is

$$A_1\frac{|\mathbf{\Sigma}|}{|\hat{\mathbf{\Sigma}}|} = \bar{e}_s$$

and moreover $A_1B_1 + A_2 = A_1B_2 + A_3 = A_1B_3 + A_4 = 0$, so the conditional expectation, summed over all t and s, reduces to

$$\begin{aligned}\sum_{t=1}^{r}\sum_{s=1}^{r} E[\epsilon_{it}\epsilon_{is}|\mathbf{X}] &= \sum_{s=1}^{r} r\bar{e}_s \\
&= r^2\left((1-\theta)^2\sigma_v^2 + \left(\frac{\theta^2}{m} + \frac{1}{r}\right)\sigma_\omega^2 + \frac{\theta^2(m-1)}{m}\psi^B + \frac{r-1}{r}\psi^A - 2\theta\psi^X\right) \\
&\equiv r^2\bar{e}\end{aligned}$$

As a result, equation (A.9) reduces to

$$\begin{aligned}\mathrm{Var}(\hat{\tau}|\mathbf{X}) = \frac{\bar{e}}{J^2Z^2}\Bigg(& \frac{1}{P^2}\sum_{i=1}^{PJ}(Z - PJ(\bar{Y}_T^B - \bar{Y}_C^B)(\bar{Y}_i^B - \bar{Y}_T^B))^2 \\
& + \frac{1}{(1-P)^2}\sum_{i=PJ+1}^{J}(-Z - (1-P)J(\bar{Y}_T^B - \bar{Y}_C^B)(\bar{Y}_i^B - \bar{Y}_C^T))^2\Bigg)\end{aligned}$$

which may be further simplified to yield

$$\begin{aligned}\mathrm{Var}(\hat{\tau}|\mathbf{X}) = & \left(\frac{1}{P(1-P)J} + \frac{(\bar{Y}_T^B - \bar{Y}_C^B)^2}{Z}\right) \\
& \times\left((1-\theta)^2\sigma_v^2 + \left(\frac{\theta^2}{m} + \frac{1}{r}\right)\sigma_\omega^2 + \frac{\theta^2(m-1)}{m}\psi^B + \frac{r-1}{r}\psi^A - 2\theta\psi^X\right)\end{aligned}$$

This is equation (1.1) in the main text, so we are done.

A.2 Time shocks not included in ANCOVA regression

Now, consider instead the ANCOVA regression model

$$Y_{it} = \alpha + \tilde{\tau} D_i + \theta \bar{Y}_i^B + \epsilon_{it} \tag{A.27}$$

where α is an intercept term and all other variables and coefficients are defined as in (A.2); this regression model, which does not account for time shocks, is identical to that analysed in Burlig et al. (2020); although, of course, the assumed DGP (A.1) is not.

Deriving the variance of the regression estimator $\hat{\tau}$ for model (A.27) follows much the same steps as the analysis in Appendix S1.1. First, projection of the outcome in (A.1) on the regressor matrix $\mathbf{X}$ corresponding to regression (A.27) yields projection coefficients $\alpha = \bar{\delta}^A - \theta\bar{\delta}^B$, where $\bar{\delta}^A = (1/r)\sum_{t=1}^{r} \delta_t$; $\tilde{\tau} = \tau$; and θ again equal to (A.3). As a result, residuals are now a direct function of the time shocks, since

$$\epsilon_{it} = \delta_t - \bar{\delta}^A + v_i + \omega_{it} - \theta(v_i + \bar{\omega}_i^B)$$

However, this makes perhaps surprisingly little difference for the calculations: in particular, note that ϵ_{it} is uncorrelated with α. Also notice that $E[\epsilon_{it}] = 0$.

Next, we will again calculate the ANCOVA variance by sandwich formula (A.4). Since $\mathbf{X}$ is now an $rJ \times 3$ matrix like that considered in Burlig et al. (2020), we may simply follow their initial calculation steps as far as equation (A.9) above. The next task, as in Appendix S1.1, is to evaluate conditional means $E[\epsilon_{it}\epsilon_{js}|\mathbf{X}]$. With time dummies no longer included in the regression, we may write any such quantity for which $i \neq j$ as

$$E[\epsilon_{it}\epsilon_{js}|\mathbf{X}] = E[\epsilon_{js}E[\epsilon_{it}|\epsilon_{js}, \bar{Y}_j^B, \bar{Y}_i^B, \bar{Y}_{-i,-j}^B, \alpha]|\bar{Y}_j^B, \bar{Y}_i^B, \bar{Y}_{-i,-j}^B, \alpha]$$

with the conditional mean for $i = j$ suitably adjusted. These means can again be calculated using formula (A.10). The covariance matrix associated with the inner (ϵ_{it}) expectation is

$$\mathbf{\Sigma}^{\mathbf{yy}} = \left(\begin{array}{cccccc:c} a_s + (r+1)d & b_s & 0 & \cdots & 0 & & 0 \\ b_s & c & d & \cdots & d & & -\theta d \\ 0 & d & c & \cdots & d & & -\theta d \\ \vdots & \vdots & \vdots & \ddots & \vdots & & \vdots \\ 0 & d & d & \cdots & c & & -\theta d \\ \hdashline 0 & -\theta d & -\theta d & \cdots & -\theta d & & \left(\frac{m}{r} + \theta^2\right) d \end{array}\right) \tag{A.28}$$

where all parameters take the same values as in Appendix S1.1 above, again implying $\sum_{s=1}^{r} b_s = 0$ for any s. Note that under the partitioning used, $\mathbf{\Sigma}_{22}^{\mathbf{yy}}$ is just the scalar

$(m/r+\theta^2)d$, making that partition straightforward to invert. In any case, when $i \neq j$, the relevant covariance vector is

$$\mathbf{\Sigma^{xy}} = \begin{pmatrix} e_{ts}^{i\neq j} & 0 & b_t & 0 & \cdots & 0 \end{pmatrix} \tag{A.29}$$

where $e_{ts}^{i\neq j} = \text{Cov}(\delta_t, \delta_s) - (\sigma_\delta^2/r)$, implying that $\sum_{t=1}^{r} e_{ts}^{i\neq j} = 0$. For $i = j$, the corresponding vector is

$$\mathbf{\Sigma^{xy}} = \begin{pmatrix} e_{ts}^{i\neq j} + e_{ts} & b_t & 0 & \cdots & 0 \end{pmatrix} \tag{A.30}$$

where $e_{ts} = \text{Cov}(\epsilon_{it}, \epsilon_{is}) - e_{ts}^{i\neq j}$ is as calculated in Appendix S1.1. Clearly, again we need consider only the first few rows of $(\mathbf{\Sigma^{yy}})^{-1}$.

