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, June 2020

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

UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

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

Claes Ek[a]

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

Abstract

Recent research by Burlig et al. (2020) has produced a useful formula for performing difference-in-differences power calculation in the presence of serially correlated errors. A similar formula for the ANCOVA estimator is shown by the authors to yield incorrect power in real data where time shocks are present. This note demonstrates that the serial-correlation-robust ANCOVA formula is in fact correct under time shocks as well. The reason that errors arise in Burlig et al. (2020) is because time shocks remain unaccounted for in the intermediate step where residual-based variance parameters are estimated from pre-existing data. When that procedure is adjusted accordingly, the serial-correlation-robust ANCOVA formula of Burlig et al. (2020) can be accurately used for power calculation.

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

JEL classification: C93, C23

In a recent paper, Burlig et al. (2020) derive a set of variance formulas for *ex-ante* power calculation in panel data with serially correlated errors. Accounting for serial correlation is important, since it is likely to occur in many real-world settings, e.g. whenever outcomes that occur close in time are more highly correlated than more distant ones. For the difference-in-differences estimator, the authors show that earlier power formulas that fail to account for serial correlation yield incorrect power. By contrast, their novel serial-correlation-robust power formula accurately predicts statistical power in simulated as well as actual data. These methods and results are likely to prove highly useful to any researcher planning experiments with multiple measurements.

The authors focus on difference-in-differences rather than the analysis-of-covariance (ANCOVA) estimator. They do note that the latter estimator is more efficient than the former,

☆This version: 25 June 2020. *Email address:* `claes.ek@economics.gu.se`

and thus may be preferable in randomized settings where time fixed effects are not needed for identification. However, when time shocks are present in the data generating process (DGP), deriving the ANCOVA regression variance for any panel length 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

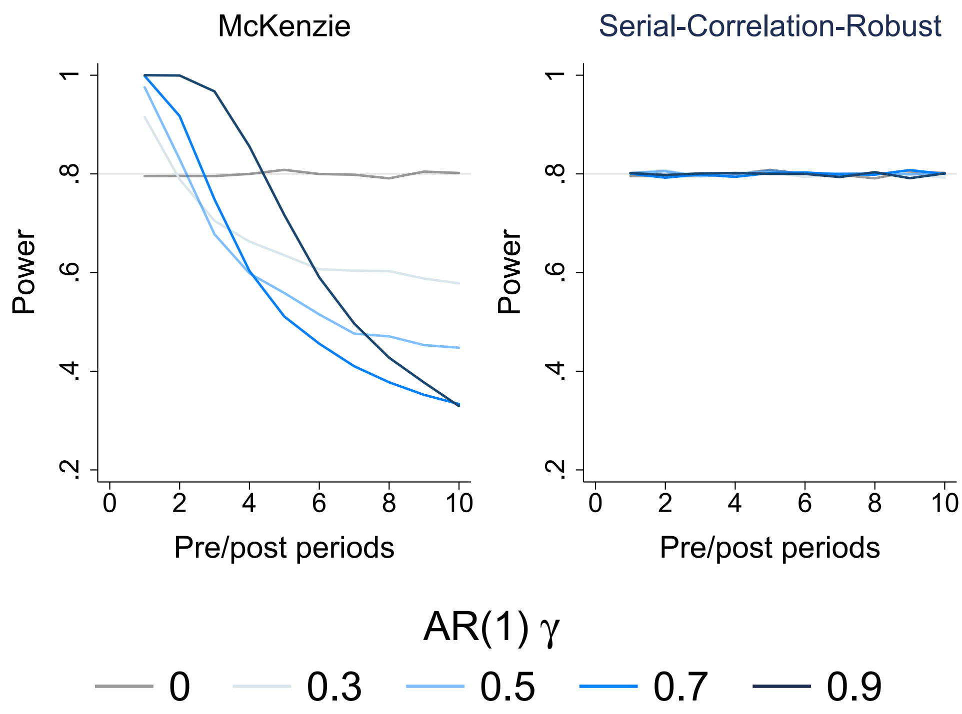
$$\begin{aligned}\mathrm{Var}\left(\hat{\tau}|\mathbf{X}\right) = &\left(\frac{1}{P(1-P)J} + \frac{\left(\bar{Y}_T^B - \bar{Y}_C^B\right)^2}{Z}\right)\\ &\times \left(\left(1-\theta\right)^2 \sigma_v^2 + \left(\frac{\theta^2}{m} + \frac{1}{r}\right)\sigma_\omega^2 + \frac{\theta^2(m-1)}{m}\psi^B + \frac{r-1}{r}\psi^A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

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 produce nominal 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 note is to demonstrate, first, that ANCOVA power formula (1) is in fact correct even in the presence of time shocks; or equivalently, that such effects do not affect ANCOVA precision. This result 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. Second, I show 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 calculations for ANCOVA in the presence of both serial correlation and time shocks. These findings should prove useful, given that ANCOVA is arguably the estimator of choice in panel experimental settings (McKenzie, 2012).

As a first indication that time shocks do not impact ANCOVA precision, consider Figure 1. It is a variation of Figure 4 in Burlig et al. (2020), where the authors check whether formula (1) accurately predicts power in simulated data. The DGP underlying the original figure includes only a single intercept rather than a set of time shocks δ_t (i.e., $\sigma_\delta^2 = 0$), and is thus consistent with the analytical model of the authors. By contrast, Figure 1 adds normally distributed time shocks with $\sigma_\delta^2 = 10$ and also estimates time fixed effects in each *ex-post* ANCOVA regression. I retain all other assumptions, steps, and parameter values underlying the original figure (as described in Appendix B.1 of Burlig et al., 2020). Despite

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 present in the data. As in Figure 4 of Burlig et al. (2020), both panels cluster standard errors by unit *ex post* and are based on 10,000 draws from a population where idiosyncratic error term ω_{it} follows an AR(1) process with autoregressive parameter γ. In the left panel, the size of the MDEs are calibrated *ex ante* using the McKenzie (2012) power formula. The right panel instead instead uses the Burlig et al. (2020) serial-correlation-robust ANCOVA power formula (1). 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 normally distributed time shocks with $\mu_\delta = 20$, $\sigma^2_\delta = 10$ are included. Despite this, the SCR ANCOVA formula yields appropriate rejection rates.

the addition of time shocks in Figure 1, rejection rates are clearly practically identical to the original figure. In particular, rejection rates corresponding to serial-correlation-robust formula (1) yield nominal power. Using other values of σ^2_δ (including very large ones, such as $\sigma^2_\delta = 1000$) does not alter these results.

In Appendix A of this note, I present analytical proofs mirroring these findings. Specifically: consider the DGP

$$Y_{it} = \delta_t + \tau D_{it} + v_i + \omega_{it} \tag{2}$$

with time shocks δ_t, distributed i.i.d. $\mathcal{N}(\mu_\delta, \sigma^2_\delta)$; treatment indicator D_{it}; unit intercept v_i; and serially correlated idiosyncratic error ω_{it}. For this model, I am able to show that ANCOVA variance is exactly equal to formula (1).[1] In fact, equation (1) applies both when

[1]Equation (2) admittedly differs slightly from the model stated as Assumption 1 in Burlig et al. (2020),

time fixed effects are included in the ANCOVA regression (see 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.[2]

While somewhat technical, the proof has an overall structure highly similar to that of Burlig et al. (2020). As in their analysis without time shocks, calculating the variance of the ANCOVA estimator involves evaluating the expression

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

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 the regression residual for unit i and period t.

The main difficulty in evaluating this expression concerns conditional means $\mathrm{E}[\epsilon_{it}\epsilon_{js}|\mathbf{X}]$. In Burlig et al. (2020), conditioning on $\mathbf{X}$ amounts to conditioning only on the baseline averages of 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. The authors then show that, under such conditions, $\sum_{t=1}^{r}\sum_{s=1}^{r}\mathrm{E}[\epsilon_{it}\epsilon_{js}|\mathbf{X}] = 0$ whenever $i \neq j$; hence, the variance of ANCOVA is composed solely of those terms where $i = j$.

By contrast, when time shocks are included in the DGP, not only must shocks t and s be added as conditioning variables in $\mathrm{E}[\epsilon_{it}\epsilon_{js}|\mathbf{X}] = 0$, but so must the baseline averages of *all* other units in the experiment. The reason is that each conditioning baseline average now provides additional information about the average pre-treatment time shocks; and those pre-treatment shocks are themselves included in both residuals.

I then show that, as a result, we no longer have $\sum_{t=1}^{r}\sum_{s=1}^{r}\mathrm{E}[\epsilon_{it}\epsilon_{js}|\mathbf{X}] = 0$ when $i \neq j$. Instead, $\sum_{t=1}^{r}\sum_{s=1}^{r}\mathrm{E}[\epsilon_{it}\epsilon_{js}|\mathbf{X}]$ takes one value when $i \neq j$, and takes the same value plus a difference term whenever $i = j$. Both expressions are otherwise invariant across i, j. The

since that DGP includes both time shocks and a constant term β. However, the discrepancy is innocuous, since it can be reconciled simply by viewing each time shock in (2) as $\delta_t = \beta + \delta_t'$, with δ_t' having mean zero and variance σ_δ^2.

[2] Running the *ex-post* regressions underlying Figure 1 without time FE confirms this point.

$i \neq j$ value reflects variation associated with the time shocks; since it is summed across both $i = j$ and $i \neq j$, it will be multiplied by

$$\sum_{i=1}^{PJ}\sum_{j=1}^{PJ} M_{ij}^T + \sum_{i=1}^{PJ}\sum_{j=PJ+1}^{J} M_{ij}^X + \sum_{i=PJ+1}^{J}\sum_{j=PJ+1}^{J} M_{ij}^C$$

which can be shown to equal zero. Thus, only the difference term remains, and that turns out to be 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.

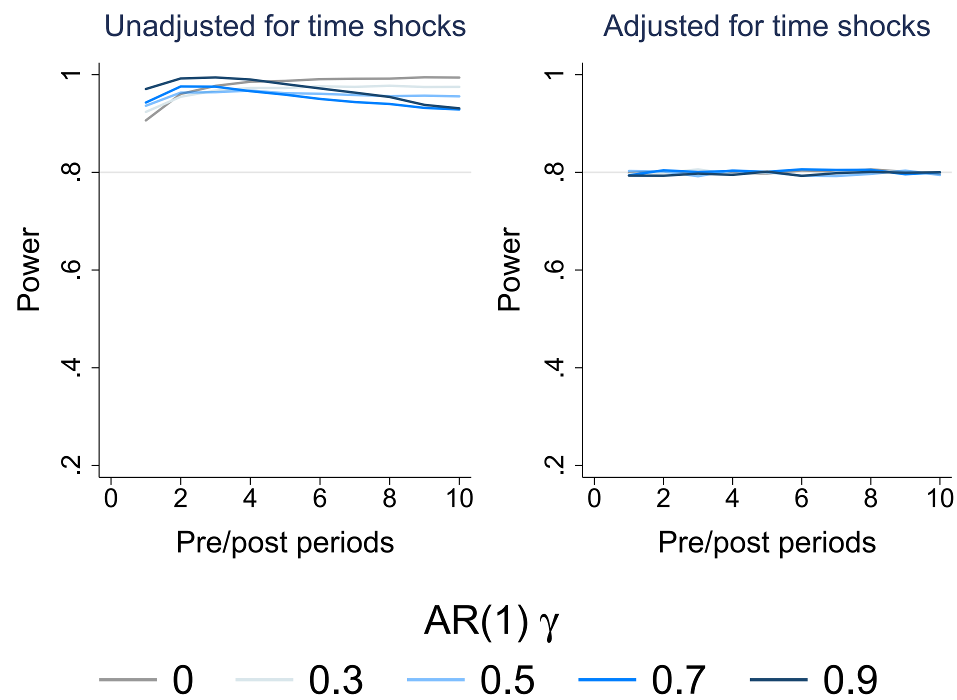
An obvious question remains: if the ANCOVA variance formula derived by Burlig et al. (2020) is correct after all, what might account for the inaccurate rejection rates they obtain using real data? The answer is the following.

