Difference in Difference Meets Generalized Least Squares: Higher Order Properties of Hypotheses Tests*

Jerry Hausman[†] Guido Kuersteiner[‡]

June 2004

Abstract

We investigate estimation and inference in difference in difference econometric models used in the analysis of treatment effects. When the innovations in such models display serial correlation, commonly used ordinary least squares (OLS) procedures are inefficient and may lead to tests with incorrect size. Implementation of feasible generalized least squares (FGLS) procedures is often hindered by too few observations in the cross section to allow for unrestricted estimation of the weight matrix without leading to tests with similar size distortions as conventional OLS based procedures. We analyze the small sample properties of FGLS based tests with a higher order Edgeworth expansion that allows us to construct a size corrected version of the test. We also address the question of optimal temporal aggregation as a method to reduce the dimension of the weight matrix. We apply our procedure to data on regulation of mobile telephone service prices. We find that a size corrected FGLS based test outperforms tests based on OLS.

*We thank Matthew Harding, Alan Grant, Anne Hall and Jie Yang for excellent research assistance. Whitney Newey engaged in helpful conversations.

[†]MIT, Department of Economics, 50 Memorial Drive, Cambridge, MA 02142; email: jhausman@mit.edu.

[‡]MIT, Department of Economics, 50 Memorial Drive, Cambridge, MA 02142, email: gkuerste@mit.edu.
Financial support from NSF Grant SES 0095132 is gratefully acknowledged.

1. Introduction

We investigate estimation and inference in difference in difference (DID) econometric models. DID models have become a widely used method to investigate changes in policy variable, which often arise from changes in legislation. An example would arise when a group of states passes new legislation that mandates firms to provide a change in benefit levels to their employees. A DID model allows estimation of the effect, if any, on an outcome variable such as wages. The typical approach is to use panel data on the 50 U.S. states for a time period of say 5 or more years to estimate a fixed effects model with fixed effects for both states and for time. A given state that adopts the legislation acts as its own "control" in the pre-legislation period while states that do not adopt the legislation act as "control observations" in the post-legislation period. The most straightforward situation occurs when all states that adopt the legislation do so in the same year. Assuming that state characteristics do not change over the period, the difference of the before and after period for the adopting states minus the difference of the before and after period for the non-adopting state yields the DID estimator. When states adopt the legislation in different periods or state characteristics change over time, a fixed effects estimator typically replaces the more straightforward DID approach, but the underlying logic remains similar.

However, this approach does not yield the best estimator in terms of efficiency or the most precise inference. Both the DID approach and the fixed effects approach do not utilize all of the time series variation in the data if the variance of the stochastic disturbances is not spherical. Consider the panel data model

$$y_{it} = T_{it}\gamma + z_{it}'\theta + \alpha_i + \varepsilon_{it};\ i = 1, ..., N;\ t = 1, ..., T \tag{1.1}$$

where T_{it} measures a policy variable or is a dummy variable for a change in policy or regulation, the parameter θ includes the time fixed effects, the α_i are the state fixed effects and ε_{it} is orthogonal to the right hand side variable, independent across i but possibly correlated with ε_{is} for all $s, t = 1, ...T$. We do not assume stationarity or any parametric form of dependence for ε_{it}. The cross-sectional sample size N is not necessarily large enough for first order asymptotic approximates to yield reliable results, T is "small", and Σ is unconstrained. Least squares (OLS) on Equation (1.1) yields unbiased estimates, but the estimate of the variance of the estimated parameters must be adjusted for accurate inference to take account of the non-diagonality of Σ, as Bertrand et. al. (2002) have recently emphasized. Otherwise, as Moulton (1986) pointed out, the unadjusted OLS standard errors often have a substantial downward bias.

However, the more efficient estimator of equation (1) would be generalized least squares (GLS) if Σ were known. Indeed, GLS is the Gauss-Markov estimator and would lead to optimal inference, e.g. uniformly most powerful tests, on the effect of the legislation. In the usual situation when Σ is unknown and needs to be estimated, the usual estimator would be "feasible" GLS (FGLS) where a consistent estimate $\hat{\Sigma}$ replaces Σ in the GLS formula. Indeed, if the estimate of Σ is unrestricted, FGLS is unbiased along with OLS and GLS. However, very few empirical examples of DID appearing in the literature use FGLS[1]. Instead, OLS is the estimator of choice.

FGLS on equation (1) is easy to implement. We estimate $\hat{\Sigma}$ from either an OLS estimator of Equation (1.1) using an approach that we develop to eliminate the bias from fixed effects estimators or we first difference the data to eliminate the state effects and proceed with the differenced model. If N were large enough, we would use the usual result that $p\lim\left[\sqrt{N}\left(\hat{\delta}_{GLS}-\hat{\delta}_{FGLS}\right)\right]=0$ where $\delta=(\theta,\gamma)$ so long as $\operatorname{plim}\hat{\Sigma}=\Sigma$. However, in many applications of DID, N is unlikely to be large enough in relation to the number of time periods T to permit the first order asymptotic approximation to be sufficiently accurate to provide accurate inference. For example, if $T=10$, the number of unknown elements in Σ is 55 compared to a sample size of 500. Thus, in this paper we use a second order Edgeworth approximation approach of Rothenberg (1988) that accounts of the uncertainty in estimating $\hat{\Sigma}$. We adjust the test statistics, which affects both the size and power of the tests. Otherwise, often the actual size of the test may considerably exceed the nominal size of the test because the usual test statistics assume that the FGLS estimator is close enough to the GLS estimator so that no adjustment for the estimation of $\hat{\Sigma}$ arises.

[1]Bertrand et. al. (2002) in their survey of the literature find only one paper out of nearly 100 papers that uses this approach where Σ is unrestricted.

Once we consider the effects of uncertainty in the estimate of $\hat{\Sigma}$, the question arises of whether a trade-off exists between some amount of averaging across time to reduce the dimension of the variance-covariance matrix needed for FGLS estimation to improve estimator efficiency. Since the number of unknown parameters in an unrestricted Σ increases at rate T^2, aggregation to reduce the dimension of Σ can lead to a significant decrease in the number of unknown parameters. The prior literature has emphasized this idea, e.g. Moulton (1986), with the DID approach of using OLS on the two before and after periods the most extreme possible approach. In this paper for a given design matrix $[T_{it}, z_{it}]$ and an unrestricted estimate $\hat{\Sigma}$, we solve for the optimal degree of aggregation using the Edgeworth expansions of Rothenberg (1988). For a given size of test calculated to second order, we choose the degree of aggregation that maximizes the power of the test statistic again calculated to second order. We demonstrate that for a commonly occurring situation where once the treatment begins in a state it continues thereafter, that small sample benefit from time aggregation can arise. We next consider the situation where the initial treatment date is the same for all states that pass the legislation. Again we find that the size of the test is affected by estimation of Σ, but we find that the higher order power is not affected. Thus, the theoretically optimal solution is not to undertake temporal aggregation. We also demonstrate that in this special situation, size corrected tests based on OLS estimation do not lead to a recommendation of temporal aggregation, in contrast to the previous literature.

In our analysis we focus on Wald tests of the hypothesis $H_0 : \gamma = 0$. Rothenberg (1984b) shows that for hypotheses only involving one dimensional parameters LR, LM and Wald tests have the same power up to order $o\left(N^{-1}\right)$ after correcting for size distortions. This result means that all three tests are affected in similar ways by the problem of estimating Σ. The focus on the Wald test is further motivated by the fact that it is the most commonly used test in practice and that the invariance properties of the LR test play a lesser role in the context of the linear restrictions we are focusing on here.

We then consider some Monte Carlo evidence on the performance of our approach and the second order Edgeworth approximations. We consider a situation with positive serial correlation across time for states, which is the usual situation found in applied research. So far in our empirical research, we have considered the single treatment date situation. For estimation with N=50 and T equal to (5,10, 15, 20) we find that size corrected FGLS, FGLS-SC, in levels does almost as well as GLS with known Σ. We also find that FGLS-SC has significantly more power than does OLS with a robustly estimated covariance matrix, robust OLS, when serial correlation in the level data is high. Thus FGLS appears to be the better estimator even with additional parameter uncertainty created by the estimated $\hat{\Sigma}$. We also consider two other versions of FGLS for 3 periods (before, change period, and after) and the "traditional" 2 period (before and after) DID approach. We find that both of these alternative approaches involving time aggregated have significantly reduced power compared to FGLS-SC. Thus, we do not recommend their use.

We next consider a first difference specification that also eliminates the fixed effects but can also lead to a reduced effect of the positive serial correlation. We now find that FGLS-SC does almost as well as GLS. For low and moderate amounts of serial correlation FGLS does significantly better than robust OLS. However, for a high degree of serial correlation robust OLS on first differences does as well (or even somewhat better) than FGLS. We then consider 3 period and 2 period time aggregation estimators[2]. We find that all size distortions have been eliminated in FGLS-SC. We also find that the 3 period version of FGLS outperforms the 2 period version by a large amount. Indeed, the 3 period aggregation FGLS-SC estimator seems to do the best of all the feasible estimators considered with correct size and maximum power.

These results suggest to use full sample FGLS-SC whenever serial correlation is high in levels. If the regressions are run in first differences the 3 period version of FGLS-SC seems to perform best. An argument for running the specification in levels can be made in cases where adjustment to the new policy takes more than one time period. In this case, the first difference specification will underestimate the total effect of the policy relative to the level specification.