Now,

$$\mathbf{\Sigma_{11}^{yy}} - \mathbf{\Sigma_{12}^{yy}}(\mathbf{\Sigma_{22}^{yy}})^{-1}\mathbf{\Sigma_{21}^{yy}} = \begin{pmatrix} a_s + (r+1)d & b_s & 0 & \cdots & 0 \\ b_s & c - \frac{r\theta^2 d}{m+r\theta^2} & \frac{md}{m+r\theta^2} & \cdots & \frac{md}{m+r\theta^2} \\ 0 & \frac{md}{m+r\theta^2} & c - \frac{r\theta^2 d}{m+r\theta^2} & \cdots & \frac{md}{m+r\theta^2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \frac{md}{m+r\theta^2} & \frac{md}{m+r+\theta^2} & \cdots & c - \frac{r\theta^2 d}{m+r\theta^2} \end{pmatrix}$$

with determinant

$$(c-d)^{J-2}\left((a_s+(r+1)d)(c-d)\left(c-d+\frac{Jmd}{m+r\theta^2}\right) - b_s^2\left(c-d+\frac{(J-1)md}{m+r\theta^2}\right)\right) \equiv |\mathbf{\Sigma}|$$

and symmetric inverse

$$
(\boldsymbol{\Sigma}_{11}^{\mathbf{yy}} - \boldsymbol{\Sigma}_{12}^{\mathbf{yy}}(\boldsymbol{\Sigma}_{22}^{\mathbf{yy}})^{-1}\boldsymbol{\Sigma}_{21}^{\mathbf{yy}})^{-1} = \frac{(c-d)^{J-2}}{|\boldsymbol{\Sigma}|}
\begin{pmatrix}
(c-d)\Big(c-d+\frac{Jmd}{m+r\theta^2}\Big) & -b_s\Big(c-d+\frac{(J-1)md}{m+r\theta^2}\Big) & \frac{md}{m+r\theta^2}b_s & \cdots & \frac{md}{m+r\theta^2}b_s \\
-b_s\Big(c-d+\frac{(J-1)md}{m+r\theta^2}\Big) & (a_s+(r+1)d)\Big(c-d+\frac{(J-1)md}{m+r\theta^2}\Big) & -\frac{md}{m+r\theta^2}(a_s+(r+1)d) & \cdots & -\frac{md}{m+r\theta^2}(a_s+(r+1)d) \\
\frac{md}{m+r\theta^2}b_s & -\frac{md}{m+r\theta^2}(a_s+(r+1)d) & (a_s+(r+1)d)\Big(c-d+\frac{(J-1)md}{m+r\theta^2}\Big)-b_s^2\Big(1+\frac{(J-2)md}{(c-d)(m+r\theta^2)}\Big) & \cdots & \frac{md}{m+r\theta^2}\Big(-a_s-(r+1)d+\frac{b_s^2}{c-d}\Big) \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\frac{md}{m+r\theta^2}b_s & -\frac{md}{m+r\theta^2}(a_s+(r+1)d) & \frac{md}{m+r\theta^2}\Big(-a_s-(r+1)d+\frac{b_s^2}{c-d}\Big) & \cdots & (a_s+(r+1)d)\Big(c-d+\frac{(J-1)md}{m+r\theta^2}\Big)-b_s^2\Big(1+\frac{(J-2)md}{(c-d)(m+r\theta^2)}\Big)
\end{pmatrix} \tag{A.31}
$$

forming the top-left partition of $(\boldsymbol{\Sigma}^{\mathbf{yy}})^{-1}$; the top-right partition is

$$
-(\boldsymbol{\Sigma}_{11}^{\mathbf{yy}} - \boldsymbol{\Sigma}_{12}^{\mathbf{yy}}(\boldsymbol{\Sigma}_{22}^{\mathbf{yy}})^{-1}\boldsymbol{\Sigma}_{21}^{\mathbf{yy}})^{-1}\boldsymbol{\Sigma}_{12}^{\mathbf{yy}}(\boldsymbol{\Sigma}_{22}^{\mathbf{yy}})^{-1} = \frac{r\theta(c-d)^{J-2}}{(m+r\theta^2)|\boldsymbol{\Sigma}|} \times
\begin{pmatrix}
-b_s(c-d) \\
(a_s+(r+1)d)(c-d) \\
(a_s+(r+1)d)(c-d)-b_s^2 \\
\vdots \\
(a_s+(r+1)d)(c-d)-b_s^2
\end{pmatrix} \tag{A.32}
$$

Combining these expressions with (A.29) produces

$$
E[\epsilon_{it}|\epsilon_{js},\mathbf{X}] = \frac{(c-d)^{J-2}}{|\boldsymbol{\Sigma}|}\Bigg(\Bigg((c-d)\bigg(c-d+\frac{Jmd}{m+r\theta^2}\bigg)e_{ts}^{i\neq j} + \frac{mb_sd}{m+r\theta^2}b_t\Bigg)\epsilon_{js}
$$

$$
+\Bigg(-b_s\bigg(c-d+\frac{(J-1)md}{m+r\theta^2}\bigg)e_{ts}^{i\neq j} - (a_s+(r+1)d)\frac{md}{m+r\theta^2}b_t\Bigg)(\bar{Y}_j^B-\mu_\delta)
$$

$$+\left(\frac{mb_sd}{m+r\theta^2}e_{ts}^{i\neq j}+\left((a_s+(r+1)d)\left(c-d+\frac{(J-1)md}{m+r\theta^2}\right)-b_s^2\left(1+\frac{(J-2)md}{(c-d)(m+r\theta^2)}\right)\right)b_t\right)$$
$$\times(\bar{Y}_i^B-\mu_\delta)$$
$$+\left(\frac{mb_sd}{m+r\theta^2}e_{ts}^{i\neq j}+\frac{md}{m+r\theta^2}\left(-a_s-(r+1)d+\frac{b_s^2}{c-d}\right)b_t\right)\sum_{k\neq i,j}(\bar{Y}_k^B-\mu_\delta)$$
$$+\left(-\frac{r\theta b_sd(c-d)}{m+r\theta^2}e_{ts}^{i\neq j}+\frac{r\theta}{m+r\theta^2}((a_s+(r+1)d)(c-d)-b_s^2)\right)(\alpha-(1-\theta)\mu_\delta)\Bigg)$$
$$=A_1^{i\neq j}\epsilon_{js}+A_2^{i\neq j}(\bar{Y}_j^B-\mu_\delta)+A_3^{i\neq j}(\bar{Y}_i^B-\mu_\delta)+A_4^{i\neq j}\sum_{k\neq i,j}(\bar{Y}_k^B-\mu_\delta)+A_5^{i\neq j}(\alpha-(1-\theta)\mu_\delta)$$

