

Working Paper in Economics No. 788

# **ANCOVA power calculation in the presence of serial correlation and time shocks: A comment on Burlig et al. (2020)**

**Claes Ek**

Department of Economics, October 2020

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



**UNIVERSITY OF GOTHENBURG**  
**SCHOOL OF BUSINESS, ECONOMICS AND LAW**

# Analysis-of-covariance power calculation in the presence of serial correlation and time shocks

Claes Ek<sup>a,1</sup>

<sup>a</sup>*Department of Economics, University of Gothenburg, P.O. Box 640, SE-405 30 Gothenburg, Sweden*

---

## Abstract

Research by Burlig et al. (2020) has produced a formula for difference-in-differences power calculation in the presence of serially correlated errors. While the preferred estimator in panel experiments is often analysis-of-covariance (ANCOVA), a similar formula for that estimator was found to yield incorrect power in real data with time shocks. This paper demonstrates that the serial-correlation-robust ANCOVA formula is correct under time shocks as well. Errors arise in Burlig et al. (2020) because time shocks remain unaccounted for in an intermediate procedure estimating residual-based variance parameters from pre-existing data. When that procedure is adjusted accordingly, the serial-correlation-robust ANCOVA formula can be accurately used for power calculation.

*Keywords:* power calculation, randomized experiments, experimental design, panel data, ANCOVA

*JEL classification:* B4, C9, C23

---

## 1 Introduction

Economists are increasingly using randomized experiments to obtain causal estimates of treatment effects (Duflo et al., 2007; Athey and Imbens, 2017). In an influential paper, McKenzie (2012) provides the variance of some common panel estimators for use in *ex-ante* power calculation, confirming that repeatedly measuring an experimental outcome may substantially improve precision (Frison and Pocock, 1992). More recently, Burlig et al. (2020) have extended those panel power formulas to allow for arbitrarily serially correlated errors. Accounting for serial correlation is important, since it is likely to occur in many real-world settings (Bertrand et al., 2004), e.g. whenever outcomes that occur close in time are more

---

<sup>☆</sup>First version: 30 June 2020. This version: 29 September 2020. Declarations of interest: none. I am grateful to Andreas Dzemeski, Peter Jochumzen, Mikael Lindahl, Eskil Rydhe, and Måns Söderbom for useful feedback and suggestions. Any remaining errors are my own.

<sup>1</sup>Corresponding author. *Tel.:* 0046 31 786 52 49. *Email address:* [claes.ek@economics.gu.se](mailto:claes.ek@economics.gu.se)

highly correlated than more distant ones. For the difference-in-differences estimator, Burlig et al. (2020) show that the McKenzie (2012) power formulas typically yield incorrect power when serial correlation is present in the data. By contrast, their novel serial-correlation-robust power formula accurately predicts statistical power in simulated as well as actual data. These results are likely to prove highly useful to any researcher planning experiments with multiple measurements.

Burlig et al. (2020) focus on difference-in-differences rather than the regression analysis-of-covariance (ANCOVA) estimator which replaces unit fixed effects (FE) with a covariate for the pre-treatment outcome average of each unit. Like McKenzie (2012), they do note that ANCOVA is the more efficient estimator, and thus that it is often preferred in randomized settings where unit FE are not needed for identification. However, the authors 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; noting such difficulties, Burlig et al. (2020) instead consider a DGP without time shocks and derive the corresponding small-sample ANCOVA variance formula

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

given as equation A61 in Burlig et al. (2020) and approximated in large samples by equation 10 of the same paper. This formula is shown to be accurate for simulated panel data, again without time shocks; however, when the authors use it to calibrate a minimum detectable effect (MDE) on real-world data, it fails to predict realized power. Burlig et al. (2020) attribute this outcome to the likely presence of time shocks in actual data and caution against using ANCOVA power calculation formulas in practice.

The purpose of this paper is to provide all steps necessary for ANCOVA power calculation in the presence of both time shocks and serially correlated errors. First, in Section 2 below, I show that ANCOVA power formula (1) is in fact correct in the presence of time shocks as well; or equivalently, that such effects do not affect ANCOVA precision. This is intuitive, since ANCOVA is a convex combination of an ex-post means comparison and difference-in-differences, both of which involve comparing means across treatment arms affected identically by the time shocks. Then, in Section 3, I demonstrate that with only a few minor adjustments to the procedures introduced by Burlig et al. (2020), formula (1) can be used to accurately

perform power calculation for ANCOVA in the presence of both serial correlation and time shocks, including for a real data set (Bloom et al., 2015). 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 a first indication that time shocks do not impact ANCOVA precision, consider Figure 1. Panel (i) of the figure replicates Figure 4 in Burlig et al. (2020), where the authors confirm that formula (1) accurately predicts power in simulated data. The DGP underlying both the original figure and panel (i) is

$$Y_{it} = \delta + \tau D_{it} + v_i + \omega_{it} \quad (2)$$

where  $\delta$  is a constant intercept term,  $D_{it}$  is a treatment indicator,  $v_i$  is a unit intercept, and  $\omega_{it}$  is a serially correlated error that is generated according to an AR(1) process, with autoregressive parameter  $\gamma$  varying across simulation sets.

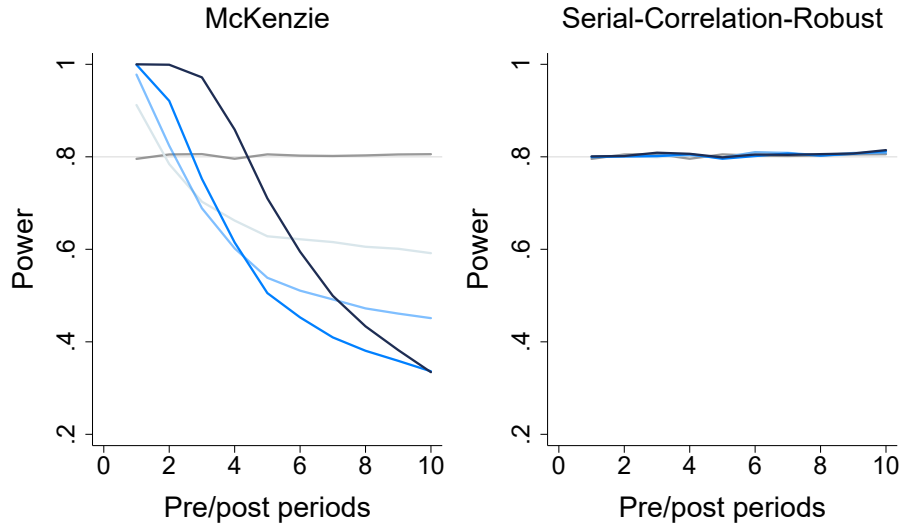
Panel (i) is based on the same set of assumptions, steps, and parameter values as the original figure; these are described in detail in Appendix B.1 of Burlig et al. (2020). To summarize, in each of the 10,000 simulation draws underlying the figure, a data set is constructed from (2), and a constant treatment effect is calibrated to imply nominal 80% power according to either the McKenzie (2012) ANCOVA variance formula (left figure), or the Burlig et al. (2020) formula (right figure), i.e. equation (1). The treatment effect is then added to all units within a randomly drawn treatment group, and an ANCOVA regression is run *ex post*, with 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) 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