In a final section we provide an application of our method to a data set for mobile telephone service prices. We exploit a 1994 FCC ruling that required all states to abolish price regulation of the mobile telephone industry. This ruling provides a natural experiment to test the hypothesis

[2]Since the data have been initially transformed to first differences, these estimators differ from the earlier fixed effects estimatiors on 2 or 3 periods.

that regulation led to higher service charges for mobile telephone services prior to 1994 in the states that had such price regulation in place. When we run robust OLS on the entire sample the t-statistic for a significant difference between pre and post regulatory regimes comes in insignificantly. We compare this result with full sample FGLS using the higher order size correction. The test statistic now indicates a significant treatment effect. Moreover, the point estimate of the FGLS regression is almost identical to the estimated price effect of regulation in an earlier cross-sectional study by Hausman (1995).

2. Tests based on OLS and GLS

2.1. Level Specification

In this section we turn to the original model formulated in levels. The analysis is complicated by the presence of fixed effects, which amongst other things complicate estimation of the weight matrix. We consider

$$y_{it} = T_{it}\gamma + z_{it}'\theta + \alpha_i + \varepsilon_{it}$$

where we define $\tilde{Y}_t = [y_{1,t}, ..., y_{n,t}]'$, $\tilde{Y} = \left[\tilde{Y}_1', ..., \tilde{Y}_T'\right]'$, $\tilde{Z}_t = [z_{1t}, ..., z_{nt}]'$ and $\tilde{Z} = \left[\tilde{Z}_1', ..., \tilde{Z}_T'\right]'$. The vector $\tilde{\Upsilon}$ is defined as $\tilde{\Upsilon}_t = [T_{1t}, ..., T_{nt}]'$, $\tilde{\Upsilon} = \left[\tilde{\Upsilon}_1', ..., \tilde{\Upsilon}_T'\right]'$ and we let $\alpha = [\alpha_1, ..., \alpha_n]'$. We assume that γ is a scalar to simplify the subsequent arguments. Then we can write $\tilde{Y} = \tilde{\Upsilon}\gamma + \tilde{Z}\theta + (\mathbf{1}_T \otimes I_n)\,\alpha + \tilde{\varepsilon}$ where $E\tilde{\varepsilon} = 0$ and $\mathrm{Var}(\tilde{\varepsilon}) = \tilde{\Sigma} \otimes I_n \equiv \tilde{\Omega}$. Also define $V_{it} = [T_{it}, Z_{it}']$ with regressor matrix $V = [V_1', ..., V_n']'$ where $V_i = [V_{i1}, ..., V_{iT}]$. We impose the following condition on the fixed effects.

Condition 1. *Conditional on $\tilde{Z}$ and $\tilde{\Upsilon}$, the fixed effects α_i are distributed normally with $E\left[\alpha_i|V_i\right] = V_i m_\alpha$ and $E\left[\alpha_i|V_i\right] = \sigma_\alpha^2$ where m_α is a vector of constants and α_i are independent across i and independent of η_{it}.*

Remark 1. *This assumption corresponds to the specification of Mundlak (1978).*

Due to the presence of fixed effects and the associated incidental parameter problem it is not possible to construct unbiased estimates of the weight matrix directly. Bias corrections may be available in some cases but they usually do not completely remove the bias and they typically depend on a stationarity assumption, which may not be accurate. This is particularly the case when a policy change occurs which is the typical situation in difference in difference regressions. Here we propose an alternative weight matrix estimator for the level case that is an unbiased estimate of a certain transformation of the weight matrix[3]. Unlike more well-known bias corrected estimators our estimator does not require the serial correlation in ε_{it} to be of a particular parametric form, nor does it require the process ε_{it} to be stationary. Absence of stationarity may lead to extremely poor performance of the usual bias corrected estimators and a stationarity assumption is inconsistent with our assumption of an unrestricted Σ.

The idea behind our estimator is to fit a misspecified OLS regression where the fixed effects are not estimated for each time period separately. The residuals from this regression are then used to compute temporal covariances. Due to the omitted fixed effects the covariances will all have the same constant in expectation. The final step consists of projecting out the common constant. Define the projection matrices M_V and $M_{\mathbf{1}_T}$ projecting onto the orthogonal complement of V and $\mathbf{1}_T$ respectively. We obtain residuals $\hat{\eta}_t = M_V \tilde{Y}_t$ and estimate the element $\sigma_{t,s}$ of $\tilde{\Sigma}$ as

$$\hat{\sigma}_{t,s} = \frac{\tilde{Y}_t' M_V \tilde{Y}_s}{\operatorname{tr}(M_V)}.$$

We now form the $T \times T$ matrix $\tilde{S}$ consisting of the elements $\hat{\sigma}_{t,s}$. We then form the estimate $\hat{\Sigma}$ of $\tilde{\Sigma}$ as

$$\hat{\Sigma} = M_{\mathbf{1}_T} \tilde{S} M_{\mathbf{1}_T}. \tag{2.1}$$

It is shown in the appendix, that $E\hat{\Sigma} = M_{\mathbf{1}_T} \tilde{\Sigma} M_{\mathbf{1}_T}$. Note that $\hat{\Sigma}$ is of rank $T-1$ due to a loss of degrees of freedom resulting from the estimation of the fixed effects. A full rank matrix can be obtained by deleting the column and row from $\hat{\Sigma}$ which amounts to only using time periods $2, ..., T$. For this purpose define the $T-1 \times T$ matrix B obtained from deleting the first row of I_T. Estimation of (3.1) can be achieved by applying the transformation

$$(BM_{\mathbf{1}_T} \otimes I_n) \tilde{Y} = \sum_{t=1}^{T} (BM_{\mathbf{1}_T} a_t \otimes Z_t) \theta + (BM_{\mathbf{1}_T} \otimes I_n) \tilde{\Upsilon} \gamma + (BM_{\mathbf{1}_T} \otimes I_n) \tilde{\varepsilon}$$

[3] A similar procedure was proposed by Kiefer (1980) but he did not establish unbiasedness.

to remove the fixed effects. The transformed innovations $(BM_{\mathbf{1}_T} \otimes I_n)\,\tilde{\varepsilon}$ have variance covariance matrix $\tilde{\Omega} = BM_{\mathbf{1}_T}\tilde{\Sigma}M_{\mathbf{1}_T}B' \otimes I_n$ such that GLS can be implemented by using the estimator $\hat{\Sigma}$[4].

For the time being the autocorrelation structure of $\tilde{\Sigma}_T$ is assumed unrestricted. Thus, OLS, GLS and FGLS are all unbiased estimators.

Estimation of the parameter γ can be done using OLS or GLS. If the dimension of T is relatively large compared to the dimension of n, it has been argued in the literature that averages across time can be used to reduce the dimensionality of the variance-covariance matrix needed for GLS estimation and for hypothesis testing in both the OLS and GLS case. In order to formalize this idea we define the $T \times r$ selector matrix C such that $C'C = I_r$.

We define $Y = (C'BM_{\mathbf{1}_T} \otimes I_n)\,\tilde{Y}$. In the simplest case where $r = 1$ and $C' = [1, ..., 1]/T$ it follows that $Y = \left[T^{-1}\sum_t y_{1,t}, ..., T^{-1}\sum_t y_{n,t}\right]'$. Similarly we define $\Upsilon = (C'BM_{\mathbf{1}_T} \otimes I_n)\,\tilde{\Upsilon}$, $Z = (C'BM_{\mathbf{1}_T} \otimes I_n)\,\tilde{Z}$ and $\varepsilon = (C'BM_{\mathbf{1}_T} \otimes I_n)\,\tilde{\varepsilon}$. Note that $E\varepsilon = 0$ and $E\varepsilon\varepsilon' = \Omega = \Sigma \otimes I_n$ where $\Sigma = C'BM_{\mathbf{1}_T}\tilde{\Sigma}M_{\mathbf{1}_T}B'C$.

The model now can be written as $Y = \Upsilon\gamma + Z\theta + \varepsilon$. If Σ were known, two tests for the hypothesis $H_0 : \gamma = \gamma_0$ could be considered. Let $\Omega_z = \Omega^{-1} - \Omega^{-1}Z(Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}$ and $M_Z = I_{nT} - Z(Z'Z)^{-1}Z'$. In Rothenberg's (1988) terminology we define the test

$$\bar{T}_1 = \frac{\left(\Upsilon'\Omega_z\Upsilon\right)^{-1}\Upsilon'\Omega_z Y - \gamma_0}{\left(\Upsilon'\Omega_z\Upsilon\right)^{-1/2}}$$

based on GLS estimation for γ which is the Gauss Markov (BLUE) estimator and the test

$$\bar{T}_2 = \frac{\left(\Upsilon'M_z\Upsilon\right)^{-1}\Upsilon'M_z Y - \gamma_0}{\left(\left(\Upsilon'M_z\Upsilon\right)^{-1}\Upsilon'M_z\Omega M_z\Upsilon\left(\Upsilon'M_z\Upsilon\right)^{-1}\right)^{1/2}}$$

which is based on the OLS estimate for γ and on robust standard errors. Under the additional assumption of Gaussian errors or under standard first order asymptotics where $n \to \infty$ it can be shown that the power of both tests depends on

$$b_1 = \frac{\gamma - \gamma_0}{\left(\Upsilon'\Omega_z\Upsilon\right)^{-1/2}}$$

[4]This approach to an unbiased estimate of Σ avoids the "Hurwicz bias" discussed in the literature because only covariances, not regression coefficients, are estimated.

and

$$b_2 = \frac{\gamma - \gamma_0}{\left(\left(\Upsilon' M_z \Upsilon\right)^{-1} \Upsilon' M_z \Omega M_z \Upsilon \left(\Upsilon' M_z \Upsilon\right)^{-1}\right)^{1/2}}.$$

From standard arguments it follows that $b_1 \geq b_2$ so that the power of the test based on GLS exceeds the power of the test based on OLS. It also follows that for $\Upsilon = (C_r' \otimes I_n)\,\tilde{\Upsilon}$ and C_r a $T \times r$ matrix with $r \leq T$ and Ω_{z_r} a corresponding matrix defined as before but based on C_r we have

$$\tilde{\Upsilon}' \Omega_{z_T} \tilde{\Upsilon} - \Upsilon' \Omega_{z_r} \Upsilon \geq 0$$

where $\geq$ stands for 'positive definite' such that from a first order asymptotic point of view it is never optimal to average the observations. This result is an application of the Gauss-Markov Theorem when Ω is known. For first order asymptotics, because of the block diagonality of the information matrix for $\hat{\gamma}$ and $\hat{\Omega}$, Ω is treated as known in the expansions.