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. In a reasonable attempt to remain consistent with their assumed time-shock free DGP, they ignore the possibility of time shocks throughout this 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 factors that *do* 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.

In Figure 2, I compare the two approaches for simulated data. The figure is based on the same model and parameters as 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 panel (a), I follow exactly the procedure described for ANCOVA in Appendix E.3 of Burlig et al. (2020);[3] as expected, this procedure ignores the presence of time shocks and consequently yields excessively high rejection rates. In panel (b), I modify

[3]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 $\frac{IT-1}{IT}$ (T being the panel length) and $\tilde{\sigma}^2_{\hat{v},S}$ by $\frac{I-1}{I}$ but leave the $\tilde{\psi}$ estimates unadjusted, in accordance with the discussion of e.g. $\mathrm{E}[\tilde{\psi}^B_{\hat{\omega}}]$ in Appendix E.3 of Burlig et al. (2020).

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

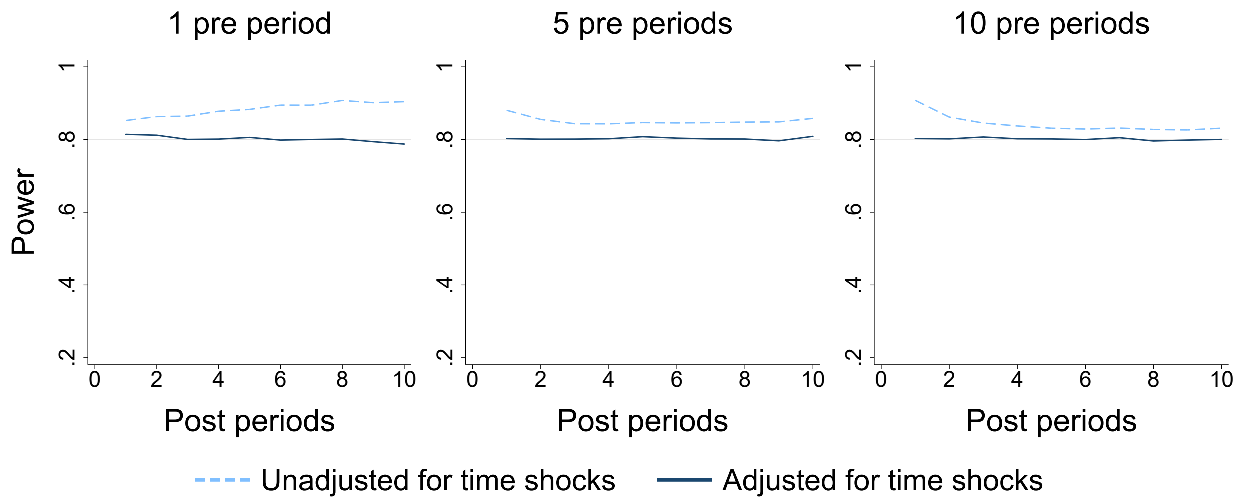


Note: The figure depicts rejection rates for the regression ANCOVA estimator when time shocks are present in the data. Both panels are based on 10,000 draws from a population where idiosyncratic error term ω_{it} follows an AR(1) process with autoregressive parameter γ. The DGP and all associated parameter values are as in Figure 1. Both panels 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 time and unit FE; minor adjustments are also made to the MDE calculation, as described in Appendix B of this note. These adjustments result in appropriate rejection rates. Both panels estimate ANCOVA *ex post*, clustering standard errors by unit; however, ANCOVA regressions include time FE only in the right-hand panel.

the procedure to correctly take time shocks into account (details are given in Appendix B of this note); when this is done, nominal power is attained.

Then, in Figure 3, I repeat the exercise for real data, specifically the Bloom et al. (2015) data set used for Figure 7 of Burlig et al. (2020). When not accounting for time shocks (dashed lines), I am able to closely replicate the original figure. When instead I account for time shocks in the proper way (solid lines), appropriate rejection rates are again achieved. This demonstrates the feasibility, adding only minor modifications, of using the Burlig et al. (2020) approach to perform an accurate ANCOVA power calculation 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 SCR 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 note. 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.

References

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, 2020.

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

Online Appendices for article "ANCOVA power calculation under serial correlation and time shocks: A comment on Burlig et al. (2020)"

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 part of Assumption 1 related to time shocks, which has 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; 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, in the first departure from the Burlig et al. (2020) model, 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 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 $\mathbf{D}$, and $J-1$ unit 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:*

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

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, first, when time shocks are included in the regression equation to be estimated; and second, when they are not. In both cases, the result will be equal to the variance calculated as equation A61 in Burlig et al. (2020).

A.1 Time shocks included in ANCOVA regression

Consider the following updated ANCOVA regression model:

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

where Y_{it}, τ, and D_i are defined as above; also,

$$\theta = \frac{m\left(\sigma_v^2 + \psi^X\right)}{m\sigma_v^2 + \sigma_\omega^2 + (m-1)\,\psi^B}$$

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

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

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}|\mathbf{X}) = (\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}'\mathrm{E}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}'|\mathbf{X}]\mathbf{X}\,(\mathbf{X}'\mathbf{X})^{-1} \tag{A.3}$$

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

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

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

$$\left.\begin{array}{cc} \sum_{i=1}^{PJ}\sum_{j=1}^{J}\sum_{t=1}^{r}\mathrm{E}[\epsilon_{it}\epsilon_{j1}|\mathbf{X}] & \sum_{i=1}^{J}\sum_{j=1}^{J}\sum_{t=1}^{r}\bar{Y}_i^B\mathrm{E}[\epsilon_{it}\epsilon_{j1}|\mathbf{X}] \\ \vdots & \vdots \\ \sum_{i=1}^{PJ}\sum_{j=1}^{J}\sum_{t=1}^{r}\mathrm{E}[\epsilon_{it}\epsilon_{jr}|\mathbf{X}] & \sum_{i=1}^{J}\sum_{j=1}^{J}\sum_{t=1}^{r}\bar{Y}_i^B\mathrm{E}[\epsilon_{it}\epsilon_{jr}|\mathbf{X}] \\ \sum_{i=1}^{PJ}\sum_{j=1}^{PJ}\sum_{t=1}^{r}\sum_{s=1}^{r}\mathrm{E}[\epsilon_{it}\epsilon_{js}|\mathbf{X}] & \sum_{i=1}^{PJ}\sum_{j=1}^{J}\sum_{t=1}^{r}\sum_{s=1}^{r}\bar{Y}_j^B\mathrm{E}[\epsilon_{it}\epsilon_{js}|\mathbf{X}] \\ \sum_{i=1}^{PJ}\sum_{j=1}^{J}\sum_{t=1}^{r}\sum_{s=1}^{r}\bar{Y}_j^B\mathrm{E}[\epsilon_{it}\epsilon_{js}|\mathbf{X}] & \sum_{i=1}^{J}\sum_{j=1}^{J}\sum_{t=1}^{r}\sum_{s=1}^{r}\bar{Y}_i^B\bar{Y}_j^B\mathrm{E}[\epsilon_{it}\epsilon_{js}|\mathbf{X}] \end{array}\right) \tag{A.4}$$

Next, consider inverting $(1/J)\mathbf{X}'\mathbf{X}$, which is the following square matrix of dimension

$r+2$:

$$\frac{1}{J}\mathbf{X}'\mathbf{X} = \begin{pmatrix} 1 & \cdots & 0 & P & \bar{Y}^B \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & 1 & P & \bar{Y}^B \\ P & \cdots & P & rP & rP\bar{Y}^B_T \\ \bar{Y}^B & \cdots & \bar{Y}^B & rP\bar{Y}^B_T & \frac{r}{J}\sum_{j=1}^{J}\left(\bar{Y}^B_i\right)^2 \end{pmatrix}$$

where, due to the inclusion of time fixed effects in the regression, the first r rows and columns of the matrix form a nested identity matrix; note that

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

where $Z = \sum_{k=1}^{PJ}\left(\bar{Y}^B_k - \bar{Y}^B_T\right)^2 + \sum_{k=PJ+1}^{J}\left(\bar{Y}^B_k - \bar{Y}^B_C\right)^2$.

The following lemmas will prove useful for inverting $(1/J)\mathbf{X}'\mathbf{X}$.

Lemma 1. *Any matrix of the form*

$$\mathbf{Y} = \begin{pmatrix} 1 & \cdots & 0 & x_1 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \vdots & 1 & x_1 \\ x_2 & \cdots & x_2 & x_3 \end{pmatrix}$$

with nested identity matrix of dimension r, has $|\mathbf{Y}| = x_3 - rx_1x_2$.

Proof. The argument is recursive: assuming the lemma holds when the nested identity matrix has dimension $r-1$, cofactor expansion along the first row of $\mathbf{Y}$ yields

$$|\mathbf{Y}| = x_3 - (r-1)x_1x_2 + (-1)^r x_1\left((-1)^{r-1}x_2|I_{r-1}|\right) = x_3 - rx_1x_2$$

where the second term relies on expansion of the $(1, r+1)$ cofactor along the first column of the corresponding submatrix; I_{r-1} is an $(r-1)$-dimensional identity matrix. Finally, if $r=1$, clearly $|\mathbf{Y}| = x_3 - x_1 x_2$. $\square$

Lemma 2. *Any matrix of the form*

$$\mathbf{Y} = \begin{pmatrix} 0 & 1 & \cdots & 0 & x_1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \vdots & 1 & x_1 \\ x_2 & x_2 & \cdots & x_2 & x_3 \\ x_4 & x_4 & \cdots & x_4 & x_5 \end{pmatrix}$$

with nested identity matrix of dimension r, has $|\mathbf{Y}| = (-1)^r (x_2 x_5 - x_3 x_4)$.

Proof. By Lemma 1, cofactor expansion along the first column yields

$$|\mathbf{Y}| = (-1)^r x_2 (x_5 - r x_1 x_4) + (-1)^{r+1} x_4 (x_3 - r x_1 x_2) = (-1)^r (x_2 x_5 - x_3 x_4)$$

$\square$

Lemma 3. *Any matrix of the form*

$$\mathbf{Y} = \begin{pmatrix} 0 & 0 & \cdots & 0 & x_1 & x_2 \\ 0 & 1 & \cdots & 0 & x_1 & x_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \vdots & 1 & x_1 & x_2 \\ x_3 & x_3 & \cdots & x_3 & x_4 & x_5 \\ x_6 & x_6 & \cdots & x_6 & x_7 & x_8 \end{pmatrix}$$

with nested identity matrix of dimension r, has $|\mathbf{Y}| = x_2 (x_3 x_7 - x_4 x_6) - x_1 (x_3 x_8 - x_5 x_6)$.

Proof. By Lemma 2, cofactor expansion along the first row yields

$$\begin{aligned} |\mathbf{Y}| &= (-1)^{r+1} x_1 [(-1)^r (x_3 x_8 - x_5 x_6)] + (-1)^{r+2} x_2 [(-1)^r (x_3 x_7 - x_4 x_6)] \\ &= x_2 (x_3 x_7 - x_4 x_6) - x_1 (x_3 x_8 - x_5 x_6) \end{aligned}$$

$\square$

Lemma 4. *Any matrix of the form*

$$\mathbf{Y} = \begin{pmatrix} 1 & \cdots & 0 & x_1 & x_2 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \vdots & 1 & x_1 & x_2 \\ x_3 & \cdots & x_3 & x_4 & x_5 \\ x_6 & \cdots & x_6 & x_7 & x_8 \end{pmatrix}$$

with nested identity matrix of dimension r, has $|\mathbf{Y}| = x_4x_8 - x_5x_7 - rx_1(x_3x_8 - x_5x_6) + rx_2(x_3x_7 - x_4x_6)$.