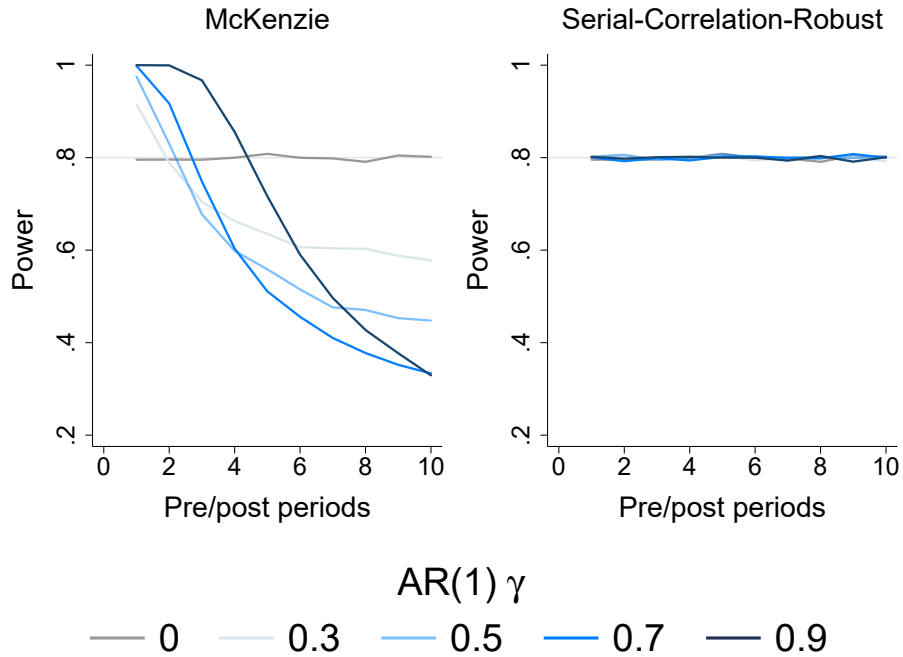
$$Y_{it} = \delta_t + \tau D_{it} + v_i + \omega_{it} \quad (3)$$

which replaces the constant term with a set time shocks  $\delta_t$ , distributed i.i.d.  $\mathcal{N}(\mu_\delta, \sigma_\delta^2)$ , and drawn once per simulated data set. (Equivalently, we may view model (2) as imposing

(i) Time shocks/intercepts not included in DGP or regression



(ii) Time shocks/intercepts included in DGP and regression



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

*Note:* The figure depicts rejection rates for the regression ANCOVA estimator when time shocks are not present in the data (panel i) as well as when they are (panel ii). All figures are based on 10,000 draws from a population where idiosyncratic error term  $\omega_{it}$  follows an AR(1) process with autoregressive parameter  $\gamma$ . All regressions cluster standard errors by unit *ex post*. In the subfigures labeled ‘McKenzie’, the size of the MDEs are calibrated *ex ante* using the McKenzie (2012) power formula. In figures labeled ‘Serial-Correlation-Robust’, the Burlig et al. (2020) serial-correlation-robust ANCOVA power formula (1) is used instead. The DGP and all associated parameter values are as in Figure 4 and Appendix B.1 of Burlig et al. (2020), with the single exception that in panel (ii), normally distributed time shocks with  $\mu_\delta = 20$ ,  $\sigma_\delta^2 = 10$  are included. Despite this, the serial-correlation-robust ANCOVA formula yields appropriate rejection rates.

$\sigma_\delta^2 = 0$ .) Specifically, in panel (ii), the  $\delta_t$  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

$$Y_{it} = \alpha_t + \tilde{\tau}D_i + \theta\bar{Y}_i^B + \epsilon_{it} \quad (4)$$

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  $\epsilon_{it}$  is the regression residual error term. This equation is estimated only on post-treatment observations, allowing the  $t$  subscript of  $D_{it}$  to be dropped. Regression (4) differs from the ANCOVA regressions run in panel (i) only in that time FE  $\alpha_t$  are included in place of a constant term. Again, 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) and thus to the original figure. In particular, rejection rates corresponding to serial-correlation-robust formula (1) remain approximately nominal. Using other values of  $\sigma_\delta^2$  (including very large ones, such as  $\sigma_\delta^2 = 1000$ ) does not alter these results, strongly suggesting that equation (1) can be used with data sets that include time shocks.

Indeed, in Appendix A of this note, I prove that this is the case: for DGP (3), ANCOVA variance is exactly equal to equation (1).<sup>2</sup> In fact, equation (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 technical and relying heavily on matrix partitioning, the proof has an overall structure highly similar to that of Burlig et al. (2020). As in their analysis of the model without time shocks, I find that calculating the variance of the ANCOVA estimator involves evaluating the expression

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

---

<sup>2</sup>Equation (3) admittedly differs slightly from the model stated as Assumption 1 in Burlig et al. (2020), since that DGP includes both time shocks and a constant term. However, the discrepancy is innocuous, since it can be reconciled simply by viewing each time shock in (3) as  $\delta_t = \mu_\delta + \delta'_t$ , with  $\delta'_t$  having mean zero and variance  $\sigma_\delta^2$ .

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  $\epsilon_{it}$  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 fixed effects and idiosyncratic errors that are assumed independent across units. Burlig et al. (2020) show that, 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).

By contrast, when time shocks are included in the DGP, not only must time fixed effects  $\alpha_1, \dots, \alpha_r$  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  $\epsilon_{it}$ . 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), concluding the proof.

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

An obvious question remains: if the ANCOVA variance formula derived by Burlig et al. (2020) is correct after all, 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) can then be accurately expressed in terms of those estimated parameters, rather than the true parameters of the DGP. In a reasonable attempt to remain consistent with their assumed time-shock free model (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 variation in  $\omega_{it}$  which *does* impact power. As a result, the ANCOVA variance calculated from 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. Indeed, Burlig et al. (2020) already do so when considering the difference-in-difference estimator. The modified procedure is described in detail in Appendix B of this paper.