We now turn to the analysis of tests where Ω is replaced with the estimator $\hat{\Omega}$ where $\hat{\Omega} = C' B M_{\mathbf{1}_T} \hat{\Sigma} M_{\mathbf{1}_T} B' C \otimes I_n$ with $\hat{\Sigma}$ defined as in (2.1).

Additional regularity conditions needed to formally justify the expansions of Rothenberg (1988) used in the development of our results are stated next.

Condition 2. *All asymptotic arguments are for T fixed and $n \to \infty$. Assume that $\tilde{\varepsilon}$ is jointly normal with $\tilde{\varepsilon} \sim N(0, \tilde{\Sigma} \otimes I_n)$. Let $X = [Z, \Upsilon]$ be a $nT \times k$ matrix. Assume that X is fixed with full column rank k. Assume that Ω is of full rank. Let $\hat{\beta} = \left(X' \hat{\Omega}^{-1} X\right)^{-1} X' \hat{\Omega}^{-1} Y$ and $\bar{\beta} = \left(X' \Omega^{-1} X\right)^{-1} X' \Omega^{-1} Y$. For any vector $c \in \mathbb{R}^k$ with $\|c\| = 1$ we assume that there exist random variables $\zeta_{1,n}$ and $\zeta_{2,n}$ such that $\sqrt{n} c' \left(\hat{\beta} - \bar{\beta}\right) = \zeta_{1,n}/\sqrt{n} + \zeta_{2,n}/n^{3/2}$ where $\zeta_{1,n}$ has bounded moments as $n \to \infty$ and $P\left[\left|\zeta_{2,n}\right| > (\log n)^q\right] = o\left(n^{-1}\right)$ for some q.*

Remark 2. *The regressors X are assumed to be fixed. Alternatively, we can regard the analysis as being conditional on a particular draw of regressors.*

Remark 3. *Note that by a Taylor expansion the term $\zeta_{1,n}$ is a polynomial function of $\tilde{\varepsilon}$ and thus has bounded moments of any order. Again by the Taylor Theorem, the remainder term $\zeta_{2,n}$ is also a polynomial in $\tilde{\varepsilon}$ as long as Σ has full rank in a neighborhood of Σ.*

In the Appendix we obtain an expression for $V_\Sigma = \mathrm{Var}(\mathrm{vec}\, C' B M_{\mathbf{1}_T} \hat{\Sigma} M_{\mathbf{1}_T} B' C)$. Let the K_{rr} be the $r^2 \times r^2$ commutation matrix of Magnus and Neudecker (1978) defined as $K_{rr} = \sum_{i,j=1}^{r} e_i e_j' \otimes e_j e_i'$ where e_i is the i-th unit vector of dimension r. Note that V_Σ is singular because of repeated elements in Σ. It then follows that

$$V_\Omega = (I_r \otimes K_{nr} \otimes I_n) \left(V_\Sigma \otimes \mathrm{vec}\, I_n \left(\mathrm{vec}\, I_n \right)' \right) (I_r \otimes K_{nr}' \otimes I_n)$$

with

$$V_\Sigma = (I_T + K_{TT}) (\Sigma \otimes \Sigma)$$

and $\Sigma = C' B M_{\mathbf{1}_T} \tilde{\Sigma} M_{\mathbf{1}_T} B' C$. We use the Edgeworth expansions of Rothenberg (1988) to obtain more precise statements about the finite sample behavior of the test statistics. For this purpose, let $\hat{\Omega}_z = \hat{\Omega}^{-1} - \hat{\Omega}^{-1} Z (Z' \hat{\Omega}^{-1} Z)^{-1} Z' \hat{\Omega}^{-1}$ and write

$$T_1 = \frac{\left(\Upsilon' \hat{\Omega}_z \Upsilon \right)^{-1} \Upsilon' \hat{\Omega}_z Y - \gamma_0}{\left(\Upsilon' \hat{\Omega}_z \Upsilon \right)^{-1/2}} = \frac{\bar{T}_1 + n^{-1/2} R}{\left(1 + n^{-1/2} S\right)^{1/2}}$$

with

$$S = \sqrt{n} \frac{\left(\Upsilon' \hat{\Omega}_z \Upsilon \right)^{-1} - \left(\Upsilon' \Omega_z \Upsilon \right)^{-1}}{\left(\Upsilon' \Omega_z \Upsilon \right)^{-1}}$$

and

$$R = \sqrt{n} \frac{\left(\Upsilon' \hat{\Omega}_z \Upsilon \right)^{-1} \Upsilon' \hat{\Omega}_z Y - \left(\Upsilon' \Omega_z \Upsilon \right)^{-1} \Upsilon' \Omega_z Y}{\left(\Upsilon' \Omega_z \Upsilon \right)^{-1/2}}.$$

We use a stochastic expansion of the variables S and R to obtain an Edgeworth expansion for the test statistic. Define $h = \Omega_z \Upsilon \left(\Upsilon' \Omega_z \Upsilon \right)^{-1}$ and $H = \Omega_z - \Omega_z \Upsilon \left(\Upsilon' \Omega_z \Upsilon \right)^{-1} \Upsilon' \Omega_z$. Then

$$S = \frac{(h' \otimes h')}{h' \Omega h} \sqrt{n}\, \mathrm{vec} \left(\hat{\Omega} - \Omega \right) - \sqrt{n} \frac{\mathrm{tr} \left[\mathrm{vec}(\Omega - \hat{\Omega})\, \mathrm{vec}(\Omega - \hat{\Omega})' \left(H \otimes h h' \right) \right]}{h' \Omega h} + o_p(n^{-1/2}) \tag{2.2}$$

and

$$R = \frac{(\varepsilon' H \otimes h')}{(h' \Omega h)^{1/2}} \sqrt{n}\, \mathrm{vec} \left(\Omega - \hat{\Omega} \right) + o_p(1) \tag{2.3}$$

where the derivation of (2.2) and (2.3) is reported in the Appendix. The distribution of S and R can be approximated by a distribution where $ER = 0$, $\operatorname{cov}(R, S) = O(n^{-1})$,

$$ES = -n^{-1/2} \frac{\operatorname{tr} V_\Omega (H \otimes hh')}{h'\Omega h} + O(n^{-1})$$

$$\operatorname{var}(S) = \frac{\operatorname{tr} V_\Omega (hh' \otimes hh')}{(h'\Omega h)^2}$$

and

$$\operatorname{var}(R) = \frac{\operatorname{tr} V_\Omega (H \otimes hh')}{h'\Omega h}.$$

Based on these approximations the test based on GLS estimates for γ and given explicitly as

$$T_1 = \frac{\left(\Upsilon'\hat{\Omega}_z\Upsilon\right)^{-1} \Upsilon'\hat{\Omega}_z Y - \gamma_0}{\left(\Upsilon'\hat{\Omega}_z\Upsilon\right)^{-1/2}}$$

then has a formal order n^{-1} Edgeworth approximation

$$\Pr(T_1 \le t) \simeq \Phi\left[t\left(1 - \frac{A_1(t)}{2n}\right) - b_1\left(1 - \frac{B_1(t)}{2n}\right)\right]$$

for arbitrary nonrandom t where $\Phi(.)$ the standard normal CDF. The functions $A_1(t)$ and $B_1(t)$ are defined as

$$A_1(t) = \frac{1}{4}\left(1 + t^2\right) \frac{\operatorname{tr} V_\Omega (hh' \otimes hh')}{(h'\Omega h)^2} + 2\frac{\operatorname{tr} V_\Omega (H \otimes hh')}{h'\Omega h}$$

and

$$B_1(t) = \frac{1}{4}t^2 \frac{\operatorname{tr} V_\Omega (hh' \otimes hh')}{(h'\Omega h)^2} + \frac{\operatorname{tr} V_\Omega (H \otimes hh')}{h'\Omega h}.$$

These results make explicit use of the fact that the weight matrix Ω is estimated without bias. Additional bias terms would need to be included in $A_1(t)$ if biased estimators were used.

Next consider the test

$$T_2 = \frac{(\Upsilon' M_z \Upsilon)^{-1} \Upsilon' M_z Y - \gamma_0}{\left((\Upsilon' M_z \Upsilon)^{-1} \Upsilon' M_z \hat{\Omega}_\eta M_z \Upsilon (\Upsilon' M_z \Upsilon)^{-1}\right)^{1/2}}$$

where $\hat{\Omega}$ is the same unbiased estimator of Ω as used for T_1.