Proof. Assuming the result holds when the nested identity matrix has dimension $r-1$, Lemma 2 implies that cofactor expansion along the first row of $\mathbf{Y}$ yields

$$\begin{aligned} |\mathbf{Y}| &= x_4x_8 - x_5x_7 - (r-1)x_1(x_3x_8 - x_5x_6) + (r-1)x_2(x_3x_7 - x_4x_6) \\ &\quad + (-1)^r x_1\left[(-1)^{r-1}(x_3x_8 - x_5x_6)\right] + (-1)^{r+1}x_2\left[(-1)^{r-1}(x_3x_7 - x_4x_6)\right] \\ &= x_4x_8 - x_5x_7 - rx_1(x_3x_8 - x_5x_6) + rx_2(x_3x_7 - x_4x_6) \end{aligned}$$

Finally, for $r = 1$, $|\mathbf{Y}| = x_4x_8 - x_5x_7 - x_1(x_3x_8 - x_5x_6) + x_2(x_3x_7 - x_4x_6)$. □

We may now proceed to invert $(1/J)\mathbf{X}'\mathbf{X}$. By Lemma 4, $|(1/J)\mathbf{X}'\mathbf{X}| = (r^2P(1-P)Z)/J$ and diagonal cofactors $C_{11} = C_{22} = ... = C_{rr} = rP\left((1/J)rZ + (1-P)\left(\bar{Y}_C^B\right)^2\right)$. Lemma 1 implies $C_{(r+1)(r+1)} = r\left(Z/J + P(1-P)\left(\bar{Y}_T^B - \bar{Y}_C^B\right)^2\right)$ and well as $C_{(r+2)(r+2)} = rP(1-P)$.

Next, Lemma 3 implies $C_{12} = rP\left((1/J)PZ + (1-P)\left(\bar{Y}_C^B\right)^2\right)$. We claim that all other cofactors C_{ij} with $i \leq r$, $j \leq r$, and $i \neq j$ will also be equal to this value. Consider such a cofactor $C_{ij} \neq C_{12}$, with $i < j$, and suppose the claim applies for some $C_{(i-1)j}$ with $i < j$, or some $C_{i(j-1)}$ with $i < j$; at least one of these cofactors must exist. Now, the $(r+1) \times (r+1)$ submatrix with determinant C_{ij} will have the first $r-1$ elements of column i, as well as the first $r-1$ elements of row $j-1$, equal to zero. Moreover, the remainder of the first $r-1$ rows and columns of the submatrix form an $(r-2)$-dimensional identity matrix into which the zeroes of column $j-1$ and row i have effectively been inserted. It follows that interchanging a single column (row) of the submatrix of $C_{(i-1)j}$ ($C_{i(j-1)}$) again yields the submatrix of C_{ij}. Since $(-1)^{i-1+j} = -(-1)^{i+j}$, we will have $C_{(i-1)j} = C_{ij}$, with an analogous statement for $C_{i(j-1)}$. For cofactors with $i > j$, the claim follows by the symmetry of $(1/J)\mathbf{X}'\mathbf{X}$.

Lemma 2 implies $C_{1(r+1)} = rP\left(-Z/J + (1-P)\bar{Y}_C^B\left(\bar{Y}_T^B - \bar{Y}_C^B\right)\right)$. Similarly to above, the submatrix of any $C_{i(r+1)}$ with $i \leq r$ inserts a vector, the first $r-1$ elements of which are equal to zero, into column i; the first r columns and $r-1$ rows are otherwise given by an $(r-1)$-dimensional identity matrix. Thus, an analogous argument to that made above implies that $C_{i(r+1)} = C_{1(r+1)} = C_{(r+1)i}$ for all $i \leq r$. Lemma 2 also implies $C_{1(r+2)} = -rP(1-P)\bar{Y}_C^B$, with all $C_{i(r+2)}$ and $C_{(r+2)i}$ where $i \leq r$ similarly equal to this quantity. Finally, by Lemma 1, $C_{(r+1)(r+2)} = C_{(r+2)(r+1)} = -rP(1-P)\left(\bar{Y}_T^B - \bar{Y}_C^B\right)$.

In summary, since $(\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{J}\left(\frac{1}{J}\mathbf{X}'\mathbf{X}\right)^{-1}$, we have

$$(\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{r(1-P)Z} \times \left(\begin{array}{ccccc} \frac{rZ}{J} + (1-P)\left(\bar{Y}_C^B\right)^2 & \cdots & \frac{PZ}{J} + (1-P)\left(\bar{Y}_C^B\right)^2 & -\frac{Z}{J} + (1-P)\bar{Y}_C^B\left(\bar{Y}_T^B - \bar{Y}_C^2\right) & -(1-P)\bar{Y}_C^B \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ \frac{PZ}{J} + (1-P)\left(\bar{Y}_C^B\right)^2 & \cdots & \frac{rZ}{J} + (1-P)\left(\bar{Y}_C^B\right)^2 & -\frac{Z}{J} + (1-P)\bar{Y}_C^B\left(\bar{Y}_T^B - \bar{Y}_C^2\right) & -(1-P)\bar{Y}_C^B \\ -\frac{Z}{J} + (1-P)\bar{Y}_C^B\left(\bar{Y}_T^B - \bar{Y}_C^B\right) & \cdots & -\frac{Z}{J} + (1-P)\bar{Y}_C^B\left(\bar{Y}_T^B - \bar{Y}_C^B\right) & \frac{Z}{PJ} + (1-P)\left(\bar{Y}_T^B - \bar{Y}_C^B\right)^2 & -(1-P)\left(\bar{Y}_T^B - \bar{Y}_C^B\right) \\ -(1-P)\bar{Y}_C^B & \cdots & -(1-P)\bar{Y}_C^B & -(1-P)\left(\bar{Y}_T^B - \bar{Y}_C^B\right) & 1-P \end{array}\right) \tag{A.5}$$

and may combine (A.5) with (A.4) to calculate element $(r+1, r+1)$ of (A.3) as

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

$$\times \Big(- Z - (1-P)J\left(\bar{Y}_T^B - \bar{Y}_C^B\right)\left(\bar{Y}_j^B - \bar{Y}_C^B\right)\Big) \times \left(\sum_{t=1}^{r} \sum_{s=1}^{r} \mathrm{E}[\epsilon_{it}\epsilon_{js}|\mathbf{X}] \right) \Bigg] \tag{A.6}$$

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.[4] To do so, we first need to compute the summed conditional means included in each of the three terms in (A.6).

For a given single conditional mean with $i \neq j$ as well as $t \neq s$,

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

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, conditioning on $\mathbf{X}$ implies conditioning on *all* baseline averages in the experiment. This is because each conditioning baseline average provides additional information about the average pre-treatment time shock included in both residuals through baseline averages i and j. Also note that while δ_s is unconditionally independent of ϵ_{it}, it must still be retained as conditioning variable in the inner expectation. The reason is somewhat subtle: conditional on ϵ_{js} and $\bar{Y}_j^B$, δ_s provides information on e.g. v_j; but conditional on $\bar{Y}_j^B$, v_j is itself informative about the pre-treatment time shocks included in ϵ_{it}.

When $i = j$ and/or $t = s$, the above expectation is adjusted accordingly. For example, when $i \neq j$ but $t = s$, we have

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

Since both residuals as well as all conditioning variables are assumed normally distributed, we may evaluate any conditional mean using the following formula:

$$\mathrm{E}[x|\mathbf{y}] = \mu_x + \mathbf{\Sigma_{xy}}\mathbf{\Sigma_{yy}^{-1}}(\mathbf{y} - \boldsymbol{\mu}_\mathbf{y}) \tag{A.7}$$

where μ_x is the mean of the normal variable x; $\mathbf{\Sigma_{xy}}$ is a row vector collecting the covariances between x and each element of the vector of normally distributed conditioning variables $\mathbf{y}$;

[4] If time shocks are assumed fixed rather than stochastic, all subsequent steps to derive the ANCOVA variance will be identical to those in Burlig et al. (2020), with the result that their formula is again appropriate.

$\boldsymbol{\Sigma}_{\mathbf{yy}}^{-1}$ is the inverted 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 both residuals have mean zero by the properties of linear projection. Also, $\mathrm{E}(\bar{Y}_i^B) = \mathrm{E}(\delta_t) = \mu_\delta$ for all i and t.

Like Burlig et al. (2020), we will begin by considering the case when $i \neq j$. For $t \neq s$, the $(J+3)$-dimensional covariance matrix corresponding to the above inner (nested) conditional expectation is

$$\boldsymbol{\Sigma}_{\mathbf{yy}}^{t\neq s} = \begin{pmatrix} a_s & me & 0 & b_s & c & \cdots & c \\ me & me & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & me & 0 & 0 & \cdots & 0 \\ b_s & 0 & 0 & d & e & \cdots & e \\ c & 0 & 0 & e & d & \cdots & e \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ c & 0 & 0 & e & e & \cdots & d \end{pmatrix} \tag{A.8}$$

where the final bottom-right J rows and columns all have d as diagonal elements, and e as off-diagonal elements. For convenience, the matrix uses the following parameter definitions.

$$\begin{aligned}
a_s &= \mathrm{Var}(\epsilon_{js}) = (1-\theta)^2\sigma_v^2 + \left(1 + \frac{\theta^2}{m}\right)\left(\sigma_\omega^2 + \sigma_\delta^2\right) - \theta \mathrm{Cov}\left(\omega_{js}, \bar{\omega}_j^B\right) + \frac{\theta^2(m-1)}{m}\psi^B \\
b_s &= \mathrm{Cov}\left(\epsilon_{js}, \bar{Y}_j^B\right) = \mathrm{Cov}\left(\omega_{js}, \bar{\omega}_j^B\right) - \psi^X - \frac{\theta\sigma_\delta^2}{m} \\
c &= \mathrm{Cov}\left(\epsilon_{js}, \bar{Y}_i^B\right) = \mathrm{Cov}\left(\epsilon_{it}, \bar{Y}_j^B\right) = -\frac{\theta\sigma_\delta^2}{m} \\
d &= \mathrm{Var}\left(\bar{Y}_i^B\right) = \frac{1}{m}\left(\sigma_\delta^2 + \sigma_\omega^2 + m\sigma_v^2 + (m-1)\psi^B\right) \\
e &= \mathrm{Cov}\left(\bar{Y}_i^B, \bar{Y}_j^B\right) = \frac{\sigma_\delta^2}{m}
\end{aligned}$$

and $\bar{\omega}_j^B = (1/m)\sum_{p=-m+1}^{0}\omega_{jp}$. Note that $\sum_s b_s = rc$.

Furthermore,

$$\boldsymbol{\Sigma}_{\mathbf{xy}}^{t\neq s} = \begin{pmatrix} f^{i\neq j} & 0 & me & c & b_t & c & \cdots & c \end{pmatrix} \tag{A.9}$$

where

$$\begin{aligned}
b_t &= \mathrm{Cov}\left(\epsilon_{it}, \bar{Y}_i^B\right) = \mathrm{Cov}\left(\omega_{it}, \bar{\omega}_i^B\right) - \psi^X - \frac{\theta\sigma_\delta^2}{m} \\
f^{i\neq j} &= \mathrm{Cov}\left(\epsilon_{it}, \epsilon_{js}\right) - \mathrm{Cov}\left(\delta_t, \delta_s\right) = \frac{\theta^2\sigma_\delta^2}{m}
\end{aligned}$$

and similarly to above, $\sum_t b_t = rc$.

When $t = s$, $\mathbf{\Sigma}_{\mathbf{yy}}^{t=s}$ is the $(J+2)$-dimensional submatrix that results when row and column 3 (corresponding to δ_t) are dropped from (A.8). We also have

$$\mathbf{\Sigma}_{\mathbf{xy}}^{t=s} = \begin{pmatrix} f^{i=j} + me & me & c & b_s & c & \cdots & c \end{pmatrix} \tag{A.10}$$

Our next objective is to invert the $\mathbf{\Sigma}_{\mathbf{yy}}$ matrices. To that end, we make use of the following lemma.

Lemma 5. *Any n-dimensional square matrix of the form*

$$\mathbf{Y}_1 = \begin{pmatrix} d & e & \cdots & e \\ e & d & \cdots & e \\ \vdots & \vdots & \ddots & \vdots \\ e & e & \cdots & d \end{pmatrix}$$

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

$$\mathbf{Y}_2 = \begin{pmatrix} e & e & \cdots & e \\ e & d & \cdots & e \\ \vdots & \vdots & \ddots & \vdots \\ e & e & \cdots & d \end{pmatrix}$$

has $|\mathbf{Y}_2| = e\,(d-e)^{n-1}$.