with $A_1^{i\neq j},\ldots,A_5^{i\neq j}$ defined accordingly. It follows that

$$E[\epsilon_{it}\epsilon_{js}|\mathbf{X}]=A_1^{i\neq j}E[\epsilon_{js}^2|\mathbf{X}]+\Big(A_2^{i\neq j}(\bar{Y}_j^B-\mu_\delta)+A_3^{i\neq j}(\bar{Y}_i^B-\mu_\delta)+A_4^{i\neq j}\sum_{k\neq i,j}(\bar{Y}_k^B-\mu_\delta)$$
$$+A_5^{i\neq j}(\alpha-(1-\theta)\mu_\delta)\Big)\times E[\epsilon_{js}|\mathbf{X}]$$

but because only $e_{ts}^{i\neq j}$ and b_t depend on t in these expressions, and because $\sum_{t=1}^r e_{ts}^{i\neq j}=\sum_{t=1}^r b_t=0$, we have

$$\sum_{t=1}^r\sum_{s=1}^r E[\epsilon_{it}\epsilon_{js}|\mathbf{X}]=\sum_{s=1}^r\Big(\sum_{t=1}^r E[\epsilon_{it}\epsilon_{js}|\mathbf{X}]\Big)=0$$

Thus, as in Appendix S1.1, the case of $i\neq j$ will not contribute to the variance of the treatment estimator.

When instead $i=j$, combining (A.31) and (A.32) with (A.30) in formula (A.10) produces

$$E[\epsilon_{it}|\epsilon_{js},\mathbf{X}]=\frac{(c-d)^{J-2}}{|\mathbf{\Sigma}|}$$
$$\times\left(\left((e_{ts}+e_{ts}^{i\neq j})(c-d)\left(c-d+\frac{Jmd}{m+r\theta^2}\right)-b_tb_s\left(c-d+\frac{(J-1)md}{m+r\theta^2}\right)\right)\epsilon_{is}\right.$$
$$+\left(-(e_{ts}+e_{ts}^{i\neq j})b_s\left(c-d+\frac{(J-1)md}{m+r\theta^2}\right)+(a_s+(r+1)d)b_t\left(c-d+\frac{(J-1)md}{m+r\theta^2}\right)\right)$$
$$\times(\bar{Y}_i^B-\mu_\delta)$$
$$+\left(\frac{md}{m+r\theta^2}(b_s(e_{ts}+e_{ts}^{i\neq j})-(a_s+(r+1)d)b_t)\right)\sum_{k\neq i}(\bar{Y}_k^B-\mu_\delta)$$
$$+\left(\frac{r\theta(c-d)}{m+r\theta^2}(-b_s(e_{ts}+e_{ts}^{i\neq j})+(a_s+(r+1)d)b_t)\right)(\alpha-(1-\theta)\mu_\delta)\Bigg)$$

where, because $\sum_{t=1}^{r} b_t = \sum_{t=1}^{r} e_{ts}^{i\neq j} = 0$,

$$
\begin{aligned}
\sum_{t=1}^{r} E[\epsilon_{it}\epsilon_{is}|\mathbf{X}] = {} & \frac{r(c-d)^{J-2}}{|\boldsymbol{\Sigma}|}\Bigg(\bar{e}_s(c-d)\left(c-d+\frac{Jmd}{m+r\theta^2}\right)E[\epsilon_{is}^2|\mathbf{X}] \\
& + \Bigg(-\bar{e}_s b_s\left(c-d+\frac{(J-1)md}{m+r\theta^2}\right)(\bar{Y}_i^B-\mu_\delta)+\frac{md}{m+r\theta^2}\bar{e}_s b_s\sum_{k\neq i}(\bar{Y}_i^B-\mu_\delta) \\
& - \frac{r\theta(c-d)}{m+r\theta^2}\bar{e}_s b_s(\alpha-(1-\theta)\mu_\delta)\Bigg)E[\epsilon_{is}|\mathbf{X}]\Bigg) \\
= {} & r\Bigg(A_1 E[\epsilon_{is}^2|\mathbf{X}] + \Bigg(A_2(\bar{Y}_i^B-\mu_\delta)+A_3\Bigg(\sum_{k\neq i}(\bar{Y}_k^B-\mu_\delta)\Bigg)+A_4(\alpha-(1-\theta)\mu_\delta)\Bigg) \\
& \times E[\epsilon_{is}|\mathbf{X}]\Bigg) \qquad \text{(A.33)}
\end{aligned}
$$

with $A_1, ..., A_4$ defined accordingly.

As for $E[\epsilon_{is}|\mathbf{X}] = E[\epsilon_{is}|\bar{Y}_i^B, \bar{Y}_{-i}^B, \alpha_1, ..., \alpha_r]$, the covariance matrix of conditioning variables has dimension $J+1$ and is given by

$$
\hat{\boldsymbol{\Sigma}}^{\mathbf{yy}} = \left(\begin{array}{cccc:c}
c & d & \cdots & d & -\theta d \\
d & c & \cdots & d & -\theta d \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
d & d & \cdots & c & -\theta d \\
\hdashline
-\theta d & -\theta d & \cdots & -\theta d & (m+\theta^2)d
\end{array}\right)
$$

while the covariance vector is $\hat{\boldsymbol{\Sigma}}^{\mathbf{xy}} = \begin{pmatrix} b_s & 0 & \cdots & 0 \end{pmatrix}$, so we need calculate only the upper part (first line) of $(\hat{\boldsymbol{\Sigma}}_{\mathbf{yy}})^{-1}$.