For the test T_2 define $x=\sqrt{n}M_z\Upsilon\left(\Upsilon' M_z\Upsilon\right)^{-1}$ where the t-th $n\times 1$ block of x is denoted as x_t. In the Appendix we show that the Edgeworth approximation of order n^{-1} of T_2 is given by

$$\Pr\left(T_2\leq t\right)=\Phi\left[t\left(1-\frac{A_2\left(t,\Omega\right)}{2n}\right)-b_2\left(1-\frac{B_2\left(t,\Omega,b_2\right)}{2n}\right)\right]$$

where

$$A_2\left(t,\Omega\right)=\frac{1}{4}\left(1+t^2\right)\frac{\operatorname{tr}V_\Omega\left(xx'\otimes xx'\right)}{\left(x'\Omega x\right)^2}$$

and

$$B_2\left(t,\Omega,b_2\right)=\frac{1}{4}t^2\frac{\operatorname{tr}V_\Omega\left(xx'\otimes xx'\right)}{\left(x'\Omega x\right)^2}.$$

The most important application of these expansions lies in the construction of a size corrected test based on the GLS estimator. Based on Rothenberg (1988) one can proceed as follows. A size corrected test is achieved by rejecting $H_0:\gamma=\gamma_0$ if

$$T_1>t_c$$

where

$$t_c=t_\alpha\left(1+\frac{A_1\left(t_\alpha\right)}{2n}\right)$$

where t_α is the critical value satisfying $\Phi\left(t_\alpha\right)=1-\alpha$. In principle similar size corrected tests could be achieved for OLS. Our analysis in the next section of a special case of particular interest reveals however, that OLS seems to be far less sensitive to the dimension of the unknown covariance matrix Σ and Monte Carlo evidence indicates that tests based on robust OLS have approximately correct size without the correction.

In general the constant $A_1\left(t_\alpha\right)$ needs to be replaced with an estimate. As Rothenberg (1988) points out this is usually without consequences. This argument remains valid under our asymptotic approximation where $n\to\infty$ while T is kept fixed. Computation of $A_1\left(t_\alpha\right)$ requires us to evaluate $\operatorname{tr}V_\Omega\left(hh'\otimes hh'\right)$ and $\operatorname{tr}V_\Omega\left(H\otimes hh'\right)$. Because the dimensions of $V_\Omega\left(hh'\otimes hh'\right)$ and $V_\Omega\left(H\otimes hh'\right)$ can be very large it is more convenient for computational purposes to use the expression

$$\begin{aligned}(2.4)\qquad \operatorname{tr}V_\Omega\left(hh'\otimes hh'\right) &= \textstyle\sum_{i,j=1}^n\left(h'\left(\Sigma\otimes b_ib_j'\right)h\right)^2\\ &\quad+\textstyle\sum_{l,m=1}^r\sum_{i,j=1}^n\left(h'\left(e_le_m'\Sigma\otimes b_ib_j'\right)h\right)\left(h'\left(e_me_l'\Sigma\otimes b_ib_j'\right)h\right)\end{aligned}$$

where b_i is the i-th unit vector of length n. Moreover one can write

$$\text{(2.5)} \quad \operatorname{tr} V_\Omega \left(H \otimes hh'\right) = \textstyle\sum_{i,j=1}^{n} \operatorname{tr}\left[\left(\Sigma \otimes b_i b_j'\right) H\right]\left(h'\left(\Sigma \otimes b_i b_j'\right) h\right)$$
$$+ \textstyle\sum_{l,m=1}^{r} \sum_{i,j=1}^{n} \operatorname{tr}\left(\left(e_l e_m' \Sigma \otimes b_i b_j'\right) H\right)\left(h'\left(e_m e_l' \Sigma \otimes b_i b_j'\right) h\right).$$

A derivation of these formulas is contained in the Appendix. Furthermore, the analysis in Section 3 considers a case where $A_1(t_\alpha)$ takes a particularly simple form that does not depend on V_Ω and thus no estimation is required.

2.2. First Difference Specification

An alternative to the level specification is a transformation to first differences. This approach is often advocated to remove fixed effects. One caveat of applying a first difference transformation is that it may lead to tests with poor power when a policy takes more than one period to take its full effect. Despite these potential problems we turn to models formulated in first differences and show that the previous results essentially remain valid without change. We thus consider the model

$$\Delta y_{it} = \Delta z_{it}'\theta + \Delta T_{it}\gamma + \Delta\varepsilon_{it}$$

where Δy_{it} is formulated in first differences to remove fixed effects and the $k-1$ dimensional exogenous regressor Δz_{it} contains time effects as well as other covariates.

We stack the observations as $\Delta\tilde{Y} = [\Delta y_{1,2}, ..., \Delta y_{n,2}, ..., \Delta y_{1T}, ..., \Delta y_{nT}]'$, $\tilde{\Upsilon}_t^\Delta = [\Delta T_{1t}, ..., \Delta T_{nt}]'$, $\tilde{\Upsilon}^\Delta = \left[\tilde{\Upsilon}_2^{\Delta\prime}, ..., \tilde{\Upsilon}_T^{\Delta\prime}\right]'$, $\tilde{Z}_t^\Delta = [\Delta z_{1t}, ..., \Delta z_{nt}]'$ and $\tilde{Z}^\Delta = \left[\tilde{Z}_2^{\Delta\prime}, ..., \tilde{Z}_T^{\Delta\prime}\right]'$ with $\Delta\tilde{\varepsilon}$ being the corresponding vector of error terms. The model then can be written as

$$\Delta\tilde{Y} = \tilde{Z}^\Delta\theta + \tilde{\Upsilon}^\Delta\gamma + \Delta\tilde{\varepsilon} \tag{2.6}$$

with $E\Delta\tilde{\varepsilon} = 0$ and $E\Delta\tilde{\varepsilon}\Delta\tilde{\varepsilon}' = \tilde{\Sigma}_{\Delta,T} \otimes I_n \equiv \tilde{\Omega}_\Delta$. If we define the $T-1 \times T$ matrix

$$B^\Delta = \begin{bmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{bmatrix}$$

then (2.6) can be obtained from the level specification by noting that $\Delta\tilde{Y} = \left(B^{\Delta} \otimes I_n\right)\tilde{Y}$, $\tilde{Z}^{\Delta} = \left(B^{\Delta} \otimes I_n\right)\tilde{Z}$, $\tilde{\Upsilon}^{\Delta} = \left(B^{\Delta} \otimes I_n\right)\tilde{\Upsilon}$ and $\Delta\tilde{\varepsilon} = \left(B^{\Delta} \otimes I_n\right)\tilde{\varepsilon}$. It thus follows that $\tilde{\Omega}_{\Delta} = B^{\Delta}\tilde{\Sigma}B^{\Delta\prime} \otimes I_n$. Moreover, because $B^{\Delta}\mathbf{1}_T = 0$ it also follows that $EB^{\Delta}\tilde{S}B^{\Delta} = B^{\Delta}\tilde{\Sigma}B^{\Delta\prime}$. In other words we continue to use the same estimator $\tilde{S}$ for the covariance matrix but apply the operator B^{Δ} when the model is specified in first differences.

As before we can then consider the transformed model

$$\Delta Y = Z^{\Delta}\theta + \Upsilon^{\Delta}\gamma + \Delta\varepsilon$$

with $\Delta Y = (C' \otimes I_n)\Delta\tilde{Y}$, $\Upsilon^{\Delta} = (C' \otimes I_n)\tilde{\Upsilon}^{\Delta}$, $Z^{\Delta} = (C' \otimes I_n)\tilde{Z}^{\Delta}$ and $\Delta\varepsilon = (C' \otimes I_n)\Delta\tilde{\varepsilon}$ such that $\Omega_{\Delta} = E\Delta\varepsilon\Delta\varepsilon' = \Sigma_{\Delta} \otimes I_n$ with $\Sigma_{\Delta} = C'\tilde{\Sigma}_{\Delta,T}C$. We impose the following additional restriction on C.