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 $\mathbf{Y}_2$):

$$\begin{aligned} |\mathbf{Y}_1| &= d\left[(d-e)^{n-2}\,(d+(n-2)\,e)\right] - (n-1)e^2\,(d-e)^{n-2} \\ &= (d-e)^{n-1}\,(d+(n-1)\,e) \end{aligned}$$

and

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

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

Lemma 5 permits calculation of a large number of similar determinants, ensuring that inverting $\mathbf{\Sigma_{yy}}$ is feasible. These corollaries are straightforward but too numerous to list in their entirety; however, the overall procedure is highly similar to that used when inverting $\mathbf{X'X}$ above. One particularly useful example follows:

$$|\mathbf{\Sigma}^{t=s}_{\mathbf{yy}}| = \begin{vmatrix} a_s & me & b_s & c & \cdots & c \\ me & me & 0 & 0 & \cdots & 0 \\ b_s & 0 & d & e & \cdots & e \\ c & 0 & e & d & \cdots & e \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ c & 0 & e & e & \cdots & d \end{vmatrix} = me \begin{vmatrix} a_s & b_s & c & \cdots & c \\ b_s & d & e & \cdots & e \\ c & e & d & \cdots & e \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c & e & e & \cdots & d \end{vmatrix} - me \begin{vmatrix} me & 0 & 0 & \cdots & 0 \\ b_s & d & e & \cdots & e \\ c & e & d & \cdots & e \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c & e & e & \cdots & d \end{vmatrix}$$

$$= a_s me\,(d-e)^{J-1}\,(d+(J-1)\,e) - b_s me \begin{vmatrix} b_s & c & \cdots & c \\ e & d & \cdots & e \\ \vdots & \vdots & \ddots & \vdots \\ e & e & \cdots & d \end{vmatrix} + (J-1)mce \begin{vmatrix} b_s & c & c & \cdots & c \\ d & e & e & \cdots & e \\ e & e & d & \cdots & e \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ e & e & e & \cdots & d \end{vmatrix}$$

$$- m^2e^2(d-e)^{J-1}(d+(J-1)e)$$

$$= me(a_s - me)(d-e)^{J-1}(d+(J-1)e) - b_s me\Big(b_s(d-e)^{J-2}(d+(J-2)e) - (J-1)ce(d-e)^{J-2}\Big)$$

$$+ (J-1)mce\Big((b_s + (J-2)c)e(d-e)^{J-2} - c(d-e)^{J-2}(d+(J-2)e)\Big)$$

$$= me(d-e)^{J-2}\Big((d-e)\big((a_s - me)(d+(J-1)e) - b_s^2 - (J-1)c^2\big) - (J-1)e(b_s - c)^2\Big)$$

$$\equiv |\mathbf{\Sigma}|$$

Notice that this determinant does not depend on t. Expressions may be similarly derived for all cofactors of the covariance matrices, yielding the symmetric inverse

$$\left(\mathbf{\Sigma}^{t\neq s}_{\mathbf{yy}}\right)^{-1} = \frac{(d-e)^{J-2}}{|\mathbf{\Sigma}|} \left(\begin{matrix} me(d-e)(d+(J-1)e) & -me(d-e)(d+(J-1)e) \\ -me(d-e)(d+(J-1)e) & (d-e)(a_s(d+(J-1)e) - b_s^2 - c^2) - (J-1)e(b_s-c)^2 \\ 0 & 0 \\ -me(b_s(d+(J-2)e) - (J-1)ce) & me(b_s(d+(J-2)e) - (J-1)ce) \\ me(b_se - cd) & -me(b_se - cd) \\ \vdots & \vdots \\ me(b_se - cd) & -me(b_se - cd) \end{matrix} \right.$$

$$
\begin{matrix}
0 & -me(b_s(d+(J-2)e)-(J-1)ce) & me(b_se-cd) \\
0 & me(b_s(d+(J-2)e)-(J-1)ce) & -me(b_se-cd) \\
\frac{|\boldsymbol{\Sigma}|}{me(d-e)^{J-2}} & 0 & 0 \\
0 & me((a_s-me)(d+(J-2)e)-(J-1)c^2) & me(-e(a_s-me)+b_sc) \\
0 & me(-e(a_s-me)+b_sc) & me\Big((a_s-me)(d+(J-2)e)-b_s^2-(J-2)c^2-\frac{(J-2)e(b_s-c)^2}{d-e}\Big) \\
\vdots & \vdots & \vdots \\
0 & me(-e(a_s-me)+b_sc) & -me\Big(e(a_s-me)-\frac{b_se(b_s-2c)+c^d}{d-e}\Big)
\end{matrix}
$$

$$
\left.
\begin{matrix}
\cdots & me(b_se-cd) \\
\cdots & -me(b_se-cd) \\
\cdots & 0 \\
\cdots & me(-e(a_s-me)+b_sc) \\
\cdots & -me\Big(e(a_s-me)-\frac{b_se(b_s-2c)+c^2d}{d-e}\Big) \\
\ddots & \vdots \\
\cdots & me\Big((a_s-me)(d+(J-2)e)-b_s^2-(J-2)c^2-\frac{(J-2)e(b_s-c)^2}{d-e}\Big)
\end{matrix}
\right) \tag{A.11}
$$

Furthermore, $\left(\boldsymbol{\Sigma}_{\mathbf{yy}}^{t=s}\right)^{-1}$ is equal to the submatrix that results when row and column 3 are dropped from (A.11). Inserting expressions (A.11) and (A.9), or $\left(\boldsymbol{\Sigma}_{\mathbf{yy}}^{t=s}\right)^{-1}$ and (A.10), into formula (A.7) yields the inner expectation $E[\epsilon_{it}|\cdot]$ under $t \neq s$ and $t = s$, respectively. It turns out that the results of both cases can be combined into the single expression

$$
\begin{aligned}
\mathrm{E}[\epsilon_{it}|\epsilon_{js},\delta_s,\delta_t,\bar{Y}_j^B,\bar{Y}_i^B,\bar{Y}_{-i,-j}^B] = {} & A_1^{i\neq j}\epsilon_{js} - A_1^{i\neq j}\left(\delta_s-\mu_\delta\right) + \delta_t - \mu_\delta + A_2^{i\neq j}\left(\bar{Y}_j^B-\mu_\delta\right) \\
& + A_3^{i\neq j}\left(\bar{Y}_i^B-\mu_\delta\right) + A_4^{i\neq j}\left(\sum_{k\neq i,j}\left(\bar{Y}_k^B-\mu_\delta\right)\right)
\end{aligned}
$$

where

$$
\begin{aligned}
A_1^{i\neq j} &= \frac{me(d-e)^{J-2}}{|\boldsymbol{\Sigma}|}\left((d-e)\Big(f^{i\neq j}(d+(J-1)e)-(J-1)c^2\Big)+cd(c-b_t-b_s)+b_tb_se\right) \\
A_2^{i\neq j} &= \frac{me(d-e)^{J-2}}{|\boldsymbol{\Sigma}|}\left(-\Big(b_s(d+(J-2)e)-(J-1)ce\Big)f^{i\neq j}+(a_s-me)(cd-b_te)\right. \\
&\quad \left. + c\Big(b_tb_s+(J-2)b_sc-(J-1)c^2\Big)\right) \\
A_3^{i\neq j} &= \frac{me(d-e)^{J-2}}{|\boldsymbol{\Sigma}|}\Bigg((b_se-cd)f^{i\neq j}+(a_s-me)((d-e)b_t+(J-1)e(b_t-c)
\end{aligned}
$$

$$+\frac{c(cd-b_se)(b_s+(J-2)c)+(J-1)b_sce(b_s-c)-(J-2)b_t(b_s^2e-2b_sce+c^2d)}{d-e}-b_tb_s^2\Bigg)$$

$$A_4^{i\neq j}=\frac{me(d-e)^{J-2}}{|\boldsymbol{\Sigma}|}\Bigg((b_se-cd)f^{i\neq j}+(a_s-me)(cd-b_te)$$

$$+\frac{(cd-b_te)(b_sc-b_s^2-c^2)+(cd-b_se)b_tc+c^2e(b_s-b_t)}{d-e}\Bigg)$$

In each of the above factors, results under $t=s$ can be obtained simply by imposing that equality. In any case, since these factors are all functions only of model parameters, it follows that the full expectation is

$$\begin{aligned}\mathrm{E}[\epsilon_{it}\epsilon_{js}|\mathbf{X}]=A_1^{i\neq j}\mathrm{E}[\epsilon_{js}^2|\delta_s,\bar{Y}_j^B,\bar{Y}_i^B,\bar{Y}_{-i,-j}^B]+\Big[-A_1^{i\neq j}(\delta_s-\mu_\delta)+\delta_t-\mu_\delta+A_2^{i\neq j}\left(\bar{Y}_j^B-\mu_\delta\right)\\+A_3^{i\neq j}\left(\bar{Y}_i^B-\mu_\delta\right)+A_4^{i\neq j}\Big(\sum_{k\neq i,j}\left(\bar{Y}_k^B-\mu_\delta\right)\Big)\Big]\times\mathrm{E}[\epsilon_{js}|\delta_s,\bar{Y}_j^B,\bar{Y}_i^B,\bar{Y}_{-i,-j}^B]\end{aligned}\tag{A.12}$$

where $\mathrm{E}[\epsilon_{js}^2|\delta_s,\bar{Y}_j^B,\bar{Y}_i^B,\bar{Y}_{-i,-j}^B]=\mathrm{Var}\left(\epsilon_{js}|\delta_s,\bar{Y}_j^B,\bar{Y}_i^B,\bar{Y}_{-i,-j}^B\right)+\left(\mathrm{E}[\epsilon_{js}|\delta_s,\bar{Y}_j^B,\bar{Y}_i^B,\bar{Y}_{-i,-j}^B]\right)^2$, and the 'outer' expectation $\mathrm{E}[\epsilon_{js}|\delta_s,\bar{Y}_j^B,\bar{Y}_i^B,\bar{Y}_{-i,-j}^B]$ may also be calculated using formula (A.7). To do so, note first that the appropriate covariance matrix of conditioning variables, which has dimension $J+1$, is now

$$\hat{\boldsymbol{\Sigma}}_{\mathbf{yy}}=\begin{pmatrix}me&0&0&\cdots&0\\0&d&e&\cdots&e\\0&e&d&\cdots&e\\\vdots&\vdots&\vdots&\ddots&\vdots\\0&e&e&\cdots&d\end{pmatrix}$$

for which Lemma 5 implies inverse

$$\hat{\boldsymbol{\Sigma}}_{\mathbf{yy}}^{-1}=\frac{(d-e)^{J-2}}{|\hat{\boldsymbol{\Sigma}}_{\mathbf{yy}}|}\tag{A.13}$$

$$\times\begin{pmatrix}(d-e)(d+(J-1)e)&0&0&\cdots&0\\0&me(d+(J-2)e)&-me^2&\cdots&-me^2\\0&-me^2&me(d+(J-2)e)&\cdots&-me^2\\\vdots&\vdots&\vdots&\ddots&\vdots\\0&-me^2&-me^2&\cdots&me(d+(J-2)e)\end{pmatrix}\tag{A.14}$$

with $|\hat{\mathbf{\Sigma}}_{\mathbf{yy}}| = me(d-e)^{J-1}\,(d+(J-1)e)$. Noting that the corresponding covariance vector is

$$\hat{\mathbf{\Sigma}}_{\mathbf{xy}} = \begin{pmatrix} me & b_s & c & \cdots & c \end{pmatrix} \tag{A.15}$$

application of formula (A.7) now yields

$$\mathrm{E}[\epsilon_{js}|\delta_s, \bar{Y}_j^B, \bar{Y}_i^B, \bar{Y}_{-i,-j}^B] = \delta_s - \mu_\delta + B_1\left(\bar{Y}_j^B - \mu_\delta\right) + B_2 \sum_{k\neq j}\left(\bar{Y}_k^B - \mu_\delta\right) \tag{A.16}$$

where

$$\begin{aligned} B_1 &= \frac{me\,(d-e)^{J-2}}{|\hat{\mathbf{\Sigma}}_{\mathbf{yy}}|}\Big(b_s\,(d+(J-2)e) - (J-1)ce\Big) \\ B_2 &= \frac{me\,(d-e)^{J-2}}{|\hat{\mathbf{\Sigma}}_{\mathbf{yy}}|}\Big(cd - b_s e\Big) \end{aligned}$$