Using partition result (A.6), we find that $\hat{\boldsymbol{\Sigma}}_{11}^{\mathbf{yy}} - \hat{\boldsymbol{\Sigma}}_{12}^{\mathbf{yy}}(\hat{\boldsymbol{\Sigma}}_{22}^{\mathbf{yy}})^{-1}\hat{\boldsymbol{\Sigma}}_{21}^{\mathbf{yy}}$ is again equal to (A.20). It follows that the determinant $|\boldsymbol{\Sigma}|$ of that matrix is given by (A.21); and its inverse, forming the top-left portion of $(\hat{\boldsymbol{\Sigma}}_{\mathbf{yy}})^{-1}$, is (A.22). The top-right partition is

$$
-(\hat{\boldsymbol{\Sigma}}_{11}^{\mathbf{yy}} - \hat{\boldsymbol{\Sigma}}_{12}^{\mathbf{yy}}(\hat{\boldsymbol{\Sigma}}_{22}^{\mathbf{yy}})^{-1}\hat{\boldsymbol{\Sigma}}_{21}^{\mathbf{yy}})^{-1}\hat{\boldsymbol{\Sigma}}_{12}^{\mathbf{yy}}(\hat{\boldsymbol{\Sigma}}_{22}^{\mathbf{yy}})^{-1} = \frac{r\theta(c-d)^{J-1}}{(m+r\theta^2)|\hat{\boldsymbol{\Sigma}}|}\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}
$$

Applying formula (A.10) then yields

$$\begin{aligned}
E[\epsilon_{is}|\mathbf{X}] &= \frac{b_s(c-d)^{J-2}}{|\hat{\mathbf{\Sigma}}|}\Bigg(\left(c-d+\frac{(J-1)md}{m+r\theta^2}\right)(\bar{Y}_i^B-\mu_\delta)-\frac{md}{m+r\theta^2}\sum_{k\neq i}(\bar{Y}_k^B-\mu_\delta) \\
&\quad + \frac{r\theta(c-d)}{m+r\theta^2}\sum_{p=1}^{r}(\alpha_p-(1-\theta)\mu_\delta)\Bigg) \\
&= B_1(\bar{Y}_i^B-\mu_\delta)+B_2\sum_{k\neq j}(\bar{Y}_k^B-\mu_\delta)+B_3\sum_{p=1}^{r}(\alpha_p-(1-\theta)\mu_\delta) \qquad \text{(A.34)}
\end{aligned}$$

where B_1, B_2 and B_3 are defined accordingly.

In addition, we apply conditional-variance formula (A.25), with $\sigma_x^2 = a_s + (r+1)d$, to compute

$$\mathrm{Var}(\epsilon_{is}|\mathbf{X}) = \frac{|\mathbf{\Sigma}|}{|\hat{\mathbf{\Sigma}}|}$$

Combining this with (A.34) and (A.33) yields the summed conditional expectation as

$$\begin{aligned}
\sum_{t=1}^{r} E[\epsilon_{it}\epsilon_{is}|\mathbf{X}] &= r\Bigg(A_1\frac{|\mathbf{\Sigma}|}{|\hat{\mathbf{\Sigma}}|}+B_1(A_1B_1+A_2)(\bar{Y}_i^B-\mu_\delta)^2+B_2(A_1B_2+A_3)\Big(\sum_{k\neq i}(\bar{Y}_k^B-\mu_\delta)\Big)^2 \\
&\quad + B_3(A_1B_3+A_4)\Big(\sum_{p=1}^{r}(\alpha_p-(1-\theta)\mu_\delta)\Big)^2 \\
&\quad + (B_1(A_1B_2+A_3)+B_2(A_1B_1+A_2))(\bar{Y}_i^B-\mu_\delta)\sum_{k\neq i}(\bar{Y}_k^B-\mu_\delta) \\
&\quad + (B_1(A_1B_3+A_4)+B_3(A_1B_1+A_2))(\bar{Y}_i^B-\mu_\delta)\sum_{p=1}^{r}(\alpha_p-(1-\theta)\mu_\delta) \\
&\quad + (B_2(A_1B_3+A_4)+B_3(A_1B_2+A_3))\sum_{k\neq i}(\bar{Y}_k^B-\mu_\delta)\sum_{p=1}^{r}(\alpha_p-(1-\theta)\mu_\delta)\Bigg)
\end{aligned}$$

which is the same expression as in Appendix S1.1, although of course the factors included differ somewhat. Nevertheless, the first term within the bracket remains

$$A_1\frac{|\mathbf{\Sigma}|}{|\hat{\mathbf{\Sigma}}|} = \bar{e}_s$$

and moreover we again find that $A_1B_1 + A_2 = A_1B_2 + A_3 = A_1B_3 + A_4 = 0$. Thus,

$$\begin{aligned}
\sum_{t=1}^{r}\sum_{s=1}^{r} E[\epsilon_{it}\epsilon_{is}|\mathbf{X}] &= \sum_{s=1}^{r} r\bar{e}_s \\
&= r^2\left((1-\theta)^2\sigma_v^2+\left(\frac{\theta^2}{m}+\frac{1}{r}\right)\sigma_\omega^2+\frac{\theta^2(m-1)}{m}\psi^B+\frac{r-1}{r}\psi^A-2\theta\psi^X\right) \\
&\equiv r^2\bar{e}
\end{aligned}$$

and all remaining steps are as in Appendix S1.1, concluding the proof.

Appendix B. Estimating an ANCOVA MDE from pre-existing data

Throughout this section, we retain model assumptions 1.1-1.5 from Appendix S1 above; this means, in particular, that time shocks remain included in the DGP. As a modification of the algorithm proposed by Burlig et al. (2020) for estimating an ANCOVA minimum detectable effect from a pre-existing data set, consider the following. Notice that Step 1 and 3 remain as originally proposed by the authors for ANCOVA: the main change is that, in Step 2, the estimating regression includes time fixed effects.

STEP 1. Determine all feasible ranges of experiments with $(m + r)$ periods, given the number of time periods in the pre-existing data set.

STEP 2. For each feasible range S:

(a) Regress the outcome variable on unit and time-period fixed effects, $Y_{it} = v_i + \delta_t + \omega_{it}$, and store the residuals. This regression includes all I available cross-sectional units, but only time periods with the specific range S.

(b) Calculate the variance of the fitted unit fixed effects $\hat{v}_i$, and store as $\tilde{\sigma}^2_{\hat{v},S}$.

(c) Calculate the variance of the stored residuals, and save as $\tilde{\sigma}^2_{\hat{\omega},S}$.

(d) For each pair of pre-treatment periods, (i.e. the first m periods in range S), calculate the the covariance between these periods' residuals. Take an unweighted average of these $m(m-1)/2$ covariances, and store as $\tilde{\psi}^B_{\hat{\omega},S}$.