Condition 3. *For all C such that $C'C = I_r$ it follows that $\left[Z_t^{\Delta}, \Upsilon_t^{\Delta}\right]$ has full column rank.*

Remark 4. *A sufficient condition for the last part is that $\left[\tilde{Z}_t^{\Delta}, \tilde{\Upsilon}_t^{\Delta}\right]$ has full column rank and that $\tilde{Z}_t^{\Delta}, \tilde{\Upsilon}_t^{\Delta}$ are stationary. Condition 3 is violated in at least one case of interest that we treat separately in Section 3.*

Because the estimator of the covariance matrix remains unbiased under the first difference transformation the expansions developed for the level case remain valid except for notational adjustments. As before we therefore define $h_{\Delta} = \Omega_z\Upsilon^{\Delta}\left(\Upsilon^{\Delta\prime}\Omega_z\Upsilon^{\Delta}\right)^{-1}$ and $H_{\Delta} = \Omega_z - \Omega_z\Upsilon^{\Delta}\left(\Upsilon^{\Delta\prime}\Omega_z\Upsilon^{\Delta}\right)^{-1}\Upsilon^{\Delta\prime}\Omega_z$. Using these results we conclude from Rothenberg (1988) and our previous analysis for the level case that

$$\Pr(T_1 \leq t) \simeq \Phi\left[t\left(1 - \frac{A_1(t)}{2n}\right) - b_1\left(1 - \frac{B_1(t)}{2n}\right)\right]$$

where

$$A_1(t) = \frac{1}{4}\left(1 + t^2\right)\frac{\operatorname{tr} V_{\Omega_{\Delta}}\left(h_{\Delta}h_{\Delta}' \otimes h_{\Delta}h_{\Delta}'\right)}{\left(h_{\Delta}'\Omega_{\Delta}h_{\Delta}\right)^2} + 2\frac{\operatorname{tr} V_{\Omega_{\Delta}}\left(H_{\Delta} \otimes h_{\Delta}h_{\Delta}'\right)}{h_{\Delta}'\Omega_{\Delta}h_{\Delta}}$$

and

$$B_1(t) = \frac{1}{4}t^2\frac{\operatorname{tr} V_{\Omega_{\Delta}}\left(h_{\Delta}h_{\Delta}' \otimes h_{\Delta}h_{\Delta}'\right)}{\left(h_{\Delta}'\Omega h_{\Delta}\right)^2} + \frac{\operatorname{tr} V_{\Omega_{\Delta}}\left(H_{\Delta} \otimes h_{\Delta}h_{\Delta}'\right)}{h_{\Delta}'\Omega h_{\Delta}}.$$

For the robust OLS based test define $x_\Delta = \sqrt{n} M_{z\Delta} \Upsilon^\Delta \left(\Upsilon^{\Delta\prime} M_{z\Delta} \Upsilon^\Delta \right)^{-1}$ and the t-th $n \times 1$ block of x_Δ is denoted as $x_{\Delta t}$. Then the Edgeworth approximation for T_2 is given by

$$\Pr\left(T_2 \leq t\right) = \Phi \left[t \left(1 - \frac{A_2\left(t, \Omega\right)}{2n} \right) - b_2 \left(1 - \frac{B_2\left(t, \Omega, b_2\right)}{2n} \right) \right]$$

where

$$A_2\left(t, \Omega\right) = \frac{1}{4}\left(1 + t^2\right) \frac{\operatorname{tr} V_\Omega \left(x_\Delta x_\Delta' \otimes x_\Delta x_\Delta'\right)}{\left(x_\Delta' \Omega x_\Delta\right)^2} \tag{2.7}$$

and

$$B_2\left(t, \Omega, b_2\right) = \frac{1}{4} t^2 \frac{\operatorname{tr} V_\Omega \left(x_\Delta x_\Delta' \otimes x_\Delta x_\Delta'\right)}{\left(x_\Delta' \Omega x_\Delta\right)^2}. \tag{2.8}$$

Size corrected tests can be constructed in the same way as before.

3. A Special Case: Same Treatment Date for all States for Which Treatment Occurs

This case implies additional structure for the regression equation that can be exploited to simplify the test and size corrections. We assume that if treatment occurs in state i it is at a fixed time τ which is known to the investigator. We assume a simplified version of the model

$$y_{it} = \alpha_i + \beta_t + T_{it}\gamma + \varepsilon_{it} \tag{3.1}$$

where α_i are individual specific fixed effects, β_t is a time effect common to all states but changing over time and T_{it} is the treatment indicator where $T_{it} = 0$ for $t < \tau$ and all i and T_{it} takes values in $\{0, 1\}$. We also assume that once treatment takes effect in state i and at time τ it remains in effect. Formally, this means that $T_{i\tau} = 1$ implies that $T_{it} = 1$ and $T_{i\tau} = 0$ implies $T_{it} = 0$ for all $t > \tau$. The innovations ε_{it}, when stacked as $\tilde{\varepsilon} = \left[\varepsilon_{11}, ..., \varepsilon_{n,1}, ..., \varepsilon_{1,T}, ..., \varepsilon_{n,T}\right]'$, satisfy $E\tilde{\varepsilon} = 0$ and $E\tilde{\varepsilon}\tilde{\varepsilon}' = \tilde{\Sigma} \otimes I_n$. We again assume that there are observable variables V_i such that $E\left[\alpha_i | V_i\right] = V_i m_\alpha$ and $\operatorname{Var}\left[\alpha_i | V_i\right] = \sigma_\alpha^2$. For simplicity we assume that V_i only contains $T_{i,1}, ..., T_{i,T}$ and a constant. We thus consider the transformed model

$$\left(BM_{\mathbf{1}_T} \otimes I_n\right) \tilde{Y} = \textstyle\sum_{t=1}^T \left(BM_{\mathbf{1}_T} a_{t,T} \otimes \mathbf{1}_n\right) \beta_t + \left(BM_{\mathbf{1}_T} \otimes I_n\right) \tilde{\Upsilon} \gamma + \left(BM_{\mathbf{1}_T} \otimes I_n\right) \tilde{\varepsilon}$$

and for the transformed model we let $Y = (C'BM_{\mathbf{1}_T} \otimes I_n)\tilde{Y}, Z = (C'BM_{\mathbf{1}_T}B'C \otimes \mathbf{1}_n), \Upsilon = (C'BM_{\mathbf{1}_T} \otimes I_n)\tilde{\Upsilon}, \varepsilon = (C'BM_{\mathbf{1}_T} \otimes I_n)\tilde{\varepsilon}$ and β a $r \times 1$ vector. The transformed model then takes the form

$$Y = (I_r \otimes \mathbf{1}_n)\beta + \Upsilon\gamma + \varepsilon.$$

The properties of the tests T_1 and T_2 are again determined by the functions A_1, B_1, A_2 and B_2. We derive these functions in the Appendix. For the test T_1 we find

$$A_1(t) = \frac{1}{2}\left(1 + t^2\right) + 2(r-1)$$

and

$$B_1(t) = \frac{1}{2}t^2 + r - 1$$

which turn out to be the same as for the first difference version of the test. Using the result for B_1 it now follows that the power of the test can be approximated by

$$\frac{\gamma - \gamma_0}{\left(\sigma^{\tau\tau}\sum_{i=1}^n \bar{x}_{i\tau}^2\right)^{-1/2}}\left(1 - \frac{1}{4n}t^2 - \frac{r-1}{2n}\right) \tag{3.2}$$

where

$$\sigma^{\tau\tau} = \xi_\tau' M_{\mathbf{1}_T} B'C\Sigma^{-1}C'BM_{\mathbf{1}_T}\xi_\tau'$$

and $\xi_\tau = \sum_{t=\tau}^T a_{t,T}$. Note that $\Sigma = C'BM_{\mathbf{1}_T}\tilde{\Sigma}M_{\mathbf{1}_T}B'C$.

We can exploit the simple structure of (3.2) to analyze the question of optimal aggregation. In our framework this amounts to choosing C optimally to maximize (3.2). The expression for approximate power shows that there is a first order effect on power determined by the efficiency of the estimator as captured by $\sigma^{\tau\tau}$. Estimation error of the elements in the optimal weight matrix affects power to order n^{-1} through the term $(r-1)/2n$. An algorithm for maximizing (3.2) thus consists in choosing C optimally for r fixed and then choosing the overall optimal $r \in (1, ..., T-1)$. Thus, for any given r, C is chosen such that

$$C_r^* = \underset{C \text{ s.t. } C'C=I_r}{\arg\max}\ \sigma^{\tau\tau} = \arg\max_C \xi_\tau' M_{\mathbf{1}_T} B'C\Sigma^{-1}C'BM_{\mathbf{1}_T}\xi_\tau'.$$

Note that $\tilde{\Sigma}^{1/2}M_{\mathbf{1}_T}B'C\Sigma^{-1}C'BM_{\mathbf{1}_T}\tilde{\Sigma}^{1/2}$ is a projection matrix. For $r = 1$ it thus follows that $\sigma_\eta^{\tau\tau}$ is maximized by minimizing $\left\|\tilde{\Sigma}^{-1/2}\xi_\tau - \tilde{\Sigma}^{1/2}M_{\mathbf{1}_T}B'C\right\|$ or equivalently $\left\|\xi_\tau - \tilde{\Sigma}M_{\mathbf{1}_T}B'C\right\|$.

This is achieved for $C_1^* = (BM_{\mathbf{1}_T}\tilde{\Sigma}M_{\mathbf{1}_T}B')^{-1}BM_{\mathbf{1}_T}\xi_\tau$. Since the projection residual is equal to zero it follows that $\sigma^{\tau\tau}$ cannot be increased further by choosing any $r > 1$. Hence the overall optimum of 3.2 is given by $r^* = 1$ and $C^* = (BM_{\mathbf{1}_T}\tilde{\Sigma}M_{\mathbf{1}_T}B')^{-1}BM_{\mathbf{1}_T}\xi_\tau$. Also note that $\sigma^{\tau\tau}$ is invariant under transformations $C^\ddagger = CO_r$ for any orthogonal matrix O_r. Solutions to the maximization problem are therefore unique subject to $C'C = I_r$ only.

This result shows that in general optimal aggregation is infeasible. The optimal matrix C depends on the unknown covariance matrix $\tilde{\Sigma}$.