The fact that none of these quantities depend on t will soon prove useful. Next, to evaluate $\mathrm{E}[\epsilon_{js}^2|\delta_s, \bar{Y}_j^B, \bar{Y}_i^B, \bar{Y}_{-i,-j}^B]$ we will also need to calculate $\mathrm{Var}\left(\epsilon_{js}|\delta_s, \bar{Y}_j^B, \bar{Y}_i^B, \bar{Y}_{-i,-j}^B\right)$. Again, because all variables involved are normally distributed, this may be done by the following conditional-variance formula:

$$\mathrm{Var}(x|\mathbf{y}) = \sigma_x^2 - \mathbf{\Sigma}_{\mathbf{xy}}\mathbf{\Sigma}_{\mathbf{yy}}^{-1}\left(\mathbf{\Sigma}_{\mathbf{xy}}\right)' \tag{A.17}$$

where σ_x^2 is the unconditional variance of x and all other quantities are as defined in (A.7). Here, $\sigma_x^2 = a_s$; combining this fact with (A.14) and (A.15) in accordance with the above formula yields

$$\mathrm{Var}\left(\epsilon_{js}|\delta_s, \bar{Y}_j^B, \bar{Y}_i^B, \bar{Y}_{-i,-j}^B\right) = \frac{|\mathbf{\Sigma}|}{|\hat{\mathbf{\Sigma}}_{\mathbf{yy}}|} \tag{A.18}$$

which also does not depend on t. Finally, inserting (A.18) and (A.16) into (A.12) and collecting terms, we find that the full expectation is

$$\begin{aligned} \mathrm{E}[\epsilon_{it}\epsilon_{js}|\mathbf{X}] = A_1^{i\neq j}\frac{|\mathbf{\Sigma}|}{|\hat{\mathbf{\Sigma}}_{\mathbf{yy}}|} &+ (\delta_t - \mu_\delta)(\delta_s - \mu_\delta) + \left(A_1^{i\neq j}B_1 + A_2^{i\neq j}\right)(\delta_s - \mu_\delta)\left(\bar{Y}_j^B - \mu_\delta\right) \\ &+ \left(A_1^{i\neq j}B_2 + A_3^{i\neq j}\right)(\delta_s - \mu_\delta)\left(\bar{Y}_i^B - \mu_\delta\right) + \left(A_1^{i\neq j}B_2 + A_4^{i\neq j}\right)(\delta_s - \mu_\delta)\sum_{k\neq i,j}\left(\bar{Y}_k^B - \mu_\delta\right) \\ &+ B_1(\delta_t - \mu_\delta)\left(\bar{Y}_j^B - \mu_\delta\right) + B_2(\delta_t - \mu_\delta)\left(\bar{Y}_i^B - \mu_\delta\right) + B_2(\delta_t - \mu_\delta)\sum_{k\neq i,j}\left(\bar{Y}_k^B - \mu_\delta\right) \\ &+ B_1\left(A_1^{i\neq j}B_1 + A_2^{i\neq j}\right)\left(\bar{Y}_j^B - \mu_\delta\right)^2 + B_2\left(A_1^{i\neq j}B_2 + A_3^{i\neq j}\right)\left(\bar{Y}_i^B - \mu_\delta\right)^2 \end{aligned}$$

$$+ B_2\left(A_1^{i\neq j}B_2 + A_4^{i\neq j}\right)\left(\sum_{k\neq i,j}\left(\bar{Y}_k^B - \mu_\delta\right)\right)^2$$
$$+ \left(B_1\left(A_1^{i\neq j}B_2 + A_3^{i\neq j}\right) + B_2\left(A_1^{i\neq j}B_1 + A_2^{i\neq j}\right)\right)\left(\bar{Y}_j^B - \mu_\delta\right)\left(\bar{Y}_i^B - \mu_\delta\right)$$
$$+ \left(B_1\left(A_1^{i\neq j}B_2 + A_4^{i\neq j}\right) + B_2\left(A_1^{i\neq j}B_1 + A_2^{i\neq j}\right)\right)\left(\bar{Y}_j^B - \mu_\delta\right)\sum_{k\neq i,j}\left(\bar{Y}_k^B - \mu_\delta\right)$$
$$+ B_2\left(A_1^{i\neq j}B_2 + A_3^{i\neq j} + A_1^{i\neq j}B_2 + A_4^{i\neq j}\right)\left(\bar{Y}_i^B - \mu_\delta\right)\sum_{k\neq i,j}\left(\bar{Y}_k^B - \mu_\delta\right)$$

which may feasibly be summed across all t and s. Recall that certain terms and factors depend only on one of the time periods; in particular, since $|\mathbf{\Sigma}|$ depends only on s, only the numerator of each A factor includes t. The summed expectation can therefore be written as

$$\sum_{t=1}^{r}\sum_{s=1}^{r}\mathrm{E}[\epsilon_{it}\epsilon_{js}|\mathbf{X}]$$
$$= \sum_{s=1}^{r}\left[\frac{|\mathbf{\Sigma}|}{|\hat{\mathbf{\Sigma}}_{\mathbf{yy}}|}\sum_{t=1}^{r}A_1^{i\neq j} + \left(B_1\sum_{t=1}^{r}A_1^{i\neq j} + \sum_{t=1}^{r}A_2^{i\neq j}\right)\left(\delta_s - \mu_\delta\right)\left(\bar{Y}_j^B - \mu_\delta\right)\right.$$
$$+ \left(B_2\sum_{t=1}^{r}A_1^{i\neq j} + \sum_{t=1}^{r}A_3^{i\neq j}\right)\left(\delta_s - \mu_\delta\right)\left(\bar{Y}_i^B - \mu_\delta\right)$$
$$+ \left(B_2\sum_{t=1}^{r}A_1^{i\neq j} + \sum_{t=1}^{r}A_4^{i\neq j}\right)\left(\delta_s - \mu_\delta\right)\sum_{k\neq i,j}\left(\bar{Y}_k^B - \mu_\delta\right)$$
$$+ B_1\left(B_1\sum_{t=1}^{r}A_1^{i\neq j} + \sum_{t=1}^{r}A_2^{i\neq j}\right)\left(\bar{Y}_j^B - \mu_\delta\right)^2 + B_2\left(B_2\sum_{t=1}^{r}A_1^{i\neq j} + \sum_{t=1}^{r}A_3^{i\neq j}\right)\left(\bar{Y}_i^B - \mu_\delta\right)^2$$
$$+ B_2\left(B_2\sum_{t=1}^{r}A_1^{i\neq j} + \sum_{t=1}^{r}A_4^{i\neq j}\right)\left(\sum_{k\neq i,j}\left(\bar{Y}_k^B - \mu_\delta\right)\right)^2$$
$$+ \left(B_1\left(B_2\sum_{t=1}^{r}A_1^{i\neq j} + \sum_{t=1}^{r}A_3^{i\neq j}\right) + B_2\left(B_1\sum_{t=1}^{r}A_1^{i\neq j} + \sum_{t=1}^{r}A_2^{i\neq j}\right)\right)\left(\bar{Y}_j^B - \mu_\delta\right)\left(\bar{Y}_i^B - \mu_\delta\right)$$
$$+ \left(B_1\left(B_2\sum_{t=1}^{r}A_1^{i\neq j} + \sum_{t=1}^{r}A_4^{i\neq j}\right) + B_2\left(B_1\sum_{t=1}^{r}A_1^{i\neq j} + \sum_{t=1}^{r}A_2^{i\neq j}\right)\right)\left(\bar{Y}_j^B - \mu_\delta\right)\sum_{k\neq i,j}\left(\bar{Y}_k^B - \mu_\delta\right)$$
$$\left.+ B_2\left(B_2\sum_{t=1}^{r}A_1^{i\neq j} + \sum_{t=1}^{r}A_3^{i\neq j} + B_2\sum_{t=1}^{r}A_1^{i\neq j} + \sum_{t=1}^{r}A_4^{i\neq j}\right)\left(\bar{Y}_i^B - \mu_\delta\right)\sum_{k\neq i,j}\left(\bar{Y}_k^B - \mu_\delta\right)\right]$$
$$+ \sum_{t=1}^{r}\left[\left(\delta_t - \mu_\delta\right)\left(\bar{Y}_j^B - \mu_\delta\right)\sum_{s=1}^{r}B_1 + \left(\delta_t - \mu_\delta\right)\left(\bar{Y}_i^B - \mu_\delta\right)\sum_{s=1}^{r}B_2\right.$$

$$+ \left(\delta_t - \mu_\delta\right) \sum_{k \neq i,j} \left(\bar{Y}_k^B - \mu_\delta\right) \sum_{s=1}^{r} B_2 \Bigg] + \sum_{t=1}^{r} \sum_{s=1}^{r} \Bigg[\left(\delta_t - \mu_\delta\right) \left(\delta_s - \mu_\delta\right) \Bigg]$$

The first term of this expression is

$$\sum_{s=1}^{r} \left(\frac{|\boldsymbol{\Sigma}|}{|\hat{\boldsymbol{\Sigma}}_{\mathbf{yy}}|} \sum_{t} A_1^{i \neq j} \right) = \sum_{s=1}^{r} r \left(f^{i \neq j} - \frac{b_s c + (J-1)\, c^2}{d + (J-1)e} \right) = r^2 \left(f^{i \neq j} - \frac{J c^2}{d + (J-1)e} \right)$$

where the equalities use $\sum_t b_t = rc$ and $\sum_s b_s = rc$, respectively. Similarly,

$$B_1 \sum_{t=1}^{r} A_1^{i \neq j} + \sum_{t=1}^{r} A_2^{i \neq j} \quad = B_2 \sum_{t=1}^{r} A_1^{i \neq j} + \sum_{t=1}^{r} A_3^{i \neq j} = B_2 \sum_{t=1}^{r} A_1^{i \neq j} + \sum_{t=1}^{r} A_4^{i \neq j} = \sum_{s=1}^{r} B_1 = \sum_{s=1}^{r} B_2$$
$$= \frac{rc}{d + (J-1)e}$$

implying that the summed expectation reduces to

$$\begin{aligned}
\sum_{t=1}^{r} \sum_{s=1}^{r} \mathrm{E}[\epsilon_{it}\epsilon_{js}|\mathbf{X}] = r^2 \left(f^{i \neq j} - \frac{J c^2}{d + (J-1)e} \right) &+ \sum_{t=1}^{r} \sum_{s=1}^{r} \Big(\left(\delta_t - \mu_\delta\right) \left(\delta_s - \mu_\delta\right) \Big) \\
&+ \frac{rc}{d + (J-1)e} \sum_{s=1}^{r} \left(\delta_s - \mu_\delta\right) \sum_{k=1}^{J} \left(\bar{Y}_k^B - \mu_\delta\right) \\
&+ \frac{rc}{d + (J-1)e} \sum_{t=1}^{r} \left(\delta_t - \mu_\delta\right) \sum_{k=1}^{J} \left(\bar{Y}_k^B - \mu_\delta\right) \\
&+ \frac{rc}{d + (J-1)e} \sum_{s=1}^{r} B_1 \left(\sum_{k=1}^{J} \left(\bar{Y}_k^B - \mu_\delta\right)^2 + \sum_{k=1}^{J} \sum_{l \neq k} \left(\bar{Y}_k^B - \mu_\delta\right) \left(\bar{Y}_l^B - \mu_\delta\right) \right) \\
&= r^2 \left(f^{i \neq j} - \frac{J^2 c^2}{d + (J-1)e} \right) + \left(\sum_{p=1}^{r} \left(\delta_p - \mu_\delta\right) + \frac{rc}{d + (J-1)e} \sum_{k=1}^{J} \left(\bar{Y}_k^B - \mu_\delta\right) \right)^2
\end{aligned}$$

which is seen to be constant across any i, j with $i \neq j$.