(e) For each pair of post-treatment periods, (i.e. the last r periods in range S), calculate the the covariance between these periods' residuals. Take an unweighted average of these $r(r-1)/2$ covariances, and store as $\tilde{\psi}^A_{\hat{\omega},S}$.[5]

STEP 3. Calculate the average of $\tilde{\sigma}^2_{\hat{v},S}$, $\tilde{\sigma}^2_{\hat{\omega},S}$, $\tilde{\psi}^B_{\hat{\omega},S}$, and $\tilde{\psi}^A_{\hat{\omega},S}$ across all ranges S, deflating $\tilde{\sigma}^2_{\hat{\omega},S}$ by $\frac{I(m+r)-1}{I(m+r)}$ and $\tilde{\sigma}^2_{\hat{v},S}$, $\tilde{\psi}^B_{\hat{\omega},S}$, and $\tilde{\psi}^A_{\hat{\omega},S}$ by $\frac{I-1}{I}$. These averages are equal in expectation to $\sigma^2_{\hat{v}}$,$\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, and $\psi^A_{\hat{\omega}}$.

[5] Burlig et al. (2020) add an additional step estimating the residual-based across-period covariance, $\tilde{\psi}^X_{\hat{\omega},S}$. However, that step turns out to be redundant, both here and in the original procedure, since $\tilde{\psi}^X_{\hat{\omega},S}$ is not used when calculating the MDE.

STEP 4. To produce the estimated MDE, plug these values into the right-hand side of equation (A.35), given below.

$$
\begin{aligned}
MDE^{est} = (t^J_{1-\kappa} - t^J_{\alpha/2}) \times \Bigg\{ & \left(\frac{1}{P(1-P)J} + \frac{(\bar{Y}^B_T - \bar{Y}^B_C)^2}{Z} \right) \times \left(\frac{I}{I-1} \right) \\
& \times \Bigg((1-\theta^2)\sigma^2_{\hat{v}} + \left(\frac{m+\theta r}{2m^2r^2} \right) ((m+r)(m+\theta r) + (1-\theta)(mr^2 - m^2 r))\sigma^2_{\hat{\omega}} \\
& + \left(\frac{m+\theta r}{2mr^2} \right) (m-1)(m+\theta r - (1-\theta)mr)\psi^B_{\hat{\omega}} \\
& + \left(\frac{m+\theta r}{2m^2 r} \right) (r-1)(m+\theta r + (1-\theta)mr)\psi^A_{\hat{\omega}} \Bigg) \Bigg\}^{1/2}
\end{aligned}
\tag{A.35}
$$

In the above equation, $t^J_{1-\kappa}$ and $t^J_{\alpha/2}$ are suitable critical values of the t distribution, and θ is expressed in terms of the residual-based parameters as

$$
\begin{aligned}
\theta = & \frac{m(4mr\sigma^2_{\hat{v}} - (m(m-r+2) + r(r-m+2))\sigma^2_{\hat{\omega}})}{2r(2m^2\sigma^2_{\hat{v}} + (m(m+1) - r(m-1))\sigma^2_{\hat{\omega}} + m(m-1)(m+1)\psi^B_{\hat{\omega}} - r(m-1)(r-1)\psi^A_{\hat{\omega}})} \\
& + \frac{m(-m(m-1)(m-r+2)\psi^B_{\hat{\omega}} - r(r-1)(r-m+2)\psi^A_{\hat{\omega}})}{2r(2m^2\sigma^2_{\hat{v}} + (m(m+1) - r(m-1))\sigma^2_{\hat{\omega}} + m(m-1)(m+1)\psi^B_{\hat{\omega}} - r(m-1)(r-1)\psi^A_{\hat{\omega}})}
\end{aligned}
\tag{A.36}
$$

The remainder of this section of the appendix mirrors the calculations in Appendix E of Burlig et al. (2020), showing that the above modified algorithm is appropriate.

First, we claim that steps 1-3 of the algorithm yield unbiased estimates of all residual-based parameters. For all estimates except $\tilde{\sigma}^2_{\hat{v}}$, the proof is identical to that provided in Appendix E.2 of Burlig et al. (2020). Furthermore, in the estimating regression,

$$
\tilde{\sigma}^2_{\hat{v}} = \frac{1}{I} \sum_{i=1}^{I} \left(\hat{v}_i - \frac{1}{I} \sum_{i=1}^{I} \hat{v}_i \right)^2
$$

which is identical to the $\sigma^2_{\hat{v}}$ estimate obtained when time FE are not included in the estimating regression of step 2a above. The proof that $E[\tilde{\sigma}^2_{\hat{v}}] = \sigma^2_{\hat{v}}$ will therefore be identical to that provided in Appendix E.3 of Burlig et al. (2020).

Next, step 4 uses these estimates to calculate the MDE. To see why this works, we first need to express each residual-based parameter as a function of the parameters of the DGP.

For $\sigma^2_{\hat{v}}$, we note that

$$\begin{aligned}
\hat{v}_i &= \frac{1}{m+r}\sum_{t=-m+1}^{r} Y_{it} - \frac{1}{I(m+r)}\sum_{i=1}^{I}\sum_{t=-m+1}^{r} Y_{it} \\
&= v_i - \frac{1}{I}\sum_{i=1}^{I} v_i + \frac{1}{m+r}\sum_{t=-m+1}^{r}\omega_{it} - \frac{1}{I(m+r)}\sum_{i=1}^{I}\sum_{t=-m+1}^{r}\omega_{it}
\end{aligned}$$

which has variance

$$\sigma^2_{\hat{v}} = \left(\frac{I-1}{I(m+r)^2}\right)\left((m+r)^2\sigma^2_v + (m+r)\sigma^2_\omega + m(m-1)\psi^B + r(r-1)\psi^A + 2mr\psi^X\right)$$

For all other parameters, we simply repeat the calculations in Appendix E.2 of Burlig et al. (2020), yielding

$$\begin{aligned}
\sigma^2_{\hat{\omega}} &= \left(\frac{I-1}{I(m+r)^2}\right)\left((m+r)(m+r-1)\sigma^2_\omega - m(m-1)\psi^B - r(r-1)\psi^A - 2mr\psi^X\right) \\
\psi^B_{\hat{\omega}} &= \left(\frac{I-1}{I(m+r)^2}\right)\left(-(m+r)\sigma^2_\omega + (r^2+2r+m)\psi^B + r(r-1)\psi^A - 2r^2\psi^X\right) \\
\psi^A_{\hat{\omega}} &= \left(\frac{I-1}{I(m+r)^2}\right)\left(-(m+r)\sigma^2_\omega + m(m-1)\psi^B + (m^2+2m+r)\psi^A - 2m^2\psi^X\right) \\
\psi^X_{\hat{\omega}} &= \left(\frac{I-1}{I(m+r)^2}\right)\left(-(m+r)\sigma^2_\omega - r(m-1)\psi^B - m(r-1)\psi^A + 2mr\psi^X\right)
\end{aligned}$$