A special case occurs when $\tilde{\Sigma} = I_T$. As it turns out this case is of particular interest for the discussion regarding difference in difference regressions. Firstly, note that for this case OLS and GLS are equivalent. However, an investigator may not be willing to assume that $\tilde{\Sigma}$ is known and instead still estimate $\tilde{\Sigma}$ in an unrestricted way and use a size corrected test based on GLS. Calculations of C^* for this case then reveal that $C^* = -[\mathbf{1}'_\tau, \mathbf{0}'_{T-1-\tau}]'$ which implies that

$$C^{*\prime}BM_{\mathbf{1}_T} = \left[-\frac{\tau}{T}\mathbf{1}'_\tau, \frac{T-\tau}{T}\mathbf{1}'_{T-\tau}\right].$$

In other words, optimal aggregation leads to the 'classical' difference in difference estimator where pre and post treatment periods are averaged and the difference between them is tested for a significant effect. A consequence of our analysis is then that this procedure is not optimal in terms of power when $\tilde{\Sigma}$ is not the identity matrix.

This result helps explain some findings in our Monte Carlo experiments where time aggregation methods to correct size distortions lead to a significant loss in power when serial correlation in ε is high but have little effect on power when serial correlation is low. Because of the large power loss when $\tilde{\Sigma} \neq I_T$ the 'classical' difference in difference approach thus can only be recommended if it is known on a priori grounds that $\tilde{\Sigma} = I_T$ holds.

For the test T_2 we find that

$$A_2(t) = \frac{1}{2}\left(1 + t^2\right)$$

and

$$B_2(t) = \frac{1}{2}t^2$$

It thus turns out that robust OLS is unaffected by the choice of dimension r as far as the second order terms are concerned.

We now turn to the first difference specification of (3.1)

$$\Delta y_{it} = \beta_t - \beta_{t-1} + \Delta T_{it}\gamma + \Delta\varepsilon_{it}.$$

Note that $\Delta T_{it} = 0$ except for $t = \tau$. The regressor matrix $\tilde{U}$ for this particular case takes on the form

$$\tilde{U} = \begin{bmatrix} \mathbf{1}_n & & 0 & 0 \\ & \ddots & & \tilde{\Upsilon}^{\Delta}_{\tau} \\ 0 & & \mathbf{1}_n & 0 \end{bmatrix}$$

where $\tilde{\Upsilon}^{\Delta}_{\tau} = [\Delta T_{1\tau}, ..., \Delta T_{n\tau}]'$ and $\mathbf{1}_n$ is a vector of ones with length n. Using the notation $a_{t,T}$ for the t-th unit vector of length $T-1$ we can define $Z^{\Delta} = \sum_{t=2}^{T}\left(C'a_{t,T}a'_{t,T}C \otimes \mathbf{1}_n\right) = I_r \otimes \mathbf{1}_n$, $Y^{\Delta} = \sum_{t=2}^{T}\left(C'a_{t,T} \otimes \tilde{Y}^{\Delta}_t\right)$ and $\Upsilon^{\Delta} = \left(C'a_{\tau,T} \otimes \tilde{\Upsilon}^{\Delta}_{\tau}\right)$. Also let $\sigma^{t,s} = a'_{t,T}C\Sigma^{-1}C'a_{s,T}$, $\sigma_{t,s} = a'_{t,T}C\Sigma C'a_{s,T}$ and $A_{t,s} = a'_{t,T}CC'a_{s,T}$ with corresponding expressions for $\hat{\sigma}^{t,s}$ and $\hat{\sigma}_{t,s}$ by replacing Σ by $\hat{\Sigma}$. Then,

$$\gamma_{GLS} = \left(\sigma^{\tau\tau}\tilde{\Upsilon}^{\Delta\prime}_{\tau}M_{\mathbf{1}_n}\tilde{\Upsilon}^{\Delta}_{\tau}\right)^{-1}\left(\textstyle\sum_{t=2}^{T}\sigma^{\tau t}\tilde{\Upsilon}^{\Delta\prime}_{\tau}M_{\mathbf{1}_n}\tilde{Y}^{\Delta}_t\right)$$

with $M_{\mathbf{1}_n} = I_n - \frac{\mathbf{1}_n\mathbf{1}'_n}{n}$. Conditional on X_τ, the variance of γ_{GLS} then is $\left(\sigma^{\tau\tau}\tilde{\Upsilon}^{\Delta\prime}_{\tau}M_{\mathbf{1}_n}\tilde{\Upsilon}^{\Delta}_{\tau}\right)^{-1}$.

The OLS estimator for this case is

$$\gamma_{OLS} = \left(A_{\tau\tau}\tilde{\Upsilon}^{\Delta\prime}_{\tau}M_{\mathbf{1}_n}\tilde{\Upsilon}^{\Delta}_{\tau}\right)^{-1}\textstyle\sum_t A_{\tau t}\tilde{\Upsilon}^{\Delta\prime}_{\tau}M_{\mathbf{1}_n}\tilde{Y}^{\Delta}_t$$

with conditional variance equal to $\sigma_{\tau\tau}A^{-2}_{\tau\tau}\left(\tilde{\Upsilon}^{\Delta\prime}_{\tau}M_{\mathbf{1}_n}\tilde{\Upsilon}^{\Delta}_{\tau}\right)^{-1}$. Based on these results the tests T_1 and T_2 specialize to

$$T_{1,\Delta} = \frac{\left(\hat{\sigma}^{\tau\tau}\tilde{\Upsilon}^{\Delta\prime}_{\tau}M_{\mathbf{1}_n}\tilde{\Upsilon}^{\Delta}_{\tau}\right)^{-1}\left(\sum_t \hat{\sigma}^{\tau t}\tilde{\Upsilon}^{\Delta\prime}_{\tau}M_{\mathbf{1}_n}\tilde{Y}^{\Delta}_t\right) - \gamma_0}{\left(\hat{\sigma}^{\tau\tau}\tilde{\Upsilon}^{\Delta\prime}_{\tau}M_{\mathbf{1}_n}\tilde{\Upsilon}^{\Delta}_{\tau}\right)^{-1/2}}$$

and

$$T_{2,\Delta} = \frac{\left(A_{\tau\tau}\tilde{\Upsilon}^{\Delta\prime}_{\tau}M_{\mathbf{1}_n}\tilde{\Upsilon}^{\Delta}_{\tau}\right)^{-1}\sum_t A_{\tau t}\tilde{\Upsilon}^{\Delta\prime}_{\tau}M_{\mathbf{1}_n}\tilde{Y}^{\Delta}_t - \gamma_0}{\hat{\sigma}^{1/2}_{\tau\tau}\left(A^2_{\tau\tau}\tilde{\Upsilon}^{\Delta\prime}_{\tau}M_{\mathbf{1}_n}\tilde{\Upsilon}^{\Delta}_{\tau}\right)^{-1/2}}.$$

It follows from arguments contained in the Appendix that

$$A_1(t) = \frac{1}{2}\left(1+t^2\right) + 2(r-1)$$

and

$$B_1(t) = \frac{1}{2}t^2 + (r-1).$$

These results hold both unconditionally and conditionally in X. They show how higher order power is affected by the design matrix C through the dimension of C and that higher order size is expected to deteriorate with more time periods as measured by the parameter r. Power up to order n^{-1} then depends on

$$\frac{\gamma - \gamma_0}{\left(\sigma^{\tau\tau}\sum_{i=1}^{n}\bar{x}_{i\tau}^2\right)^{-1/2}}\left(1 - \frac{1}{4n}t^2 - \frac{r-1}{2n}\right) \tag{3.3}$$

with $\sigma^{\tau\tau} = a'_{\tau,T}C(C'\tilde{\Sigma}C)^{-1}C'a_{\tau,T}$. For $r = 1$ the optimal C_1^* is found by minimizing $\left\|\tilde{\Sigma}^{-1/2}a_{\tau,T} - \tilde{\Sigma}^{1/2}C\right\|$ which leads to $C_1^* = \tilde{\Sigma}^{-1}a_{\tau,T}$. Since the residual from this projection is identical to zero it follows that $\sigma^{\tau\tau}$ can not be further increased by increasing r. The overall optimum of 3.3 is then found by choosing $r^* = 1$ and $C_1^* = \tilde{\Sigma}^{-1}a_{\tau,T}$. It should be noted that the optimal choice of C_1^* is the same as the infeasible GLS estimator. Since C_1^* depends on unknown parameters implementing this approach leads back to feasible GLS on the full sample with all time periods. In other words, as we have seen for the level case before, temporal aggregation is only recommended under special circumstances where $\tilde{\Sigma}$ is known a priori.

For the test T_2 we show in the Appendix that

$$A_2(t,\Omega) = \frac{1}{2}\left(1+t^2\right)$$

and

$$B_2(t,\Omega,b_2) = \frac{1}{2}t^2.$$

The power of the OLS based test is therefore again independent of the temporal aggregation as far as second order terms are concerned. Moreover, size is not affected by the number of time periods which suggest that temporal aggregation is not justified for the robust OLS based tests as far as achieving correct size is concerned. The power function can be approximated by

$$\frac{\gamma-\gamma_0}{\sigma_{\tau\tau}^{1/2}\left(A_{\tau\tau}^2\sum_{i=1}^n\bar{x}_{i\tau}^2\right)^{-1/2}}\left(1-\frac{1}{8n}t^2\right)$$

which is dominated by the power curve of the GLS based test. Size corrections can again be based on $A_2(t,\Omega)$.

4. Monte Carlo

We first consider some Monte Carlo evidence on the performance of our approach and the second order Edgeworth approximations. We consider a situation with positive serial correlation across time for states, which is the usual situation found in applied research. So far in our empirical research, we have considered the single treatment date situation. Our Monte Carlo design uses N=50 and T = (5,10, 15, 20) and the first order serial correlation, $\rho=[0,0.4,0.8,0.9]$.