Having calculated the sum of conditional means for the case of different experimental units, we now move on to consider the $i = j$ case, for which

$$\mathrm{E}[\epsilon_{it}\epsilon_{is}|\mathbf{X}] = \mathrm{E}[\epsilon_{is}\mathrm{E}[\epsilon_{it}|\epsilon_{is}, \delta_s, \delta_t, \bar{Y}_i^B, \bar{Y}_{-i}^B]|\delta_s, \bar{Y}_i^B, \bar{Y}_{-i}^B]$$

where, similarly to before, $\bar{Y}_{-i}^B$ is the set of all baseline averages belonging to units other than i. It is straightforward to confirm that, both when $t \neq s$ and $t = s$, the variance-covariance matrices $\boldsymbol{\Sigma}_{\mathbf{yy}}$ of the inner (ϵ_{it}) expectation are identical to their counterparts when $i \neq j$.

It follows, of course, that the corresponding inverse matrices are also identical. By contrast, the covariance vectors are now different from before; when $t \neq s$, we have

$$\mathbf{\Sigma}_{\mathbf{xy}}^{t\neq s} = \begin{pmatrix} f_{ts}^{i=j} & 0 & me & b_t & c & \cdots & c \end{pmatrix}$$

where

$$\begin{aligned} f_{ts}^{i=j} &= \text{Cov}\left(\epsilon_{it}, \epsilon_{is}\right) - \text{Cov}\left(\delta_t, \delta_s\right) \\ &= \left(1-\theta\right)^2 \sigma_v^2 + \frac{\theta^2}{m}\left(\sigma_\omega^2 + \sigma_\delta^2\right) + \frac{\theta^2\left(m-1\right)}{m}\psi^B + \text{Cov}\left(\omega_{it}, \omega_{is}\right) - \theta\text{Cov}\left(\omega_{it}, \bar{\omega}_i^B\right) \\ &- \theta\text{Cov}\left(\omega_{is}, \bar{\omega}_i^B\right) \end{aligned}$$

which we may also note implies

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

In any case, when $t = s$ we have

$$\mathbf{\Sigma}_{\mathbf{xy}}^{t=s} = \begin{pmatrix} f_{ts}^{i=j} + me & me & b_s & c & \cdots & c \end{pmatrix}$$

Applying formula (A.7) to calculate the inner expectation, it is again the case that the $t \neq s$ and $t = s$ cases may be collected into a single expression, namely

$$\begin{aligned} \text{E}[\epsilon_{it} | \epsilon_{is}, \delta_s, \delta_t, \bar{Y}_i^B, \bar{Y}_{-i}^B] = & A_1^{i=j}\epsilon_{is} - A_1^{i=j}\left(\delta_s - \mu_\delta\right) + \delta_t - \mu_\delta \\ & + A_2^{i=j}\left(\bar{Y}_i^B - \mu_\delta\right) + A_3^{i=j}\left(\sum_{k\neq i}\left(\bar{Y}_k^B - \mu_\delta\right)\right) \end{aligned}$$

where

$$\begin{aligned} A_1^{i=j} = & \frac{me(d-e)^{J_2}}{|\mathbf{\Sigma}|}\Bigg(\left(d-e\right)\left(d + (J-1)e\right) f_{ts}^{i=j} - b_t\Big(b_s\left(d + (J-2)e\right) - (J-1)ce\Big) \\ & + (J-1)c\left(b_s e - cd\right)\Bigg) \end{aligned}$$

$$A_2^{i=j} = \frac{me(d-e)^{J_2}}{|\mathbf{\Sigma}|}\Bigg(-\Big(b_s\left(d+(J-2)e\right)-(J-1)ce\Big) f_{ts}^{i=j}$$

$$+ b_t\Big((a_s - me)(d+(J-2)e)-(J-1)c^2\Big) + (J-1)c\left(-e(a_s-me)+b_s c\right)\Bigg)$$

$$A_3^{i=j} = \frac{me(d-e)^{J_2}}{|\mathbf{\Sigma}|}\Bigg((b_s e - cd) f_{ts}^{i=j} + (a_s - me)(cd - b_t e) + b_s c(b_t - b_s)\Bigg)$$

As before, only the numerator of each factor depends on t.

As for the outer (ϵ_{is}) expectation, both covariance matrix $\hat{\mathbf{\Sigma}}_{\mathbf{yy}}$ and covariance vector $\hat{\mathbf{\Sigma}}_{\mathbf{xy}}$ are identical to those of the $i \neq j$ case. It follows that $\mathrm{E}[\epsilon_{is}|d_s, \bar{Y}_i^B, \bar{Y}_{-i}^B]$ as well as $\mathrm{Var}[\epsilon_{is}|d_s, \bar{Y}_i^B, \bar{Y}_{-i}^B]$ are again equal to (A.16) and (A.18). Combining this fact with the above expressions, then, the summed full expectation when $i = j$ is

$$\begin{aligned}
&\sum_{t=1}^{r}\sum_{s=1}^{r}\mathrm{E}[\epsilon_{it}\epsilon_{is}|\mathbf{X}] \\
&= \sum_{s=1}^{r}\Bigg[\frac{|\mathbf{\Sigma}|}{|\hat{\mathbf{\Sigma}}_{\mathbf{yy}}|}\sum_{t=1}^{r}A_1^{i=j} + \left(B_1\sum_{t=1}^{r}A_1^{i=j} + \sum_{t=1}^{r}A_2^{i=j}\right)(\delta_s - \mu_\delta)\left(\bar{Y}_i^B - \mu_\delta\right) \\
&+ \left(B_2\sum_{t=1}^{r}A_1^{i=j} + \sum_{t=1}^{r}A_3^{i=j}\right)(\delta_s - \mu_\delta)\sum_{k\neq i}\left(\bar{Y}_k^B - \mu_\delta\right) \\
&+ B_1\left(B_1\sum_{t=1}^{r}A_1^{i=j} + \sum_{t=1}^{r}A_2^{i=j}\right)\left(\bar{Y}_i^B - \mu_\delta\right)^2 + B_2\left(B_2\sum_{t=1}^{r}A_1^{i=j} + \sum_{t=1}^{r}A_3^{i=j}\right)\left(\sum_{k\neq i}\left(\bar{Y}_k^B - \mu_\delta\right)\right)^2 \\
&+ \left(B_1\left(B_2\sum_{t=1}^{r}A_1^{i=j} + \sum_{t=1}^{r}A_3^{i=j}\right) + B_2\left(B_1\sum_{t=1}^{r}A_1^{i=j} + \sum_{t=1}^{r}A_2^{i=j}\right)\right)\left(\bar{Y}_i^B - \mu_\delta\right)\sum_{k\neq i}\left(\bar{Y}_k^B - \mu_\delta\right)\Bigg] \\
&+ \sum_{t=1}^{r}\left[(\delta_t - \mu_\delta)\left(\bar{Y}_i^B - \mu_\delta\right)\sum_{s=1}^{r}B_1 + (\delta_t - \mu_\delta)\sum_{k\neq i}\left(\bar{Y}_k^B - \mu_\delta\right)\sum_{s=1}^{r}B_2\right] \\
&+ \sum_{t=1}^{r}\sum_{s=1}^{r}\Big[(\delta_t - \mu_\delta)(\delta_s - \mu_\delta)\Big]
\end{aligned} \tag{A.19}$$

but

$$\sum_{s=1}^{r}\left(\frac{|\mathbf{\Sigma}|}{|\hat{\mathbf{\Sigma}}_{\mathbf{yy}}|}\sum_{t=1}^{r}A_1^{i=j}\right) = \sum_{s=1}^{r}\left(r\bar{f}_s^{i=j} - \frac{r\left(b_s c + (J-1)\,c^2\right)}{d+(J-1)e}\right) = r^2\left(\bar{f}^{i=j} - \frac{Jc^2}{d+(J-1)e}\right)$$

where

$$\bar{f}^{i=j} = \frac{1}{r}\sum_{s=1}^{r} \bar{f}_s^{i=j} = \left((1-\theta)^2 \sigma_v^2 + \left(\frac{\theta^2}{m} + \frac{1}{r}\right)\sigma_\omega^2 + \frac{\theta^2\sigma_\delta^2}{m} + \frac{\theta^2(m-1)}{m}\psi^B + \frac{r-1}{r}\psi^A - 2\theta\psi^X \right)$$

and furthermore

$$B_1 \sum_{t=1}^{r} A_1^{i=j} + \sum_{t=1}^{r} A_2^{i=j} = B_2 \sum_{t=1}^{r} A_1^{i=j} + \sum_{t=1}^{r} A_3^{i=j} = \frac{rc}{d+(J-1)e}$$

Inserting these expressions into (A.19) yields the summed expectation as

$$\sum_{t=1}^{r}\sum_{s=1}^{r} \mathrm{E}[\epsilon_{it}\epsilon_{is}|\mathbf{X}] = r^2\left(\bar{f}^{i=j} - \frac{Jc^2}{d+(J-1)e}\right) + \left(\sum_{p=1}^{r}(\delta_p - \mu_\delta) + \frac{rc}{d+(J-1)e}\sum_{k=1}^{J}\left(\bar{Y}_k^B - \mu_\delta\right)\right)^2$$

This differs from the summed expectation we found when $i \neq j$ only by $r^2\left(\bar{f}^{i=j} - f^{i\neq j}\right)$. Clearly, everything but this difference is constant across all i, j and will thus be summed across both $i \neq j$ and $i = j$. Referring back to equation (A.6), the constant will therefore be multiplied with

$$\begin{aligned}
&\frac{1}{P^2}\sum_{i=1}^{PJ}\sum_{j=1}^{PJ}\left[\Big(Z - PJ\left(\bar{Y}_T^B - \bar{Y}_C^B\right)\left(\bar{Y}_i^B - \bar{Y}_T^B\right)\Big)\Big(Z - PJ\left(\bar{Y}_T^B - \bar{Y}_C^B\right)\left(\bar{Y}_j^B - \bar{Y}_T^B\right)\Big)\right] \\
&+ \frac{2}{P(1-P)}\sum_{i=1}^{PJ}\sum_{j=PJ+1}^{J}\left[\Big(Z - PJ\left(\bar{Y}_T^B - \bar{Y}_C^B\right)\left(\bar{Y}_i^B - \bar{Y}_T^B\right)\Big)\right. \\
&\left.\times\Big(-Z - (1-P)J\left(\bar{Y}_T^B - \bar{Y}_C^B\right)\left(\bar{Y}_j^B - \bar{Y}_C^B\right)\Big)\right] \\
&+ \frac{1}{(1-P)^2}\sum_{i=PJ+1}^{J}\sum_{j=PJ+1}^{J}\left[\Big(-Z - (1-P)J\left(\bar{Y}_T^B - \bar{Y}_C^B\right)\left(\bar{Y}_i^B - \bar{Y}_C^B\right)\Big)\right. \\
&\left.\times\Big(-Z - (1-P)J\left(\bar{Y}_T^B - \bar{Y}_C^B\right)\left(\bar{Y}_j^B - \bar{Y}_C^B\right)\Big)\right] \\
&= \frac{1}{P^2}\sum_{i=1}^{PJ}\left[PJZ\Big(Z - PJ\left(\bar{Y}_T^B - \bar{Y}_C^B\right)\left(\bar{Y}_i^B - \bar{Y}_T^B\right)\Big)\right] \\
&+ \frac{2}{P(1-P)}\sum_{i=1}^{PJ}\left[-(1-P)JZ\Big(Z - PJ\left(\bar{Y}_T^B - \bar{Y}_C^B\right)\left(\bar{Y}_i^B - \bar{Y}_T^B\right)\Big)\right] \\
&+ \frac{1}{(1-P)^2}\sum_{i=PJ+1}^{J}\left[-(1-P)JZ\Big(-Z - (1-P)J\left(\bar{Y}_T^B - \bar{Y}_C^B\right)\left(\bar{Y}_i^B - \bar{Y}_C^B\right)\Big)\right]
\end{aligned}$$

$$= 0$$

As a result, we are left only with

$$\text{Var}\left(\hat{\tau}|\mathbf{X}\right) = \frac{\bar{f}^{i=j} - f^{i\neq j}}{J^2 Z^2} \left\{ \frac{1}{P^2} \sum_{i=1}^{PJ} \Big(Z - PJ\left(\bar{Y}_T^B - \bar{Y}_C^B\right)\left(\bar{Y}_i^B - \bar{Y}_T^B\right) \Big)^2 \right.$$
$$\left. + \frac{1}{(1-P)^2} \sum_{i=PJ+1}^{J} \Big(-Z - (1-P)J\left(\bar{Y}_T^B - \bar{Y}_C^B\right)\left(\bar{Y}_i^B - \bar{Y}_C^T\right) \Big)^2 \right\} \tag{A.20}$$

but

$$\bar{f}^{i=j} - f^{i\neq j} = (1-\theta)^2 \sigma_v^2 + \left(\frac{\theta^2}{m} + \frac{1}{r}\right)\sigma_\omega^2 + \frac{\theta^2(m-1)}{m}\psi^B + \frac{r-1}{r}\psi^A - 2\theta\psi^X$$

implying that (A.20) is exactly equal to equation A61 in Burlig et al. (2020) and may likewise be simplified to yield