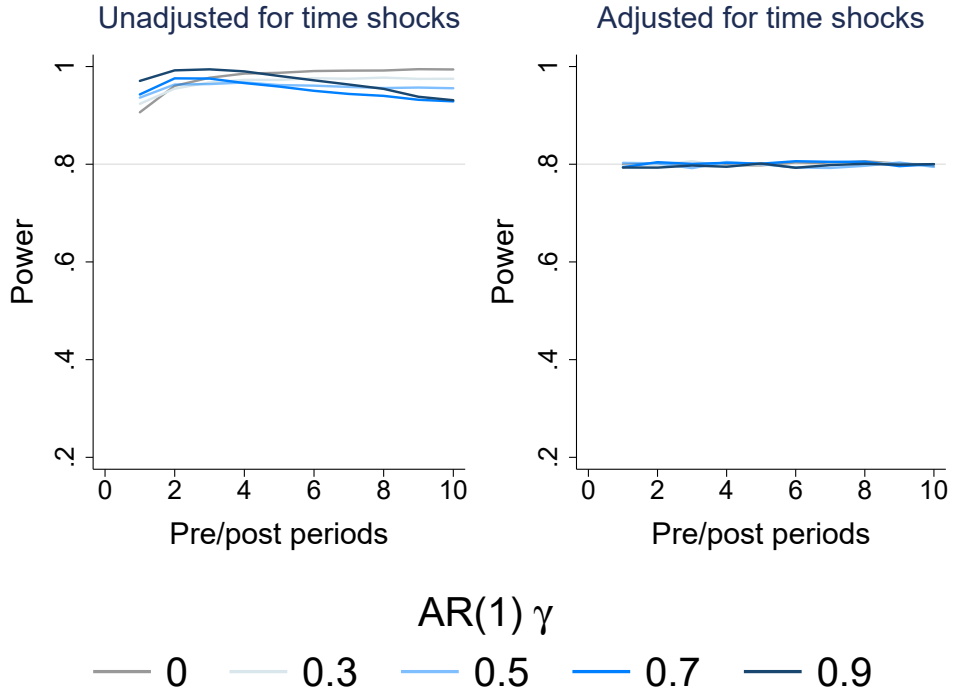
In Figure 2, I compare the two approaches for simulated data. The figure is based on the same model and parameters as panel (ii) of Figure 1; but instead of computing an MDE directly from the parameters of the DGP, I use a set of residual-based parameters estimated from each simulated data set. In the left-hand panel of Figure 2, I follow the estimation procedure described for ANCOVA in Appendix E.3 of Burlig et al. (2020);<sup>3</sup> as expected, this procedure ignores the presence of time shocks and consequently yields excessively high rejection rates. In the right-hand panel, I use the modified approach that correctly accounts for time shocks, and attain nominal power.

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, I retain all procedures and steps used in Figure 2 except those for generating the data: in particular, I again calibrate a constant treatment effect by combining power formula (1) with a set of residual-based parameters estimated from the data, and the effect is then added to (and estimated from) a randomly drawn set of ‘treated’ units. 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.

---

<sup>3</sup>For each simulated data set, I estimate  $\tilde{\sigma}_{\hat{v},S}^2$ ,  $\tilde{\sigma}_{\hat{\omega},S}^2$ ,  $\tilde{\psi}_{\hat{\omega},S}^A$ , and  $\tilde{\psi}_{\hat{\omega},S}^B$  only once, with estimation range  $S$  and sample size  $I$  given by all periods and all units in the data, respectively.  $\tilde{\sigma}_{\hat{v},S}^2$  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  $\tilde{\sigma}_{\hat{\omega},S}^2$  by  $\frac{IT-1}{IT}$  ( $T$  being the panel length) and  $\tilde{\sigma}_{\hat{v},S}^2$  by  $\frac{I-1}{I}$  but leave the  $\tilde{\psi}$  estimates unadjusted, in accordance with the discussion of e.g.  $E[\tilde{\psi}_{\hat{\omega}}^B]$  in Appendix E.3 of Burlig et al. (2020).



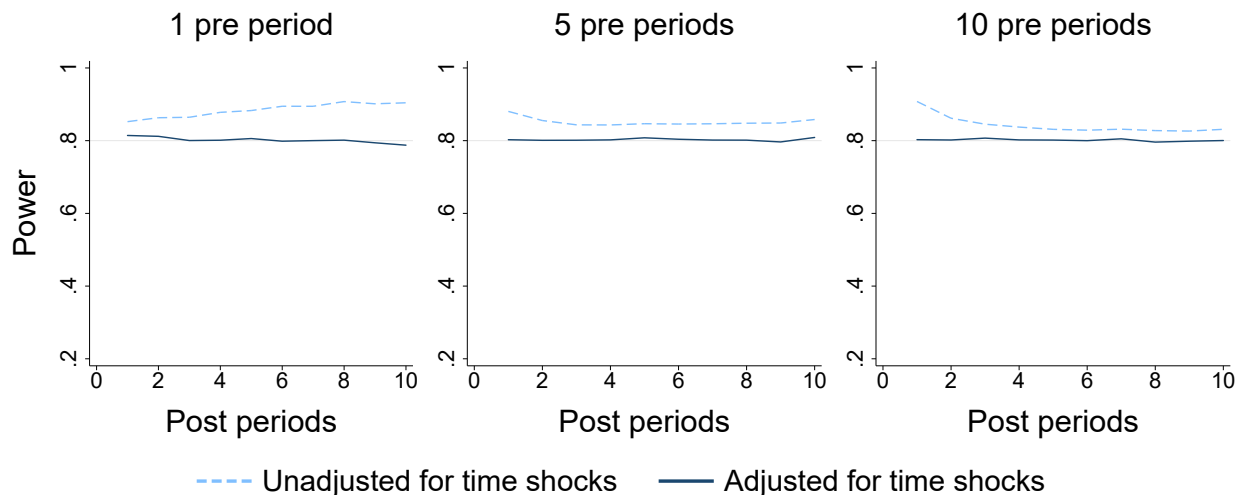


**Figure 2:** Accounting for time shocks when estimating residual-based parameters: simulated data

*Note:* The figure depicts rejection rates for the regression ANCOVA estimator when time shocks are present in the data. Both panels are based on 10,000 draws from a population where idiosyncratic error term  $\omega_{it}$  follows an AR(1) process with autoregressive parameter  $\gamma$ . The DGP and all associated parameter values are as in panel (ii) of Figure 1. Both panels above calibrate an MDE appropriate for serial-correlation-robust power calculation using estimates of residual-based parameters. In the left-hand panel, this procedure is based on a regression of  $Y_{it}$  on unit fixed effects only, in accordance with the approach described in Appendix E.3 of Burlig et al. (2020). In the right-hand panel, the regression is on both unit and time FE; minor adjustments are also made to the MDE calculation, as described in Appendix B of this paper. These adjustments result in appropriate rejection rates. Both panels estimate ANCOVA *ex post*, clustering standard errors by unit; however, the regressions include time FE only in the right-hand panel. Only the adjusted procedure attains nominal power.