Comparing with the corresponding expressions in Appendix E.3 of Burlig et al. (2020), we note the single difference that all residual-based parameters $\sigma^2_{\hat{v}}, \sigma^2_{\hat{\omega}}, \psi^B_{\hat{\omega}}, \psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$ are now multiplied by $(I-1)/I$, while this was true only for $\sigma^2_{\hat{v}}$ in the original procedure. In any case, we now seek coefficients k_v, k_ω, k_B, k_A, and k_X that allow us to express the serial-correlation-robust ANCOVA variance in terms of the residual-based parameters rather than the true parameters. The coefficients will be given by any solution to the following equation:

$$\begin{aligned}
&k_v\sigma^2_{\hat{v}} + k_\omega\sigma^2_{\hat{\omega}} + k_B\psi^B_{\hat{\omega}} + k_A\psi^A_{\hat{\omega}} + k_X\psi^X_{\hat{\omega}} \\
&= (1-\theta)^2\sigma^2_v + \left(\frac{\theta^2}{m} + \frac{1}{r}\right)\sigma^2_\omega + \frac{\theta^2(m-1)}{m}\psi^B + \frac{r-1}{r}\psi^A - 2\theta\psi^X
\end{aligned}$$

This implies the equation system

$$\begin{pmatrix} k_v & k_\omega & k_B & k_A & k_X \end{pmatrix}\mathbf{\Gamma} = \begin{pmatrix} (1-\theta)^2 & \frac{m+\theta^2 r}{mr} & \frac{(m-1)\theta^2}{m} & \frac{r-1}{r} & -2\theta \end{pmatrix}$$

where

$$\mathbf{\Gamma} = \frac{I-1}{I(m+r)^2}\begin{pmatrix} (m+r)^2 & m+r & m(m-1) & r(r-1) & 2mr \\ 0 & (m+r)(m+r-1) & -m(m-1) & -r(r-1) & -2mr \\ 0 & -(m+r) & r^2+2r+m & r(r-1) & -2r^2 \\ 0 & -(m+r) & m(m-1) & m^2+2m+r & -2m^2 \\ 0 & -(m+r) & -r(m-1) & -m(r-1) & 2mr \end{pmatrix}$$

Although the equation system has infinite solutions, we follow Burlig et al. (2020) in selecting the one where $k_X = 0$. This yields

$$\begin{aligned} k_v &= \left(\frac{I}{I-1}\right)(1-\theta)^2 \\ k_\omega &= \left(\frac{I}{I-1}\right)\frac{m+\theta r}{2m^2r^2}((m+r)(m+\theta r)+(1-\theta)(mr^2-m^2r)) \\ k_B &= \left(\frac{I}{I-1}\right)\frac{m+\theta r}{2mr^2}(m-1)(m+\theta r-(1-\theta)mr) \\ k_A &= \left(\frac{I}{I-1}\right)\frac{m+\theta r}{2m^2r}(r-1)(m+\theta r+(1-\theta)mr) \\ k_X &= 0 \end{aligned}$$

which implies equation (A.35) may be used to compute the MDE. Similarly to above, the only difference between this solution and that of the original procedure is that all coefficients (rather than just k_v) now include the factor $\frac{I}{I-1}$.

Finally, we must also express θ in terms of the residual-based parameters. This requires choosing coefficients $k_v^N, k_\omega^N, k_B^N, k_A^N, k_X^N$ (corresponding to the numerator of θ) as well as $k_v^D, k_\omega^D, k_B^D, k_A^D, k_X^D$ (corresponding to the denominator) such that

$$\theta = \frac{m\sigma_v^2 + m\psi^X}{m\sigma_v^2 + \sigma_\omega^2 + (m-1)\psi^B} = \frac{k_v^N\sigma_{\hat{v}}^2 + k_\omega^N\sigma_{\hat{\omega}}^2 + k_B^N\psi_{\hat{\omega}}^B + k_A^N\psi_{\hat{\omega}}^A + k_X^N\psi_{\hat{\omega}}^X}{k_v^D\sigma_{\hat{v}}^2 + k_\omega^D\sigma_{\hat{\omega}}^2 + k_B^D\psi_{\hat{\omega}}^B + k_A^D\psi_{\hat{\omega}}^A + k_X^D\psi_{\hat{\omega}}^X}$$

For the numerator, the solution where $k_X^N = 0$ is

$$\begin{aligned}
k_v^N &= \left(\frac{I}{I-1}\right) m \\
k_\omega^N &= -\left(\frac{I}{I-1}\right) \frac{1}{4r}(m(m-r+2)+r(r-m+2)) \\
k_B^N &= -\left(\frac{I}{I-1}\right) \frac{m}{4r}(m-1)(m-r+2) \\
k_A^N &= -\left(\frac{I}{I-1}\right) \frac{1}{4}(r-1)(r-m+2) \\
k_X^N &= 0
\end{aligned}$$

For the denominator, the solution where $k_X^D = 0$ is

$$\begin{aligned}
k_v^D &= \left(\frac{I}{I-1}\right) m \\
k_\omega^D &= \left(\frac{I}{I-1}\right) \frac{1}{2m}(m(m+1)-r(m-1)) \\
k_B^D &= \left(\frac{I}{I-1}\right) \frac{1}{2}(m+1)(m-1) \\
k_A^D &= -\left(\frac{I}{I-1}\right) \frac{r}{2m}(m-1)(r-1) \\
k_X^D &= 0
\end{aligned}$$

which gives θ as equation (A.36). Again, these solutions differ from the original results only in that all coefficients (rather than just the k_v coefficients) include $I/(I-1)$.

Appendix C. Simulation details

This section describes the simulation procedure underlying each of the figures included in the main text. Each figure is based on multiple simulation sets of 10,000 data draws each. Across the sets, I vary time periods m, r, and autocorrelation parameter γ (in Figure 1 and 2 only). All other parameter values are held constant: in particular, $\alpha = 0.05$; $\beta = 0.8$; $J = 500$; $P = 1/2$; $\delta = \mu_\delta = 21$; $\sigma_\delta^2 = 10$; $\mu_v = 100$; $\sigma_v^2 = 80$; and $\sigma_\omega^2 = 10$.

C.1 Figure 1

For the upper half of the figure, the following algorithm is used; the procedure is here very nearly identical to Burlig et al. (2020) Appendix B.1.[6]

Step 1. In each simulation run, use DGP

$$Y_{it} = \delta + v_i + \omega_{it}$$

to generate a data set with $m + r = 2m$ periods and J units. To this effect, values of v_i are drawn J times, while for each i, ω_{it} is generated by the AR(1) process $\omega_{it} = \gamma\omega_{i(t-1)} + \xi_{it}$, where ξ_{it} is distributed $\mathcal{N}(0, \sigma_\omega^2(1-\gamma^2))$. To allow the data time to fully reflect the autoregressive process, a 'burn-in' duration of 100 periods is added prior to the first pre-treatment period and then discarded.