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

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

A.2 Time shocks not included in ANCOVA regression

Now, consider instead the ANCOVA regression model

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

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 analyzed in Burlig et al. (2020); although, of course, the assumed DGP (A.1) is not. Again, we will calculate the ANCOVA variance by sandwich formula (A.3). However, 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.6) above.

The next task is to evaluate conditional means $\text{E}[\epsilon_{it}\epsilon_{js}|\mathbf{X}]$, as in the previous section. With time FE no longer included in the regression, we may write any such quantity for which $i \neq j$ as

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

These means can again be calculated using formula (A.7). Here, the covariance matrix associated with the inner (ϵ_{it}) expectation is

$$\mathbf{\Sigma_{yy}} = \begin{pmatrix} a_s & b_s & c & \cdots & c \\ b_s & d & e & \cdots & e \\ c & e & d & \cdots & e \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c & e & e & \cdots & d \end{pmatrix} \tag{A.21}$$

where all parameters are defined as in section A.1 above. Lemma 5 may again be used to calculate corresponding inverse

$$\mathbf{\Sigma_{yy}^{-1}} = \frac{(d-e)^{J-2}}{|\mathbf{\Sigma_{yy}}|} \left(\begin{matrix} (d-e)(d+(J-1)e) & -b_s(d+(J-2)e)+(J-1)ce \\ -b_s(d+(J-2)e)+(J-1)ce & a_s(d+(J-2)e)-(J-1)c^2 \\ b_se-cd & b_sc-a_se \\ \vdots & \vdots \\ b_se-cd & b_sc-a_se \end{matrix} \right.$$

$$\left. \begin{matrix} b_se-cd & \cdots & b_se-cd \\ b_sc-a_se & \cdots & b_sc-a_se \\ a_s(d+(J-2)e)-b_s^2-(J-2)c^2-\frac{(J-2)e(b_s-c)^2}{d-e} & \cdots & -a_se+\frac{b_se(b_s-c)-c(b_se-cd)}{d-e} \\ -a_se+\frac{b_se(b_s-c)-c(b_se-cd)}{d-e} & \cdots & a_s(d+(J-2)e)-b_s^2-(J-2)c^2-\frac{(J-2)e(b_s-c)^2}{d-e} \end{matrix} \right) \tag{A.22}$$

with

$$|\mathbf{\Sigma_{yy}}| = (d-e)^{J-2}\Big[(d-e)\big(a_s(d+(J-1)e)-b_s^2-(J-1)c^2\big)-(J-1)e(b_s-c)^2\Big]$$

We also note that

$$\mathbf{\Sigma_{xy}} = \begin{pmatrix} g_{ts}^{i\neq j} & c & b_t & c & \cdots & c \end{pmatrix} \tag{A.23}$$

where

$$g_{ts}^{i\neq j} = \mathrm{Cov}(\epsilon_{it}, \epsilon_{js}) = \frac{\theta^2\sigma_\delta^2}{m} + \mathrm{Cov}(\delta_t, \delta_s)$$

implying

$$\sum_{t=1}^{r} g_{ts}^{i\neq j} = r\left(\frac{\theta^2}{m} + \frac{1}{r}\right)\sigma_\delta^2 \equiv r\bar{g}^{i\neq j}$$

Combining (A.22) and (A.23) in accordance with (A.7) yields the full expectation as

$$\begin{aligned}
\mathrm{E}[\epsilon_{it}\epsilon_{js}|\mathbf{X}] = {} & A_1^{i\neq j}\mathrm{E}[\epsilon_{js}^2|\bar{Y}_j^B, \bar{Y}_i^B, \bar{Y}_{-i,-j}] \\
& + \left(A_2^{i\neq j}\left(\bar{Y}_j^B - \mu_\delta\right) + A_3^{i\neq j}\left(\bar{Y}_i^B - \mu_\delta\right) + A_4^{i\neq j}\sum_{k\neq i,j}\left(\bar{Y}_k^B - \mu_\delta\right)\right) \\
& \times \mathrm{E}[\epsilon_{js}|\bar{Y}_j^B, \bar{Y}_i^B, \bar{Y}_{-i,-j}]
\end{aligned} \tag{A.24}$$

where

$$\begin{aligned}
A_1^{i\neq j} &= \frac{(d-e)^{J-2}}{|\mathbf{\Sigma_{yy}}|}\left((d-e)\left(g_{ts}^{i\neq j}(d+(J-1)e) - (J-2)c^2\right) + e(b_s b_t + c^2) - (b_s + b_t)cd\right) \\
A_2^{i\neq j} &= \frac{(d-e)^{J-2}}{|\mathbf{\Sigma_{yy}}|}\left(g_{ts}^{i\neq j}\left(-b_s(d+(J-2)e) + (J-1)ce\right) + a_s(cd - b_t e)\right. \\
&\quad \left. - (J-1)c^3 + b_t b_s c + (J-2)b_s c^2\right) \\
A_3^{i\neq j} &= \frac{(d-e)^{J-2}}{|\mathbf{\Sigma_{yy}}|}\left(g_{ts}^{i\neq j}(b_s e - cd) + a_s\left(b_t(d+(J-2)e) - (J-1)ce\right) + b_s c^2\right. \\
&\quad \left. - b_t b_s^2 - (J-2)b_t c^2 + \frac{J-2}{d-e}\left(c\left(b_s^2 e - 2b_s ce + c^2 d\right) - b_t e(b_s - c)^2\right)\right) \\
A_4^{i\neq j} &= \frac{(d-e)^{J-2}}{|\mathbf{\Sigma_{yy}}|}\left(g_{ts}^{i\neq j}(b_s e - cd) + a_s(cd - b_t e) + b_s c^2 - b_s^2 c - (J-2)c^3\right. \\
&\quad \left. + \frac{(b_t + (J-3)c)\left(b_s^2 e - 2b_s ce + c^2 d\right) - (J-2)ce(b_s - c)^2}{d-e}\right)
\end{aligned}$$

To calculate the two remaining conditional expectations in (A.24), note first that the corresponding covariance matrix of conditioning variables is now simply

$$\hat{\mathbf{\Sigma}}_{\mathbf{yy}} = \begin{pmatrix} d & e & \cdots & e \\ e & d & \cdots & e \\ \vdots & \vdots & \ddots & \vdots \\ e & e & \cdots & d \end{pmatrix}$$

for which Lemma 5 implies inverse

$$\hat{\mathbf{\Sigma}}_{\mathbf{yy}}^{-1} = \frac{(d-e)^{J-2}}{|\hat{\mathbf{\Sigma}}_{\mathbf{yy}}|} \begin{pmatrix} d+(J-2)e & -e & \cdots & -e \\ -e & d+(J-2)e & \cdots & -e \\ \vdots & \vdots & \ddots & \vdots \\ -e & -e & \cdots & d+(J-2)e \end{pmatrix} \tag{A.25}$$

with $|\hat{\mathbf{\Sigma}}_{\mathbf{yy}}| = (d-e)^{J-1}(d+(J-1)e)$. The appropriate covariance vector for ϵ_{js} is

$$\hat{\mathbf{\Sigma}}_{\mathbf{xy}} = \begin{pmatrix} b_s & c & \cdots & c \end{pmatrix}$$

so applying formula (A.7) yields

$$\mathrm{E}[\epsilon_{js}|\bar{Y}_j^B, \bar{Y}_i^B, \bar{Y}_{-i,-j}^B] = B_1\left(\bar{Y}_j^B - \mu_\delta\right) + B_2 \sum_{k\neq j}\left(\bar{Y}_k^B - \mu_\delta\right)$$

where

$$\begin{aligned} B_1 &= \frac{(d-e)^{J-2}}{|\hat{\mathbf{\Sigma}}_{\mathbf{yy}}|}\Big(b_s(d+(J-2)e) - (J-1)ce\Big) \\ B_2 &= \frac{(d-e)^{J-2}}{|\hat{\mathbf{\Sigma}}_{\mathbf{yy}}|}\Big(cd - b_s e\Big) \end{aligned}$$

Also, application of formula (A.17) yields

$$\mathrm{Var}(\epsilon_{js}|\bar{Y}_j^B, \bar{Y}_i^B, \bar{Y}_{-i,-j}^B) = \frac{|\mathbf{\Sigma}_{\mathbf{yy}}|}{|\hat{\mathbf{\Sigma}}_{\mathbf{yy}}|}$$

Thus, after combining with (A.24) and collecting terms, we find that the full expectation term is

$$\begin{aligned} \mathrm{E}[\epsilon_{it}\epsilon_{js}|\mathbf{X}] = A_1^{i\neq j}\frac{|\mathbf{\Sigma}_{\mathbf{yy}}|}{|\hat{\mathbf{\Sigma}}_{\mathbf{yy}}|} &+ B_1(A_1^{i\neq j}B_1 + A_2^{i\neq j})\left(\bar{Y}_j^B - \mu_\delta\right)^2 + B_2(A_1^{i\neq j}B_2 + A_3^{i\neq j})\left(\bar{Y}_i^B - \mu_\delta\right)^2 \\ &+ B_2(A_1^{i\neq j}B_2 + A_3^{i\neq j})\left(\sum_{k\neq i,j}\left(\bar{Y}_k^B - \mu_\delta\right)\right)^2 \\ &+ \Big(B_1(A_1^{i\neq j}B_2 + A_3^{i\neq j}) + B_2(A_1^{i\neq j}B_1 + A_2^{i\neq j})\Big)\left(\bar{Y}_j^B - \mu_\delta\right)\left(\bar{Y}_i^B - \mu_\delta\right) \\ &+ \Big(B_1(A_1^{i\neq j}B_2 + A_4^{i\neq j}) + B_2(A_1^{i\neq j}B_1 + A_2^{i\neq j})\Big)\left(\bar{Y}_j^B - \mu_\delta\right)\sum_{k\neq i,j}\left(\bar{Y}_k^B - \mu_\delta\right) \end{aligned}$$

$$B_2\left(A_1^{i\neq j}B_2+A_3^{i\neq j}+A_1^{i\neq j}B_2+A_4^{i\neq j}\right)\left(\bar{Y}_i-\mu_\delta\right)\sum_{k\neq i,j}\left(\bar{Y}_k^B-\mu_\delta\right) \tag{A.26}$$

Since

$$\sum_{t=1}^{r}\sum_{s=1}^{r}\left(A_1^{i\neq j}\frac{|\mathbf{\Sigma_{yy}}|}{|\hat{\mathbf{\Sigma}}_{\mathbf{yy}}|}\right)=\sum_{s=1}^{r}r\left(\bar{g}^{i\neq j}-\frac{b_s c+(J-1)c^2}{d+(J-1)e}\right)=r^2\left(\bar{g}^{i\neq j}-\frac{Jc^2}{d+(J-1)e}\right)$$

and furthermore

$$B_1\sum_{t=1}^{r}A_1^{i\neq j}+\sum_{t=1}^{r}A_2^{i\neq j}=B_2\sum_{t=1}^{r}A_1^{i\neq j}+\sum_{t=1}^{r}A_3^{i\neq j}=B_2\sum_{t=1}^{r}A_1^{i\neq j}+\sum_{t=1}^{r}A_4^{i\neq j}=\frac{rc}{d+(J-1)e}$$

it follows that summing (A.26), first across t and then across s, produces

$$\sum_{t=1}^{r}\sum_{s=1}^{r}\mathrm{E}[\epsilon_{it}\epsilon_{js}|\mathbf{X}]=r^2\left[\bar{g}^{i\neq j}-\frac{Jc^2}{d+(J-1)e}+\left(\frac{c}{d+(J-1)e}\sum_{k=1}^{J}\left(\bar{Y}_k^B-\mu_\delta\right)\right)^2\right]$$

which we note is invariant across any i,j with $i\neq j$.