## 4 Conclusion

This short paper has demonstrated the feasibility, adding only minor modifications, of using the Burlig et al. (2020) approach to perform an accurate ANCOVA power calculation that is robust to time shocks as well as serial correlation. It seems likely that the Stata packages introduced by the authors could be similarly modified, usefully expanding the power-calculation toolkit available to experimenters even further.



**Figure 3:** Accounting for time shocks when estimating residual-based parameters: real data

*Note:* Each panel simulates experiments with a certain number of pre-treatment periods  $m \in \{1, 5, 10\}$ . Horizontal axes vary the number of post-treatment periods ( $1 \leq r \leq 10$ ). In each panel, both lines calibrate an MDE using the serial-correlation-robust ANCOVA formula in combination with estimates of residual-based parameters from the Bloom et al. (2015) data set. Lines labeled ‘Unadjusted for time shocks’ replicate the original Burlig et al. (2020) approach where time shocks are ignored in the parameter-estimation step. Lines labeled ‘Adjusted for time shocks’ follow the procedure outlined in Appendix B of this paper. Both cases estimate ANCOVA *ex post*, clustering standard errors by unit; however, only the ‘Adjusted for time shocks’ lines include time FE in the ANCOVA regression. Only the adjusted procedure attains nominal power.

## References

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

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 “ANCOVA power calculation in the presence of serial correlation and time shocks”

## 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. All model assumptions in Burlig et al. (2020) are retained as well as repeated 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** (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  $t$ ;  $\tau$  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.*  $\mathcal{N}(0, \sigma_v^2)$ ;  $\omega_{it}$  is an idiosyncratic error term distributed (not necessarily *i.i.d.*)  $\mathcal{N}(0, \sigma_\omega^2)$ . Finally, in a departure from the Burlig et al. (2020) model,  $\delta_t$  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 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 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 4** (Independence across units).  $E[\omega_{it}\omega_{js}|\mathbf{X}_r] = 0, \forall i \neq j, \forall t, s$ .

**Assumption 5** (Uniform covariance structures). *Define:*

$$\begin{aligned}\psi_i^B &\equiv \frac{2}{m(m-1)} \sum_{t=-m+1}^{-1} \sum_{s=t+1}^0 \text{Cov}(\omega_{it}, \omega_{is} | \mathbf{X}_r) \\ \psi_i^A &\equiv \frac{2}{r(r-1)} \sum_{t=1}^{r-1} \sum_{s=t+1}^r \text{Cov}(\omega_{it}, \omega_{is} | \mathbf{X}_r) \\ \psi_i^X &\equiv \frac{1}{mr} \sum_{t=-m+1}^0 \sum_{s=1}^r \text{Cov}(\omega_{it}, \omega_{is} | \mathbf{X}_r)\end{aligned}$$

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 A.1, we consider the case when time shocks are included in the regression equation; then, in Appendix A.2, we consider the case when they are not. In both cases, the result will be equal to variance equation (1) in the main text, calculated as equation A61 in Burlig et al. (2020).

#### 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  $\epsilon_{it}$  is the regression residual error term. Finally,  $\alpha_t$  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  $E[\mathbf{X}'\epsilon] = 0$ , where  $\mathbf{X}$  is the regressor matrix in (A.2), and  $\epsilon$  is the full vector of

residuals. This equation system<sup>4</sup> includes the conditions  $\epsilon_{it} = 0$  for all  $i$  and  $t$ . Moreover, its solution yields the set of projection parameters to which the ANCOVA (i.e. OLS) regression estimator will converge.<sup>5</sup> These coefficients are  $\alpha_t = \delta_t - \theta\bar{\delta}^B$ , where  $\bar{\delta}^B = \frac{1}{m} \sum_{p=-m+1}^0 \delta_p$ ;  $\tilde{\tau} = \tau$ ; and

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

As a result, we also have

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

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  $\hat{\beta}$  the set of regression coefficients from OLS estimation of (A.2) given (A.1), the coefficient covariance matrix is given by the sandwich formula

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

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

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

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

<sup>4</sup>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  $i$ :  $\text{E}[\mathbf{X}'_i\boldsymbol{\epsilon}_i] = 0$ .

<sup>5</sup>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  $\delta_t$  as nonstochastic; otherwise, for instance, every  $\alpha_t$  will equal the same value. All other computations throughout this proof do treat the time shocks as random.

$$\left( \begin{array}{cc}
\sum_{i=1}^{PJ} \sum_{j=1}^J \sum_{t=1}^r \mathbf{E}[\epsilon_{it}\epsilon_{j1}|\mathbf{X}] & \sum_{i=1}^J \sum_{j=1}^J \sum_{t=1}^r \bar{Y}_i^B \mathbf{E}[\epsilon_{it}\epsilon_{j1}|\mathbf{X}] \\
\vdots & \vdots \\
\sum_{i=1}^{PJ} \sum_{j=1}^J \sum_{t=1}^r \mathbf{E}[\epsilon_{it}\epsilon_{jr}|\mathbf{X}] & \sum_{i=1}^J \sum_{j=1}^J \sum_{t=1}^r \bar{Y}_i^B \mathbf{E}[\epsilon_{it}\epsilon_{jr}|\mathbf{X}] \\
\sum_{i=1}^{PJ} \sum_{j=1}^{PJ} \sum_{t=1}^r \sum_{s=1}^r \mathbf{E}[\epsilon_{it}\epsilon_{js}|\mathbf{X}] & \sum_{i=1}^{PJ} \sum_{j=1}^J \sum_{t=1}^r \sum_{s=1}^r \bar{Y}_j^B \mathbf{E}[\epsilon_{it}\epsilon_{js}|\mathbf{X}] \\
\sum_{i=1}^{PJ} \sum_{j=1}^J \sum_{t=1}^r \sum_{s=1}^r \bar{Y}_j^B \mathbf{E}[\epsilon_{it}\epsilon_{js}|\mathbf{X}] & \sum_{i=1}^J \sum_{j=1}^J \sum_{t=1}^r \sum_{s=1}^r \bar{Y}_i^B \bar{Y}_j^B \mathbf{E}[\epsilon_{it}\epsilon_{js}|\mathbf{X}]
\end{array} \right) \quad (\text{A.5})$$

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

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

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  $\mathbf{G}$ ,

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

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

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<sup>6</sup> is the inverse of

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

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

**Lemma 1.** *Any  $n$ -dimensional square matrix of the form*

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

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  $\mathbf{Y}_2$ ):

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

---

<sup>6</sup>Because we are interested in element  $(r+1, r+1)$  of  $\text{Var}(\hat{\tau}|\mathbf{X})$ , calculating this final submatrix is, strictly speaking, unnecessary and is done for completeness only.