Step 2. Calculate calibrated MDEs

$$\tau^{McK} = (t_{1-\beta}^J + t_{\alpha/2}^J)\sqrt{\left(\frac{1}{P(1-P)J} + \frac{(\bar{Y}_T^B - \bar{Y}_C^B)^2}{Z}\right) \times \frac{\sigma_\omega^2}{r} \times \frac{\sigma_\omega^2 + (m+r)\sigma_v^2}{\sigma_\omega^2 + m\sigma_v^2}}$$

$$\tau^{SCR} = (t_{1-\beta}^J + t_{\alpha/2}^J)\left(\frac{1}{P(1-P)J} + \frac{(\bar{Y}_T^B - \bar{Y}_C^B)^2}{Z}\right)^{1/2}$$
$$\times \left((1-\theta)^2\sigma_v^2 + \left(\frac{\theta^2}{m} + \frac{1}{r}\right)\sigma_\omega^2 + \frac{\theta^2(m-1)}{m}\psi^B + \frac{r-1}{r}\psi^A - 2\theta\psi^X\right)^{1/2}$$

where $t_{1-\beta}^J$ and $t_{\alpha/2}^J$ are appropriate critical values of the t distribution. These expressions correspond to the ANCOVA variance given in McKenzie (2012) and equation (1.1) in the main text, respectively. $\bar{Y}_T^B$, $\bar{Y}_C^B$, and Z are calculated from the data generated in Step 1. Also, since $m = r$ in these simulations,

$$\psi^B = \psi^A = \frac{2\sigma_\omega^2}{m(m-1)}\sum_{p=1}^{m-1}(m-p)\gamma^p$$
$$\psi^X = \frac{\sigma_\omega^2}{m^2}\left(\sum_{p=1}^{m} p\gamma^p + \sum_{p=m+1}^{2m-1}(2m-p)\gamma^p\right)$$

where, if $m = 1$, $\psi^B = \psi^A = 0$, and $\psi^X = \sigma_\omega^2\gamma$.

[6] The following differences apply. In Step 2, Burlig et al. (2020) do not include the small-sample correction term $(\bar{Y}_T^B - \bar{Y}_C^B)^2/Z$ in the McKenzie formula. Also, in Step 4, they randomize a treatment vector with PJ ones, whereas I select the first PJ units for treatment. Both changes have negligible impact on the results.

STEP 3. Calculate unit-specific pre-treatment averages $\bar{Y}_i^B = (1/m)\sum_{t\leq 0} Y_{it}$ and drop all m pre-treatment periods from the data.

STEP 4. For integer PJ, assign all units with $i \leq PJ$ to treatment by letting $D_i = 1$; all other units have $D_i = 0$. Then, for all i and t, construct

$$Y_{it}^{McK} = Y_{it} + \tau^{McK} D_i$$
$$Y_{it}^{SCR} = Y_{it} + \tau^{SCR} D_i$$

STEP 5. Estimate the following pair of regressions

$$Y_{it}^{McK} = \alpha + \tilde{\tau} D_i + \theta \bar{Y}_i^B + \epsilon_{it}$$
$$Y_{it}^{SCR} = \alpha + \tilde{\tau} D_i + \theta \bar{Y}_i^B + \epsilon_{it}$$

with robust standard errors clustered at the unit level. In the top half of Figure 1, rejection rates for the former (latter) regression are reported on the left-hand (right-hand) side.

For the bottom half of Figure 1, two alterations are made to the above procedure. First, in Step 1, δ is replaced by δ_t, distributed $\mathcal{N}(\mu_\delta, \sigma_\delta^2)$; values of δ_t are thus drawn $2m$ times. Second, in Step 5, time FE α_t replace the constant α in both regressions.

C.2 Figure 2

For Figure 2, the following simulation procedure is used.

STEP 1. In each simulation run, use DGP

$$Y_{it} = \delta_t + v_i + \omega_{it}$$

to generate a data set with $2m$ time periods and J units. To this effect, values of δ_t are drawn $2m$ times, v_i is drawn J times, while for each i, ω_{it} is generated by an AR(1) process, just as in Section S3.1.

STEP 2. Compute two MDEs by estimating residual-based parameters from the data set generated in Step 1. One, termed τ^{adj}, uses the algorithm described in Appendix S2, using $I = J$ and a single range S equal to all $2m$ periods. The other, denoted τ^{BPW}, follows the original Burlig et al. (2020) procedure for ANCOVA, again with $I = J$ and S given by all periods. As described in Appendix S2, this latter algorithm does not include time FE in the estimating regression, and also uses a slightly different formula for the MDE.

Step 3. Calculate unit-specific pre-treatment averages $\bar{Y}_i^B = (1/m)\sum_{t\leq 0} Y_{it}$ and drop all m pre-treatment periods from the data.

Step 4. For integer PJ, assign all units with $i \leq PJ$ to treatment by letting $D_i = 1$; all other units have $D_i = 0$. Then, for all i and t, construct

$$\begin{aligned} Y_{it}^{BPW} &= Y_{it} + \tau^{BPW} D_i \\ Y_{it}^{adj} &= Y_{it} + \tau^{adj} D_i \end{aligned}$$

Step 5. Estimate the following pair of regressions

$$\begin{aligned} Y_{it}^{BPW} &= \alpha + \tilde{\tau} D_i + \theta \bar{Y}_i^B + \epsilon_{it} \\ Y_{it}^{adj} &= \alpha_t + \tilde{\tau} D_i + \theta \bar{Y}_i^B + \epsilon_{it} \end{aligned}$$

with robust standard errors clustered at the unit level. Again, rejection rates for the former (latter) regression are reported in the left-hand (right-hand) part of the figure.

C.3 Figure 3

Finally, in Figure 3, the following procedure is used, once the Bloom et al. (2015) data set has been loaded. Note that the data include 79 units and 48 time periods.

Step 1. Estimate two sets of residual-based parameters from the pre-existing data. One will be used for computing τ^{BPW} and is estimated using the original Burlig et al. (2020) procedure, while the other is for calculating τ^{adj} and uses the algorithm described in Appendix S2. In both cases, $I = 79$, and ranges S are given by all feasible data subsets of length $m + r$, as described in Section S2.