In order to asses the different procedures numerically we now make more specific assumptions about the generating process. We assume that

$$\begin{aligned}\varepsilon_{it} &= \rho\varepsilon_{it-1}+u_{it}\\ &\quad\varepsilon_{i0}\tilde{}N\left(0,\frac{1}{1-\rho^2}\right)\end{aligned}$$

where $u_{it}\tilde{}N(0,1)$ is iid both across i and t. We assume that $\alpha_i\tilde{}N(0,1)$, $\beta_t\tilde{}N(0,1)$ and generate

$$y_{it}=\alpha_i+\beta_t+T_{it}\gamma+\varepsilon_{it} \tag{4.1}$$

where the treatment T_{it} is drawn in a two stage process. First we draw treated states i with probability p. Then we draw a common treatment time τ randomly from $[T/4],...,T-[T/4]$ where $[a]$ is the largest integer smaller than a. Then $T_{it}=1$ if $t\geq\tau$ and i is a selected state and $T_{it}=0$ otherwise. We generate 500 random samples for parameter values $\gamma=[0,.1,.6,1]$ and $\rho=[0,0.4,0.8,0.9]$. Note that α_i and β_t are drawn before we generate the 500 Monte Carlo samples, ie. they are fixed parameters for all the Monte Carlo samples.

In Tables 1a-1d we find that size corrected FGLS, FGLS-SC, in levels does almost as well as GLS with Σ known in terms of power. However, for T = 20 we find that the size correction based on the second order Edgeworth expansion does not completely solve the size distortion problem when serial correlation is very high. For $\rho = .9$ the actual size is 0.088 when the nominal size is 0.05. While this amounts to a small size distortion, overall the performance of the size correction is remarkable even when $T = 20$. For the T = 20 case the number of unknown elements of $\hat{\Sigma}$ is 210 which is over 20% as large as the total number of observations. Apparently, such a large number of unknown parameters causes a slight inaccuracy of the Edgeworth approximation when ρ is close to one. When T = 10 so that the number of unknown elements of $\hat{\Sigma}$ is 55 which is 11% as large as the total number of observations, the size correction is very accurate for all values of ρ we consider. We also find that FGLS-SC has significantly more power than does OLS with a robustly estimated covariance matrix, which we call robust OLS[5]. For example, in the situation of $\rho = 0.9$ in Table 1d, FGLS-SC often has almost 2 times more power than robust OLS for the cases of T = 10 or 15. Thus FGLS appears to be the better estimator even with additional parameter uncertainty created by the estimated $\hat{\Sigma}$. In summary, we do recommend FGLS-SC even for "large $\hat{\Sigma}$" because the remaining size distortion is negligible.

However, also note that in Table 1d that non-robust OLS on the entire sample has an actual size level that exceeds 0.25 for when T = 10, 15, and 20 although the nominal size level is only 0.05. Thus, as the previous literature found, OLS cannot be used without a correction to the estimated variance matrix of the estimates or severe size distortions may result[6].

We also consider two other versions of FGLS for 3 periods (before, change period, and after) and the "traditional" 2 periods (before and after) DID approach. We find that both of these alternative approaches involving time aggregation have significantly reduced power compared to FGLS-SC. We find that the power of FGLS-SC on the full sample is often 50%-100% higher than the 2 or 3 period time aggregated version. Thus, we do not recommend their use. The traditional DID approach loses too much power to solve the problem of a consistent estimate of

[5] Robust OLS is the estimator studied by Bertrand et. al. (2002).

[6] Significant size distortion for OLS also occur when $\rho = 0.4$ although they are not as severe.

the variance of the estimated parameters.

We next consider in Tables 2a-2d a first difference specification that also eliminates the fixed effects but can also lead to a reduced effect of the positive serial correlation. Note that because of the way we estimate the covariance matrix for both the level and the first difference specification the two are numerically identical for the full sample specification. This is because $BM_{\mathbf{1}_T}$ and B^{Δ} map into the same $T-1$ dimensional subspace and on that subspace the tests are invariant to orthogonal rotations.

We thus only consider 3 period and 2 period time aggregation estimators[7]. The middle period of the 3 period specification of time aggregation has the first difference of the single time period when the treatment occurs. The treatment effect parameter appears only in this period because first difference eliminates it in all other periods. However, the before and after periods still lead to an efficiency improvement in FGLS estimation because of the correlation of the stochastic disturbances. We find that all size distortions have been eliminated in FGLS-SC. We also find that FGLS does approximately as well as non-size corrected GLS, because the size corrections are now quite small.

We also find that the 3 period version of FGLS outperforms the 2 period version by a large amount. Indeed, the 3 period aggregation FGLS-SC estimator seems to do the best of all the feasible GLS estimators considered with correct size and maximum power. Nevertheless, a word of caution with regard to the first difference transformation is in place. If the effect on the treatment group occurs only with a time lag after the policy change then the 3 period version of the first difference specification is not expected to have as much power and the level specification is likely to be preferred in terms of power. Considering that FGLS-SC performs well for the level specification in terms of power and size we tend to recommend its use over the three period first difference specification.

Recently Hansen (2004) proposed an alternative solution to the size problem of GLS based hypothesis tests. He proposes to fit a parametric model, in his case an AR(p) model, to the

[7]Since the data have been initially transformed to first differences, these estimators differ from the earlier fixed effects estimators on 2 or 3 periods.

serial correlation process of ε_{it}. With these parametric estimates an estimate of $\tilde{\Sigma}$ is constructed. Since the work of Nickell (1981) it is well known that parametric estimates of serial correlation are inconsistent in the presence of individual specific fixed effects. In order to implement a parametric estimator of $\tilde{\Sigma}$ a bias correction of the parameter estimates is therefore needed. Based on Hahn and Kuersteiner's (2002) method of bias correction Hansen (2004) uses a formula of Nickell (1981) to correct for bias. We replicate this procedure by computing the residuals $\hat{\varepsilon} = M_{\tilde{x}}\tilde{Y}$ where $M_{\tilde{x}}$ is the projection onto the orthogonal space spanned by $\tilde{X} = \left[\tilde{Z}, \tilde{\Upsilon}\right]$. We then fit a panel AR(1) model to $\hat{\varepsilon}_{it} = \rho\hat{\varepsilon}_{it-1} + \hat{\eta}_{it}$. Since $\hat{\varepsilon}_{it}$ is already demeaned from $M_{\tilde{x}}\tilde{Y}$ this panel AR(1) estimator is essentially the within estimator of the model $y_{it} = \alpha_i + \rho y_{it-1} + z_{it}'\theta - \rho z_{it-1}'\theta + \eta_{it}$ where η_{it} is iid if ε_{it} is an AR(1) process. Once we obtain an estimate $\hat{\rho}$ we subtract the estimated bias and form an estimate $\tilde{\Sigma}_{\hat{\rho}}$ of $\tilde{\Sigma}$ based on the assumption that ε_{it} indeed follows an AR(1) process. We then construct $\hat{\Sigma} = C'BM_{\mathbf{1}_T}\tilde{\Sigma}_{\hat{\rho}}M_{\mathbf{1}_T}B'C$ or $\hat{\Sigma} = C'B^{\Delta}\tilde{\Sigma}_{\hat{\rho}}B^{\Delta\prime}C$ depending on whether the model is estimated in first differences or in levels. Note that we estimate $\hat{\rho}$ on the full sample even if $r = 2$ or 3 because it was shown by Hahn and Kuersteiner (2002) that the performance of the bias correction improves with larger T. For the first difference formulation of the model we could also estimate the serial correlation coefficient using a GMM procedure. Because first differencing induces an endogeneity bias, GMM tends to be heavily biased especially when serial correlation is large. For a discussion of these problems see Hahn, Hausman and Kuersteiner (2000) and Hahn and Kuersteiner (2002). For these reasons we do not consider GMM estimators here.

A potential problem of the parametric estimator for $\tilde{\Sigma}$ lies in its stationarity and functional form assumption. If the parametric model is misspecified the resulting GLS estimator is inefficient and the estimated standard errors generally are incorrect. In Monte Carlo simulations not reported here we found that if ε_{i0} is not drawn from its stationary distribution, the performance of the parametric covariance matrix estimator can be quite poor. It also suffers from the disadvantage that for small T the parametric estimator tends to be more biased and thus has inferior small sample behavior even when the model is stationary.

Monte Carlo designs where only the initial observation is not drawn from the stationary distribution tend to be quite artificial and have the disadvantage that the form of the non-stationarity mostly affects observations at the beginning of the sample. In order to have a more realistic design we estimate a simple treatment model for the dataset on cellular telephone service prices analyzed in more detail in Section 5. We estimate the variance covariance matrix of the residuals from that model. Denote the estimated variance covariance matrix by $\hat{\Omega}$. The sample size is $T = 11$. We then draw $\varepsilon_i\tilde{\ }N(0, \hat{\Omega})$ for $\varepsilon_i = (\varepsilon_{i1}, ..., \varepsilon_{i,11})'$. The variables α_i, β_t and the treatment indicators T_{it} are generated as before where now the treatment time τ is fixed at $\tau = 8$ to coincide with the change in regulation of the actual sample. The outcome variable y_{it} is then defined as in 4.1 where we again vary the size of the treatment effect measured by $\gamma = [0, .1, .6, 1]$. Results for 500 Monte Carlo replications on samples of size $n = 50$ and $T = 11$ are reported in Table 3. It is striking that under the more realistic correlation patterns for the residuals which are not well approximated by an AR(1) or for that matter any stationary parametric time series model, the parametric estimator (GLS-AR) performs quite poorly with size distortions in the range of 20%. This contrasts with the unrestricted covariance matrix estimator used in GLS with a size correction. This procedure continues to have approximately correct size as it did previously in stationary designs. Thus, imposing a stationary AR(p) specification in a non-stationary situation may not solve the problem of obtaining the correct size of tests, which is the most important problem for applied research of policy evaluation.