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  $(\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{J}(\frac{1}{J}\mathbf{X}'\mathbf{X})^{-1}$ , we have

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

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

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

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

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  $i \neq j$ ,

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

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)  $\epsilon_{it}$  is unconditionally uncorrelated with baseline averages other than  $\bar{Y}_i^B$ , evaluating the mean of  $\epsilon_{it}$  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  $\alpha_t$ ) 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 other components of  $\epsilon_{it}$ .

When instead  $i = j$ , we have

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

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:<sup>7</sup>

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

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

$$\boldsymbol{\Sigma}^{\mathbf{y}\mathbf{y}} = \begin{pmatrix} a_s & b_s & 0 & \cdots & 0 & | & 0 & 0 & \cdots & 0 \\ b_s & c & d & \cdots & d & | & -\theta d & -\theta d & \cdots & -\theta d \\ 0 & d & c & \cdots & d & | & -\theta d & -\theta d & \cdots & -\theta d \\ \vdots & \vdots & \vdots & \ddots & \vdots & | & \vdots & \vdots & & \vdots \\ 0 & d & d & \cdots & c & | & -\theta d & -\theta d & \cdots & -\theta d \\ \hline 0 & -\theta d & -\theta d & \cdots & -\theta d & | & (m + \theta^2)d & \theta^2 d & \cdots & \theta^2 d \\ 0 & -\theta d & -\theta d & \cdots & -\theta d & | & \theta^2 d & (m + \theta^2)d & \cdots & \theta^2 d \\ \vdots & \vdots & \vdots & & \vdots & | & \vdots & \vdots & \ddots & \vdots \\ 0 & -\theta d & -\theta d & \cdots & -\theta d & | & \theta^2 d & \theta^2 d & \cdots & (m + \theta^2)d \end{pmatrix} \quad (\text{A.11})$$

$$\equiv \begin{pmatrix} \boldsymbol{\Sigma}_{11}^{\mathbf{y}\mathbf{y}} & \boldsymbol{\Sigma}_{12}^{\mathbf{y}\mathbf{y}} \\ \boldsymbol{\Sigma}_{21}^{\mathbf{y}\mathbf{y}} & \boldsymbol{\Sigma}_{22}^{\mathbf{y}\mathbf{y}} \end{pmatrix}$$

For convenience, the matrix uses the following definitions.

$$a_s = \text{Var}(\epsilon_{js}) = (1 - \theta)^2 \sigma_v^2 + \left(1 + \frac{\theta^2}{m}\right) \sigma_\omega^2 - \theta \text{Cov}(\omega_{js}, \bar{\omega}_j^B) + \frac{\theta^2 (m - 1)}{m} \psi^B$$

$$b_s = \text{Cov}(\epsilon_{js}, \bar{Y}_j^B) = \text{Cov}(\omega_{js}, \bar{\omega}_j^B) - \psi^X$$

---

<sup>7</sup>If one assumes that all  $\delta_t$  are fixed rather than random, then pre-treatment time shocks cease to be informative regarding the residuals. Thus,  $\mathbb{E}[\epsilon_{it}|\epsilon_{js}, \mathbf{X}] = \mathbb{E}[\epsilon_{it}|\bar{Y}_i^B]$  for  $i \neq j$ ,  $\mathbb{E}[\epsilon_{it}|\epsilon_{is}, \mathbf{X}] = \mathbb{E}[\epsilon_{it}|\epsilon_{is}, \bar{Y}_i^B]$  for  $i = j$ , and moreover  $\mathbb{E}[\epsilon_{is}|\mathbf{X}] = \mathbb{E}[\epsilon_{is}|\bar{Y}_i^B]$  while also  $\mathbb{E}[\epsilon_{is}^2|\mathbf{X}] = \mathbb{E}[\epsilon_{is}^2|\bar{Y}_i^B]$ . It is reasonably simple to confirm these statements using (A.10). As a result, 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 A.2.

$$c = \text{Var}(\bar{Y}_i^B) = \frac{1}{m} (\sigma_\delta^2 + \sigma_\omega^2 + m\sigma_v^2 + (m-1)\psi^B)$$

$$d = \text{Cov}(\bar{Y}_i^B, \bar{Y}_j^B) = \frac{\sigma_\delta^2}{m}, \text{ for } i \neq j$$

with  $\bar{\omega}_j^B = (1/m) \sum_{p=-m+1}^0 \omega_{jp}$ . Notice that  $\sum_{s=1}^r b_s = 0$  for any  $s$ . Furthermore, when  $i \neq j$ ,

$$\Sigma^{\mathbf{xy}} = \begin{pmatrix} 0 & 0 & b_t & 0 & \cdots & 0 \end{pmatrix} \quad (\text{A.12})$$

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

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

where

$$e_{ts} = \text{Cov}(\epsilon_{it}, \epsilon_{is})$$

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

which we may also note implies

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

$$\equiv r \bar{e}_s$$

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  $(\Sigma^{\mathbf{yy}})^{-1}$ .

First, we use Lemma 1 to calculate

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

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

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

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

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

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

$$\left( \Sigma_{11}^{yy} - \Sigma_{12}^{yy} (\Sigma_{22}^{yy})^{-1} \Sigma_{21}^{yy} \right)^{-1} = \frac{(c-d)^{J-2}}{|\Sigma|} \begin{pmatrix} (c-d) \left( c-d + \frac{Jmd}{m+r\theta^2} \right) & -b_s \left( c-d + \frac{(J-1)md}{m+r\theta^2} \right) \\ -b_s \left( c-d + \frac{(J-1)md}{m+r\theta^2} \right) & a_s \left( c-d + \frac{(J-1)md}{m+r\theta^2} \right) \\ \frac{md}{m+r\theta^2} b_s & -\frac{md}{m+r\theta^2} a_s \\ \vdots & \vdots \\ \frac{md}{m+r\theta^2} b_s & -\frac{md}{m+r\theta^2} a_s \end{pmatrix} \begin{pmatrix} \frac{md}{m+r\theta^2} b_s & \cdots & \frac{md}{m+r\theta^2} b_s \\ -\frac{md}{m+r\theta^2} a_s & \cdots & -\frac{md}{m+r\theta^2} a_s \\ a_s \left( c-d + \frac{(J-1)md}{m+r\theta^2} \right) - b_s^2 \left( 1 + \frac{(J-2)md}{(c-d)(m+r\theta^2)} \right) & \cdots & \frac{md}{m+r\theta^2} \left( -a_s + \frac{b_s^2}{c-d} \right) \\ \vdots & \ddots & \vdots \\ \frac{md}{m+r\theta^2} \left( -a_s + \frac{b_s^2}{c-d} \right) & \cdots & a_s \left( c-d + \frac{(J-1)md}{m+r\theta^2} \right) - b_s^2 \left( 1 + \frac{(J-2)md}{(c-d)(m+r\theta^2)} \right) \end{pmatrix} \quad (\text{A.16})$$