Step 2. In each simulation run, randomly select a consecutive time period of length $m + r$ from the data, dropping all other periods. This subset of the data will be used as the experimental data set of the current simulation run. Then, randomly assign exactly PJ units to treatment ($D_i = 1$).[7]

Step 3. Calculate τ^{BPW} and τ^{adj}, using values of $\bar{Y}_T^B$, $\bar{Y}_C^B$, and Z corresponding to the current experimental data set.

[7] In practice, only 40/79 units are assigned to treatment; when calculating the MDEs in Step 3, the value of P is adjusted accordingly.

STEP 4. For the current experimental data, calculate unit-specific pre-treatment averages $\bar{Y}_i^B = (1/m) \sum_{t \leq 0} Y_{it}$ and drop all m pre-treatment periods from the data.

STEP 5. For all i and t in the remaining data, construct

$$Y_{it}^{BPW} = Y_{it} + \tau^{BPW} D_i$$
$$Y_{it}^{adj} = Y_{it} + \tau^{adj} D_i$$

STEP 6. Estimate the following pair of regressions

$$Y_{it}^{BPW} = \alpha + \tilde{\tau} D_i + \theta \bar{Y}_i^B + \epsilon_{it}$$
$$Y_{it}^{adj} = \alpha_t + \tilde{\tau} D_i + \theta \bar{Y}_i^B + \epsilon_{it}$$

with robust standard errors clustered at the unit level. In Figure 3, rejection rates for the former (latter) regression are given as dashed (solid) lines.

C.4 Detailed simulation results

The rejection rates summarized in each figure of the main text are reported numerically in Tables 1, 2, and 3 on the following pages.

Table 1: Rejection rates: Figure 1.

	McKenzie (2012) formula					Equation (1.1)				
	AR(1) γ					AR(1) γ				
$m = r$	0	0.3	0.5	0.7	0.9	0	0.3	0.5	0.7	0.9
Panel (i)										
1	0.795	0.909	0.976	0.998	1.000	0.795	0.799	0.801	0.798	0.806
2	0.800	0.783	0.818	0.917	1.000	0.800	0.799	0.794	0.804	0.806
3	0.800	0.702	0.686	0.738	0.966	0.800	0.800	0.803	0.792	0.803
4	0.803	0.656	0.604	0.606	0.858	0.803	0.801	0.804	0.795	0.806
5	0.805	0.628	0.542	0.509	0.716	0.805	0.799	0.799	0.800	0.801
6	0.798	0.622	0.511	0.448	0.585	0.798	0.801	0.805	0.801	0.798
7	0.800	0.604	0.478	0.401	0.494	0.800	0.803	0.796	0.802	0.795
8	0.804	0.599	0.465	0.373	0.433	0.804	0.800	0.801	0.804	0.810
9	0.802	0.591	0.457	0.352	0.373	0.802	0.801	0.802	0.795	0.807
10	0.801	0.586	0.446	0.328	0.330	0.801	0.796	0.795	0.797	0.798
Panel (ii)										
1	0.796	0.915	0.976	0.999	1.000	0.796	0.802	0.802	0.801	0.802
2	0.796	0.789	0.830	0.917	1.000	0.796	0.807	0.805	0.792	0.798
3	0.796	0.705	0.678	0.748	0.967	0.796	0.795	0.795	0.799	0.801
4	0.800	0.663	0.599	0.603	0.856	0.800	0.802	0.799	0.794	0.802
5	0.808	0.635	0.559	0.511	0.717	0.808	0.803	0.807	0.802	0.800
6	0.800	0.607	0.515	0.456	0.590	0.800	0.794	0.801	0.803	0.800
7	0.798	0.604	0.476	0.410	0.497	0.798	0.800	0.801	0.799	0.794
8	0.791	0.603	0.471	0.378	0.428	0.791	0.802	0.802	0.799	0.804
9	0.805	0.588	0.453	0.352	0.377	0.805	0.800	0.799	0.808	0.791
10	0.802	0.578	0.448	0.334	0.329	0.802	0.792	0.802	0.800	0.801

Table 2: Rejection rates: Figure 2.

	τ^{BPW}					τ^{adj}				
	AR(1) γ					AR(1) γ				
$m = r$	0	0.3	0.5	0.7	0.9	0	0.3	0.5	0.7	0.9
1	0.906	0.924	0.936	0.943	0.971	0.798	0.796	0.803	0.794	0.793
2	0.961	0.955	0.963	0.976	0.992	0.802	0.798	0.802	0.804	0.793
3	0.977	0.968	0.964	0.975	0.994	0.800	0.806	0.792	0.801	0.797
4	0.986	0.973	0.967	0.966	0.990	0.798	0.796	0.805	0.803	0.795
5	0.987	0.973	0.962	0.959	0.981	0.797	0.800	0.800	0.801	0.801
6	0.991	0.976	0.961	0.950	0.972	0.804	0.793	0.793	0.806	0.793
7	0.992	0.975	0.958	0.944	0.963	0.802	0.799	0.792	0.805	0.798
8	0.992	0.977	0.956	0.940	0.954	0.806	0.801	0.797	0.805	0.801
9	0.995	0.975	0.957	0.932	0.938	0.801	0.802	0.804	0.796	0.799
10	0.994	0.975	0.956	0.929	0.931	0.798	0.799	0.795	0.800	0.800

Table 3: Rejection rates: Figure 3.

	$m = 1$		$m = 5$		$m = 10$	
r	τ^{BPW}	τ^{adj}	τ^{BPW}	τ^{adj}	τ^{BPW}	τ^{adj}
1	0.850	0.809	0.881	0.805	0.911	0.807
2	0.860	0.807	0.855	0.803	0.865	0.805
3	0.867	0.806	0.845	0.800	0.844	0.801
4	0.880	0.809	0.845	0.803	0.841	0.805
5	0.884	0.805	0.845	0.804	0.835	0.804
6	0.892	0.802	0.842	0.798	0.833	0.803
7	0.894	0.801	0.845	0.798	0.832	0.803
8	0.903	0.800	0.847	0.800	0.829	0.800
9	0.900	0.790	0.850	0.799	0.828	0.798
10	0.907	0.790	0.855	0.802	0.829	0.797

References

Bloom, N., J. Liang, J. Roberts, and ZJ Ying. 2015. Does working from home work? Evidence from a Chinese experiment. *The Quarterly Journal of Economics 130*(1), 165–218.

Burlig, F., L. Preonas, and M. Woerman. 2020. Panel data and experimental design. *Journal of Development Economics 144*, 102458.

McKenzie, D.. 2012. Beyond baseline and follow-up: The case for more T in experiments. *Journal of Development Economics 99*(2), 210–221.