Moving on to the $i=j$ case, conditional expectations are now

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

where the covariance matrix $\mathbf{\Sigma_{yy}}$ relevant to the inner expectation is again (A.21). As for $\mathbf{\Sigma_{xy}}$, it is now

$$\mathbf{\Sigma_{xy}}=\begin{pmatrix} g_{ts}^{i=j} & b_t & c & \cdots & c\end{pmatrix} \tag{A.27}$$

where

$$\begin{aligned}
g_{ts}^{i=j}&=\mathrm{Cov}(\epsilon_{it},\epsilon_{is})\\
&=(1-\theta)^2\sigma_v^2+\frac{\theta^2}{m}\left(\sigma_\omega^2+\sigma_\delta^2\right)+\mathrm{Cov}(\delta_t,\delta_s)+\frac{\theta^2\,(m-1)}{m}\psi^B+\mathrm{Cov}\left(\omega_{it},\omega_{is}\right)\\
&-\theta\mathrm{Cov}\left(\omega_{it},\bar{\omega}_i^B\right)-\theta\mathrm{Cov}\left(\omega_{is},\bar{\omega}_i^B\right)
\end{aligned}$$

with

$$\sum_{t=1}^{r}g_{ts}^{i=j}=r\Bigg((1-\theta)^2\sigma_v^2+\left(\frac{\theta^2}{m}+\frac{1}{r}\right)\left(\sigma_\omega^2+\sigma_\delta^2\right)+\frac{\theta^2(m-1)}{m}\psi^B+\frac{\sum_{p\neq s}\mathrm{Cov}\left(\omega_{ip},\omega_{is}\right)}{r}$$

$$
\begin{aligned}
&- \theta\psi^X - \theta \mathrm{Cov}\left(\omega_{is}, \bar{\omega}_i^B\right) \Bigg) \\
&\equiv r\bar{g}_s^{i=j}
\end{aligned}
$$

Combining (A.10) with (A.21) in formula (A.7) now yields

$$
\mathrm{E}[\epsilon_{it}\epsilon_{is}|\mathbf{X}] = A_1^{i=j}\mathrm{E}[\epsilon_{is}^2|\bar{Y}_i^B, \bar{Y}_{-i}^B] + \left(A_2^{i=j}\left(\bar{Y}_i^B - \mu_\delta\right) + A_3^{i=j}\sum_{k\neq i}\left(\bar{Y}_k^B - \mu_\delta\right)\right)\mathrm{E}[\epsilon_{is}|\bar{Y}_i^B, \bar{Y}_{-i}^B]
$$

where

$$
\begin{aligned}
A_1^{i=j} &= \frac{(d-e)^{J-2}}{|\mathbf{\Sigma_{yy}}|}\Bigg(g_{ts}^{i=j}(d-e)(d+(J-1)e) - b_t b_s d - (J-2)b_s b_t e \\
&\quad + (J-1)c\Big((b_t+b_s)e - cd\Big)\Bigg) \\
A_2^{i=j} &= \frac{(d-e)^{J-2}}{|\mathbf{\Sigma_{yy}}|}\Bigg(g_{ts}^{i=j}\Big(- b_s(d+(J-2)e) + (J-1)ce\Big) + a_s\Big(b_t(d+(J-2)e) - (J-1)ce\Big) \\
&\quad + (J-1)c^2(b_s - b_t)\Bigg) \\
A_3^{i=j} &= \frac{(d-e)^{J-2}}{|\mathbf{\Sigma_{yy}}|}\Bigg(g_{ts}^{i=j}(b_s e - cd) + a_s(cd - b_t e) + b_t b_s c - b_s^2 c\Bigg)
\end{aligned}
$$

It is also simple to check that $\hat{\mathbf{\Sigma}}_{\mathbf{yy}}$, and $\hat{\mathbf{\Sigma}}_{\mathbf{xy}}$, corresponding to the conditional mean of ϵ_{is}, are both unchanged compared to the $i \neq j$ case. It follows that the full expectation term is

$$
\begin{aligned}
\mathrm{E}[\epsilon_{it}\epsilon_{is}|\mathbf{X}] &= \frac{|\mathbf{\Sigma_{yy}}|}{|\hat{\mathbf{\Sigma}}_{\mathbf{yy}}|}A_1^{i=j} + B_1(A_1^{i=j}B_1 + A_2^{i=j})\left(\bar{Y}_i^B - \mu_\delta\right)^2 \\
&\quad + B_2(A_1^{i=j}B_2 + A_3^{i=j})\left(\sum_{k\neq i}\left(\bar{Y}_k^b - \mu_\delta\right)\right)^2 \\
&\quad + \Big(B_1(A_1^{i=j}B_2 + A_3^{i=j}) + B_2(A_1^{i=j}B_1 + A_2^{i=j})\Big)\left(\bar{Y}_i^B - \mu_\delta\right)\sum_{k\neq i}\left(\bar{Y}_k^B - \mu_\delta\right)
\end{aligned}
$$

but, similarly to above,

$$
\sum_{s=1}^{r}\left(\frac{|\mathbf{\Sigma_{yy}}|}{|\hat{\mathbf{\Sigma}}_{\mathbf{yy}}|}\sum_{t=1}^{r}A_1^{i=j}\right) = \sum_{s=1}^{r}\left(r\bar{g}_s^{i=j} - \frac{r\left(b_s c + (J-1)\,c^2\right)}{d+(J-1)e}\right) = r^2\left(\bar{g}^{i=j} - \frac{Jc^2}{d+(J-1)e}\right)
$$

for

$$\bar{g}^{i=j} = \frac{1}{r}\sum_{s=1}^{r}\bar{g}_s^{i=j} = \left(\left(1-\theta\right)^2\sigma_v^2 + \left(\frac{\theta^2}{m}+\frac{1}{r}\right)\left(\sigma_\omega^2+\sigma_\delta^2\right) + \frac{\theta^2(m-1)}{m}\psi^B + \frac{r-1}{r}\psi^A - 2\theta\psi^X\right)$$

Also, since

$$B_1\sum_{t=1}^{r}A_1^{i=j} + \sum_{t=1}^{r}A_2^{i=j} = B_2\sum_{t=1}^{r}A_1^{i=j} + \sum_{t=1}^{r}A_3^{i=j} = \frac{rc}{d+(J-1)e}$$

we can finally compute the summed expectation as

$$\sum_{t=1}^{r}\sum_{s=1}^{r}\mathrm{E}[\epsilon_{it}\epsilon_{is}|\mathbf{X}] = r^2\left[\bar{g}^{i=j} - \frac{Jc^2}{d+(J-1)e} + \left(\frac{c}{d+(J-1)e}\sum_{k=1}^{J}\left(\bar{Y}_k^B - \mu_\delta\right)\right)^2\right]$$

Similarly to before, this is invariant across all i, j with $i = j$. Moreover, because everything except $r^2(\bar{g}^{i=j} - \bar{g}^{i\neq j})$ will be summed across both the $i \neq j$ and the $i = j$ cases, thus canceling out in equation (A.6), and furthermore because

$$\bar{g}^{i=j} - \bar{g}^{i\neq j} = \left(1-\theta\right)^2\sigma_v^2 + \left(\frac{\theta^2}{m}+\frac{1}{r}\right)\sigma_\omega^2 + \frac{\theta^2(m-1)}{m}\psi^B + \frac{r-1}{r}\psi^A - 2\theta\psi^X$$

we conclude that the appropriate ANCOVA variance formula will again be that derived by Burlig et al. (2020).

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, 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]

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 $\psi^A_{\hat{\omega}}$.

4. To produce the estimated MDE, plug these values into

$$
\begin{aligned}
MDE^{est} = \left(t^J_{1-\kappa} - t^J_{\alpha/2}\right) \times \Bigg\{ & \left(\frac{1}{P(1-P)J} + \frac{\left(\bar{Y}^B_T - \bar{Y}^B_C\right)^2}{Z}\right) \times \left(\frac{I}{I-1}\right) \\
& \times \Bigg[(1-\theta^2)\sigma^2_{\hat{v}} + \left(\frac{m+\theta r}{2m^2r^2}\right)\left((m+r)\left(m+\theta r\right) + (1-\theta)(mr^2 - m^2r)\right)\sigma^2_{\hat{\omega}} \\
& + \left(\frac{m+\theta r}{2mr^2}\right)(m-1)\left(m+\theta r - (1-\theta)mr\right)\psi^B_{\hat{\omega}} \\
& + \left(\frac{m+\theta r}{2m^2r}\right)(r-1)\left(m+\theta r + (1-\theta)mr\right)\psi^A_{\hat{\omega}}\Bigg]\Bigg\}^{1/2}
\end{aligned} \tag{A.28}
$$

where $t^J_{1-\kappa}$ 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

$$
\begin{aligned}
\theta = & \frac{m\left[4mr\sigma^2_{\hat{v}} - \left(m\left(m-r+2\right) + r(r-m+2)\right)\sigma^2_{\hat{\omega}}\right]}{2r\left[2m^2\sigma^2_{\hat{v}} + (m(m+1) - r(m-1))\sigma^2_{\hat{\omega}} + (m(m-1)(m+1)\psi^B_{\hat{\omega}} - r(m-1)(r-1)\psi^A_{\hat{\omega}}\right]} \\
& + \frac{m\left[-m(m-1)(m-r+2)\psi^B_{\hat{\omega}} - r(r-1)(r-m+2)\psi^A_{\hat{\omega}}\right]}{2r\left[2m^2\sigma^2_{\hat{v}} + (m(m+1) - r(m-1))\sigma^2_{\hat{\omega}} + (m(m-1)(m+1)\psi^B_{\hat{\omega}} - r(m-1)(r-1)\psi^A_{\hat{\omega}}\right]}
\end{aligned} \tag{A.29}
$$

[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.

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,

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

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 $\mathrm{E}[\tilde{\sigma}^2_{\hat{v}}] = \sigma^2_{\hat{v}}$ is 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

$$\begin{aligned}\hat{v}_i &= \frac{1}{m+r}\sum_{t=-m+1}^{r} Y_{it} - \frac{1}{I(m+r)}\sum_{i=1}^{I}\sum_{t=-m+1}^{r} Y_{it}\\ &= v_i - \frac{1}{I}\sum_{i=1}^{I} v_i + \frac{1}{m+r}\sum_{t=-m+1}^{r}\omega_{it} - \frac{1}{I(m+r)}\sum_{i=1}^{I}\sum_{t=-m+1}^{r}\omega_{it} \qquad \text{(A.30)}\end{aligned}$$

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 $\frac{I-1}{I}$, 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 SCR 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

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

which implies equation (A.28) 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, as in Burlig et al. (2020), 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

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

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

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

which gives θ as equation (A.29). Again, these solutions differ from the original results only in that all coefficients (rather than just the k_v coefficients) include $\frac{I}{I-1}$.