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

$$- \left( \Sigma_{11}^{yy} - \Sigma_{12}^{yy} (\Sigma_{22}^{yy})^{-1} \Sigma_{21}^{yy} \right)^{-1} \Sigma_{12}^{yy} (\Sigma_{22}^{yy})^{-1}$$

$$= \frac{\theta(c-d)^{J-2}}{(m+r\theta^2)|\Sigma|} \times \begin{pmatrix} -b_s(c-d) & -b_s(c-d) & \cdots & -b_s(c-d) \\ a_s(c-d) & a_s(c-d) & \cdots & a_s(c-d) \\ a_s(c-d) - b_s^2 & a_s(c-d) - b_s^2 & \cdots & a_s(c-d) - b_s^2 \\ \vdots & \vdots & & \vdots \\ a_s(c-d) - b_s^2 & a_s(c-d) - b_s^2 & \cdots & a_s(c-d) - b_s^2 \end{pmatrix} \quad (\text{A.17})$$

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

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

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

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

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

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

and the case of  $i \neq j$  will not contribute to  $\text{Var}(\hat{\tau}|\mathbf{X})$  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

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

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

$$\begin{aligned}
\sum_{t=1}^r E[\epsilon_{it}\epsilon_{is}|\mathbf{X}] &= \frac{r(c-d)^{J-2}}{|\boldsymbol{\Sigma}|} \left[ \bar{e}_s(c-d) \left( c-d + \frac{Jmd}{m+r\theta^2} \right) E[\epsilon_{is}^2|\mathbf{X}] \right. \\
&+ \left[ -\bar{e}_s b_s \left( c-d + \frac{(J-1)md}{m+r\theta^2} \right) (\bar{Y}_i^B - \mu_\delta) + \frac{md}{m+r\theta^2} \bar{e}_s b_s \sum_{k \neq i} (\bar{Y}_k^B - \mu_\delta) \right. \\
&\left. \left. - \frac{\theta(c-d)}{m+r\theta^2} \bar{e}_s b_s \sum_{p=1}^r (\alpha_p - (1-\theta)\mu_\delta) \right] E[\epsilon_{is}|\mathbf{X}] \right] \\
&= r \left[ A_1 E[\epsilon_{is}^2|\mathbf{X}] + \left[ A_2 (\bar{Y}_i^B - \mu_\delta) + A_3 \sum_{k \neq i} (\bar{Y}_k^B - \mu_\delta) \right. \right. \\
&\left. \left. + A_4 \sum_{p=1}^r (\alpha_p - (1-\theta)\mu_\delta) \right] E[\epsilon_{is}|\mathbf{X}] \right] \tag{A.19}
\end{aligned}$$

with  $A_1, \dots, A_4$  defined accordingly. The next step is to calculate the ‘outer’ expectation  $E[\epsilon_{is}|\mathbf{X}] = E[\epsilon_{is}|\bar{Y}_i^B, \bar{Y}_{-i}^B, \alpha_1, \dots, \alpha_r]$ , 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

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

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

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

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

$$(c - d)^{J-1} \left( c - d + \frac{Jmd}{m + r\theta^2} \right) \equiv |\hat{\Sigma}| \quad (\text{A.21})$$

and application of Lemma 1 shows that its inverse, forming the top-left partition of  $(\hat{\Sigma}^{\mathbf{y}\mathbf{y}})^{-1}$ , is

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

where, as usual, all (off)diagonal elements equal the same value. Finally, the top-right



partition of  $(\hat{\Sigma}_{\mathbf{y}\mathbf{y}})^{-1}$  is

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

Application of formula (A.10) now yields

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

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

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

$$\text{Var}(x|\mathbf{y}) = \sigma_x^2 - \Sigma_{\mathbf{x}\mathbf{y}} \Sigma_{\mathbf{y}\mathbf{y}}^{-1} (\Sigma_{\mathbf{x}\mathbf{y}})' \quad (\text{A.25})$$

where  $\sigma_x^2$  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  $\hat{\Sigma}^{\mathbf{x}\mathbf{y}}$ , (A.22) and (A.23) in accordance with the above formula yields

$$\text{Var}(\epsilon_{is}|\mathbf{X}) = \frac{|\Sigma|}{|\hat{\Sigma}|} \quad (\text{A.26})$$

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

$$\sum_{t=1}^r \mathbb{E}[\epsilon_{it}\epsilon_{is}|\mathbf{X}] = r \left[ A_1 \frac{|\Sigma|}{|\hat{\Sigma}|} + B_1 (A_1 B_1 + A_2) (\bar{Y}_i^B - \mu_\delta)^2 + B_2 (A_1 B_2 + A_3) \left( \sum_{k \neq i} (\bar{Y}_k^B - \mu_\delta) \right)^2 \right]$$

$$\begin{aligned}
& + B_3 (A_1 B_3 + A_4) \left( \sum_{p=1}^r (\alpha_p - (1 - \theta) \mu_\delta) \right)^2 \\
& + \left( B_1 (A_1 B_2 + A_3) + B_2 (A_1 B_1 + A_2) \right) (\bar{Y}_i^B - \mu_\delta) \sum_{k \neq i} (\bar{Y}_k^B - \mu_\delta) \\
& + \left( B_1 (A_1 B_3 + A_4) + B_3 (A_1 B_1 + A_2) \right) (\bar{Y}_i^B - \mu_\delta) \sum_{p=1}^r (\alpha_p - (1 - \theta) \mu_\delta) \\
& + \left( B_2 (A_1 B_3 + A_4) + B_3 (A_1 B_2 + A_3) \right) \sum_{k \neq i} (\bar{Y}_k^B - \mu_\delta) \sum_{p=1}^r (\alpha_p - (1 - \theta) \mu_\delta) \Big]
\end{aligned}$$

The first term within the bracket is

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

and moreover  $A_1 B_1 + A_2 = A_1 B_1 + A_3 = A_1 B_3 + A_4 = 0$ , so the conditional expectation, summed over all  $t$  and  $s$ , reduces to

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

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

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

which may be further simplified to yield

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

This is the same small-sample ANCOVA variance formula as derived by Burlig et al. (2020).

## A.2 Time shocks not included in ANCOVA regression

Now, consider instead the ANCOVA regression model

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

where  $\alpha$  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 analyzed 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 A.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  $\theta$  again equal to (A.3). As a result, residuals now 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  $\epsilon_{it}$  is uncorrelated with  $\alpha$ . Also notice that  $E[\epsilon_{it}] = 0$ .

Next, we will again calculate the ANCOVA variance by sandwich formula (A.4). Since  $\mathbf{X}'\mathbf{X}$  is now the 3-by-3 matrix 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 A.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  $(\epsilon_{it})$  expectation is

$$\Sigma^{\mathbf{y}\mathbf{y}} = \left( \begin{array}{ccccc|c} a_s + (r+1)d & b_s & 0 & \cdots & 0 & 0 \\ b_s & c & d & \cdots & d & -\theta d \\ 0 & d & c & \cdots & d & -\theta d \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & d & d & \cdots & c & -\theta d \\ \hline 0 & -\theta d & -\theta d & \cdots & -\theta d & (\frac{m}{r} + \theta^2) d \end{array} \right) \quad (\text{A.28})$$

where all parameters take the same values as in Appendix A.1 above, again implying  $\sum_{s=1}^r b_s = 0$  for any  $s$ . Note that under the partitioning used,  $\Sigma_{22}^{\mathbf{y}\mathbf{y}}$  is just the scalar  $(\frac{m}{r} + \theta^2) d$ , making that partition straightforward to invert. In any case, when  $i \neq j$ , the relevant covariance vector is

$$\Sigma^{\mathbf{x}\mathbf{y}} = \left( e_{ts}^{i \neq j} \quad 0 \quad b_t \quad 0 \quad \cdots \quad 0 \right) \quad (\text{A.29})$$

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

$$\Sigma^{\mathbf{x}\mathbf{y}} = \left( e_{ts}^{i=j} + e_{ts} \quad b_t \quad 0 \quad \cdots \quad 0 \right) \quad (\text{A.30})$$

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

Now,

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

with determinant

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

and symmetric inverse

$$\begin{aligned}
\left(\Sigma_{11}^{yy} - \Sigma_{12}^{yy} (\Sigma_{22}^{yy})^{-1} \Sigma_{21}^{yy}\right)^{-1} &= \frac{(c-d)^{J-2}}{|\Sigma|} \begin{pmatrix} (c-d) \left(c-d + \frac{Jmd}{m+r\theta^2}\right) & -b_s \left(c-d + \frac{(J-1)md}{m+r\theta^2}\right) \\ -b_s \left(c-d + \frac{(J-1)md}{m+r\theta^2}\right) & (a_s + (r+1)d) \left(c-d + \frac{(J-1)md}{m+r\theta^2}\right) \\ \frac{md}{m+r\theta^2} b_s & -\frac{md}{m+r\theta^2} (a_s + (r+1)d) \\ \vdots & \vdots \\ \frac{md}{m+r\theta^2} b_s & -\frac{md}{m+r\theta^2} (a_s + (r+1)d) \\ & \dots \\ & -\frac{md}{m+r\theta^2} (a_s + (r+1)d) \\ (a_s + (r+1)d) \left(c-d + \frac{(J-1)md}{m+r\theta^2}\right) - b_s^2 \left(1 + \frac{(J-2)md}{(c-d)(m+r\theta^2)}\right) & \dots \\ \vdots & \ddots \\ \frac{md}{m+r\theta^2} \left(-a_s - (r+1)d + \frac{b_s^2}{c-d}\right) & \dots \\ & \dots \\ & \frac{md}{m+r\theta^2} b_s \\ & -\frac{md}{m+r\theta^2} (a_s + (r+1)d) \\ \frac{md}{m+r\theta^2} \left(-a_s - (r+1)d + \frac{b_s^2}{c-d}\right) & \\ \vdots & \\ (a_s + (r+1)d) \left(c-d + \frac{(J-1)md}{m+r\theta^2}\right) - b_s^2 \left(1 + \frac{(J-2)md}{(c-d)(m+r\theta^2)}\right) & \end{pmatrix} \quad (\text{A.31})
\end{aligned}$$

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

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

Combining these expressions with (A.29) produces

$$\begin{aligned}
E[\epsilon_{it} | \epsilon_{js}, \mathbf{X}] &= \frac{(c-d)^{J-2}}{|\Sigma|} \left[ \left[ (c-d) \left(c-d + \frac{Jmd}{m+r\theta^2}\right) e_{ts}^{i \neq j} + \frac{mb_s d}{m+r\theta^2} b_t \right] \epsilon_{js} \right. \\
&+ \left. \left[ -b_s \left(c-d + \frac{(J-1)md}{m+r\theta^2}\right) e_{ts}^{i \neq j} - (a_s + (r+1)d) \frac{md}{m+r\theta^2} b_t \right] (\bar{Y}_j^B - \mu_\delta) \right]
\end{aligned}$$

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

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

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

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

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

Thus, as in Appendix A.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

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

$$\begin{aligned}
& + \left[ \frac{md}{m+r\theta^2} \left( b_s \left( e_{ts} + e_{ts}^{i \neq j} \right) - (a_s + (r+1)d)b_t \right) \right] \sum_{k \neq i} (\bar{Y}_k^B - \mu_\delta) \\
& + \left[ \frac{r\theta(c-d)}{m+r\theta^2} \left( -b_s \left( e_{ts} + e_{ts}^{i \neq j} \right) + (a_s + (r+1)d)b_t \right) \right] (\alpha - (1-\theta)\mu_\delta) \Big]
\end{aligned}$$

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

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

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

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

$$\hat{\Sigma}^{\mathbf{y}\mathbf{y}} = \begin{pmatrix} c & d & \cdots & d & | & -\theta d \\ d & c & \cdots & d & | & -\theta d \\ \vdots & \vdots & \ddots & \vdots & | & \vdots \\ d & d & \cdots & c & | & -\theta d \\ \hline -\theta d & -\theta d & \cdots & -\theta d & | & (m+\theta^2)d \end{pmatrix}$$

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

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

forming the top-left portion of  $(\hat{\Sigma}_{\mathbf{y}\mathbf{y}})^{-1}$ , is (A.22). The top-right partition is

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

Applying formula (A.10) then yields

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

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

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

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

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



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

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

and moreover we again find that  $A_1 B_1 + A_2 = A_1 B_1 + A_3 = A_1 B_3 + A_4 = 0$ . Thus,

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

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

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

Throughout this section, we retain model assumptions 1-5 from Appendix A of this note; 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 minimum detectable effects (MDE) from a pre-existing data set, consider the following. (Notice that steps 1 and 3 remain as originally proposed by the authors.)

1. Determine all feasible ranges of experiments with  $(m+r)$  periods, given the number of time periods in the pre-existing data set.
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  $\tilde{\sigma}_{\hat{v},S}^2$ .
  - (c) Calculate the variance of the stored residuals, and save as  $\tilde{\sigma}_{\omega,S}^2$ .

- (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  $\tilde{\psi}_{\hat{\omega},S}^B$ .
- (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  $\tilde{\psi}_{\hat{\omega},S}^A$ .<sup>8</sup>
3. Calculate the average of  $\tilde{\sigma}_{\hat{v},S}^2$ ,  $\tilde{\sigma}_{\hat{\omega},S}^2$ ,  $\tilde{\psi}_{\hat{\omega},S}^B$ , and  $\tilde{\psi}_{\hat{\omega},S}^A$  across all ranges  $S$ , deflating  $\tilde{\sigma}_{\hat{\omega},S}^2$  by  $\frac{I(m+r)-1}{I(m+r)}$  and  $\tilde{\sigma}_{\hat{v},S}^2$ ,  $\tilde{\psi}_{\hat{\omega},S}^B$ , and  $\tilde{\psi}_{\hat{\omega},S}^A$  by  $\frac{I-1}{I}$ . These averages are equal in expectation to  $\sigma_{\hat{v}}^2, \sigma_{\hat{\omega}}^2$ ,  $\psi_{\hat{\omega}}^B$ , and  $\psi_{\hat{\omega}}^A$ .
4. To produce the estimated MDE, plug these values into

$$\begin{aligned}
MDE^{est} &= (t_{1-\kappa}^J - t_{\alpha/2}^J) \times \left\{ \left( \frac{1}{P(1-P)J} + \frac{(\bar{Y}_T^B - \bar{Y}_C^B)^2}{Z} \right) \times \left( \frac{I}{I-1} \right) \right. \\
&\quad \times \left[ (1 - \theta^2) \sigma_{\hat{v}}^2 + \left( \frac{m + \theta r}{2m^2 r^2} \right) ((m + r)(m + \theta r) + (1 - \theta)(mr^2 - m^2 r)) \sigma_{\hat{\omega}}^2 \right. \\
&\quad + \left( \frac{m + \theta r}{2mr^2} \right) (m - 1)(m + \theta r - (1 - \theta)mr) \psi_{\hat{\omega}}^B \\
&\quad \left. \left. + \left( \frac{m + \theta r}{2m^2 r} \right) (r - 1)(m + \theta r + (1 - \theta)mr) \psi_{\hat{\omega}}^A \right] \right\}^{1/2} \tag{A.35}
\end{aligned}$$

where  $t_{1-\kappa}^J$  and  $t_{\alpha/2}^J$  are suitable critical values of the  $t$  distribution, and  $\theta$  is expressed in terms of the residual-based parameters as

$$\begin{aligned}
\theta &= \frac{m [4mr\sigma_{\hat{v}}^2 - (m(m - r + 2) + r(r - m + 2)) \sigma_{\hat{\omega}}^2]}{2r [2m^2\sigma_{\hat{v}}^2 + (m(m + 1) - r(m - 1))\sigma_{\hat{\omega}}^2 + (m(m - 1)(m + 1)\psi_{\hat{\omega}}^B - r(m - 1)(r - 1)\psi_{\hat{\omega}}^A)]} \\
&\quad + \frac{m [-m(m - 1)(m - r + 2)\psi_{\hat{\omega}}^B - r(r - 1)(r - m + 2)\psi_{\hat{\omega}}^A]}{2r [2m^2\sigma_{\hat{v}}^2 + (m(m + 1) - r(m - 1))\sigma_{\hat{\omega}}^2 + (m(m - 1)(m + 1)\psi_{\hat{\omega}}^B - r(m - 1)(r - 1)\psi_{\hat{\omega}}^A)]} \tag{A.36}
\end{aligned}$$

---

<sup>8</sup>Burlig et al. (2020) add an additional step estimating the residual-based across-period covariance,  $\tilde{\psi}_{\hat{\omega},S}^X$ . However, that step turns out to be redundant, both here and in the original procedure, since  $\tilde{\psi}_{\hat{\omega},S}^X$  is not used when calculating the MDE.

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  $\tilde{\sigma}_v^2$ , the proof is identical to that provided in Appendix E.2 of Burlig et al. (2020). Furthermore, in the estimating regression,

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

which is identical to the  $\sigma_v^2$  estimate obtained when time FE are not included in the estimating regression of step 2a above. The proof that  $E[\tilde{\sigma}_v^2] = \sigma_v^2$  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  $\sigma_v^2$ , 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

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

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

$$\begin{aligned} \sigma_\omega^2 &= \left( \frac{I-1}{I(m+r)^2} \right) \left( (m+r)(m+r-1) \sigma_\omega^2 - m(m-1) \psi^B - r(r-1) \psi^A - 2mr \psi^X \right) \\ \psi_\omega^B &= \left( \frac{I-1}{I(m+r)^2} \right) \left( -(m+r) \sigma_\omega^2 + (r^2 + 2r + m) \psi^B + r(r-1) \psi^A - 2r^2 \psi^X \right) \\ \psi_\omega^A &= \left( \frac{I-1}{I(m+r)^2} \right) \left( -(m+r) \sigma_\omega^2 + m(m-1) \psi^B + (m^2 + 2m + r) \psi^A - 2m^2 \psi^X \right) \end{aligned}$$

$$\psi_{\omega}^X = \left( \frac{I-1}{I(m+r)^2} \right) (-(m+r)\sigma_{\omega}^2 - r(m-1)\psi^B - m(r-1)\psi^A + 2mr\psi^X)$$

Comparing with the corresponding expressions in Appendix E.3 of Burlig et al. (2020), we note the single difference that all residual-based parameters  $\sigma_v^2, \sigma_{\omega}^2, \psi_{\omega}^B, \psi_{\omega}^A$ , and  $\psi_{\omega}^X$  are now multiplied by  $\frac{I-1}{I}$ , while this was true only for  $\sigma_v^2$  in the original procedure. In any case, we now seek coefficients  $k_v, k_{\omega}, k_B, k_A$ , 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:

$$\begin{aligned} & k_v\sigma_v^2 + k_{\omega}\sigma_{\omega}^2 + k_B\psi_{\omega}^B + k_A\psi_{\omega}^A + k_X\psi_{\omega}^X \\ &= (1-\theta)^2\sigma_v^2 + \left( \frac{\theta^2}{m} + \frac{1}{r} \right) \sigma_{\omega}^2 + \frac{\theta^2(m-1)}{m}\psi^B + \frac{r-1}{r}\psi^A - 2\theta\psi^X \end{aligned}$$

This implies the equation system

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

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

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

$$k_A = \left( \frac{I}{I-1} \right) \frac{m + \theta r}{2m^2 r} (r-1)(m + \theta r + (1-\theta)mr)$$

$$k_X = 0$$

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  $\theta$  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  $\theta$ ) as well as  $k_v^D, k_\omega^D, k_B^D, k_A^D, k_X^D$  (corresponding to the denominator) such that

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

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

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

$$k_X^D = 0$$

which gives  $\theta$  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  $\frac{I}{I-1}$ .