5. Effect of Regulation on Cellular Telephone Service Prices

In the U.S. for the first 12 years of operation, 1983-1995, cellular telephone operated as a duopoly. However, the two facilities-based carriers were required to sell cellular airtime to resellers who also sold cellular service to consumers. In the U.S. each of 51 state regulatory commissions decided on whether to regulate cellular prices or to use market outcomes.[8] In an interesting natural experiment 26 states regulated cellular prices, while the other 25 did not. In Table 5.1 we list monthly service prices in 1994 for the least expensive plan for average usage of

[8]In the U.S. the District of Columbia acts as the 51st state.

Table 5.1: Average Cellular Prices in the Top 10 MSAs: 1994
160 minutes of use (80% peak)

MSA No.	MSA	Monthly Price	Regulated
1.	New York	$110.77	Yes
2.	Los Angeles	99.99	Yes
3.	Chicago	58.82	
4.	Philadelphia	80.98	
5.	Detroit	66.76	
6.	Dallas	59.78	
7.	Boston	82.16	Yes
8.	Washington	76.89	
9.	San Francisco	99.47	Yes
10.	Houston	80.33	

160 minutes per month (80% peak)[9] for up to a 1-year contract in the 10 largest MSAs, which are the metropolitan areas where cellular licenses were granted.[10]

The fact that price regulation goes along with higher monthly service prices is evident from Table 5.1. Every regulated price in Table 5.1 is greater than every unregulated price in Table 5.1. The average price of regulated MSAs is $98.10 while the average price of unregulated MSAs is $70.59, which is a difference of $27.51 per month or 39%.

Table 5.1 demonstrates that price regulation of cellular telephone was associated with higher prices for consumers in the U.S. However, other factors such as higher costs in the regulated states could be the reason for the higher prices. Hausman (1995) used a cross-section approach to quantify the higher prices that consumers pay in regulated states. He specified a model of cellular prices in the top 30 MSAs where the right had side variable included MSA pop-

[9]This usage, 160 minutes per month, was the approximate average usage of cellular customers in 1994.

[10]While in most other countries national cellular licenses were granted, the US has followed the framework of granting licenses on a significant disaggregated geographical level.

ulation, average commuting time, average MSA income, and an index of constructions costs.[11] These top 30 MSAs contain about 41% of the entire U.S. population and about 60% of cellular subscribers in 1994. Hausman treated price regulation as a jointly endogenous variable and used instrumental variables in estimation.[12] The estimated coefficient of the price regulation variable is 0.149, which means that regulated states had cellular prices that are 15% higher, holding other economic factors equal. The coefficient is estimated very precisely (standard error = 0.052) and the finding is highly statistically significant (t statistic = 2.87). Thus, states that regulate had significantly higher cellular prices in large MSAs.

To explore this issue further Hausman also collected data from cellular companies for the years 1989-93 and run a similar regression. Over this time period price regulation led to a higher price of 14.2% which is again estimated quite precisely (standard error = .029) and is very statistically significant (t statistic = 4.9). Thus, the results of the effect of price regulation are very similar for the period 1989-93 and for the single year 1994. However, these estimates could be objected to (and were objected to by defenders of price regulation) on the grounds that unmeasured variables led to higher prices in the regulatory states. Since the regulatory status of the states did not change over time, this possible objection was untestable.

However, a "natural experiment" occurred that allowed a further test of the regulatory hypothesis. In 1993 U.S. Congress instructed the Federal Communications Commission (FCC) to deregulate cellular prices unless a given state that was regulating cellular prices could show price regulation was "necessary".[13] Eight states petitioned the FCC to continue price regulation, and the FCC turned them down in late 1994. One state appealed, but regulation completely ended in 1995. Thus, Congress and the FCC provided a natural experiment that permitted an analysis of how cellular prices changed in the regulated and unregulated states, after price regulation was prohibited.

[11] See Hausman (1995)

[12] The instruments were state tax rates and whether the state regulated paging prices. By 1994 paging had numerous competitors in each MSAs and no economic reason existed to regulate paging prices.

[13] In the U.S. a dual regulatory framework exists where the FCC, at the national level, and each state has regulatory authority over telecommunications. However, each state regulatory body must implement FCC rules.

A complicating factor arose because cellular prices decreased significantly in 1995-96 both because of new PCS entry and because of deregulation.[14] Thus, the econometric specification, as in equation (1.1) has a fixed effect for each MSA, and a time effect for each year, which allows for the effect of new entry. A single indicator variable allows for the effect of price regulation. The econometric specification was estimated over 11 years of data with 7 years prior to the end of price regulation and 4 years after the end of regulation. Given the 30 MSAs we have a total of 330 observations.

First, we estimate the model by the traditional "differences in differences" OLS approach. That is, we average across all observations for a given MSA during the regulatory period and also average across observations for the post-regulatory period and compare the change in average price for regulated MSAs to the non-regulated MSAs after price regulation ended. The point estimate is 0.180, consistent with the earlier estimates that regulation led to higher prices for consumers. However, the estimated t-statistic is 1.35, which is not significant at usual test levels. When OLS is run on the complete sample so that T = 11, the estimated OLS t-statistic is 2.11, which would indicate statistical significance. However, the estimated robust t-statistic that allows for a non-diagonal covariance matrix is 1.65, which again indicates a lack of statistical significance. Thus, if OLS is used on the complete sample the effect of non-independence across periods can affect inference in important ways.

We now use FGLS on the entire sample. We allow for an unrestricted covariance matrix and estimate it using an unbiased estimator. The GLS point estimate is 0.150, which is very close to the 0.149 estimate from the original cross section specification from 1994 before price regulation was prohibited. The conventional first-order GLS t-statistic that does not account for estimation of the covariance matrix is 3.68. However, the second order approximation that accounts for estimation of the covariance matrix, is 2.67, which yields a p-value of 0.996 indicating a highly significant result. Thus, the "natural experiment" of the end of price regulation demonstrates the effect of regulation on prices, and the result is less subject to criticism of omitted or unmeasured

[14] PCS is a "second generation" cellular technology. The FCC auctioned off additional spectrum, which permitted entry of additional cellular service providers. Hausman (2002) discusses the new entry in greater detail.

variables. Taking account of the estimated covariance matrix is also important and has an important effect on the estimated precision of the estimator.

We next consider FGLS on the 2 period specification, where FGLS accounts for correlation across periods rather than taking an unweighted average across periods as does the difference in differences approach. The 2 period FGLS estimate is 0.140 with the estimated t-statistic of 1.72, which is greater than the difference in differences t-statistic but is still below convention significant levels. We lastly consider FGLS on a 3 period specification, where the periods are during price regulation, the year of the change, and the period follow regulation. The FGLS point estimate is 0.160 with an estimated t-statistic of 1.94, slightly below conventional level of statistical significance. Thus, in this application GLS on the entire sample appears to be the best estimator. However, using different "cuts" of the data permit additional estimates, which allow for specification tests following Hausman (1978). The specification tests do not reject the orthogonality of the econometric specification, as expected given the rather close point estimates using the three different approaches. The economic conclusion is that state regulators, by attempting to protect cellular resellers from competition by the two facilities based carriers, led to significantly higher prices to consumers.

6. Conclusions

We derive higher order expansions of the distribution of the t-statistic for the significance of treatment variables in difference in difference regressions. When serial correlation in the errors is present, standard OLS based inference leads to tests with distorted size. Robust OLS does not suffer from this problem and is shown to be immune to a dimension problem when N, the number of cross-sectional units, is small relative to the number of time periods. A more efficient procedure is GLS. Our expansions show, that unlike robust OLS, feasible GLS does suffer from a many parameter problem and exhibits severe small sample size distortions when N is not large enough. Using our expansions we obtain a size correction for FGLS.

We find that size corrected FGLS, FGLS-SC, in levels does almost as well as GLS with known

Σ. We also find that FGLS-SC has significantly more power than does OLS with a robustly estimated covariance matrix, robust OLS, when serial correlation in the level data is high. Thus FGLS appears to be the better estimator even with additional parameter uncertainty created by the estimated $\hat{\Sigma}$. We also consider two other versions of FGLS for 3 periods (before, change period, and after) and the "traditional" 2 period (before and after) DID approach. We find that both of these alternative approaches involving time aggregated have significantly reduced power compared to FGLS-SC. Thus, we do not recommend their use.

The first difference specification also eliminates the fixed effects but can also lead to a reduced effect of the positive serial correlation. We now find that FGLS-SC does almost as well as GLS. We then consider 3 period and 2 period time aggregation estimators. We find that all size distortions have been eliminated in FGLS-SC. We also find that the 3 period version of FGLS outperforms the 2 period version by a large amount. Unlike in the case of the level specification, the loss of power for the 3 period version of feasible GLS using first differenced data is negligible.

These results suggest to use full sample FGLS-SC whenever serial correlation is high in levels. If the regressions are run in first differences the 3 period version of FGLS-SC seems to perform best. An argument for running the specification in levels can be made in cases where adjustment to the new policy takes more than one time period. In this case, the first difference specification will underestimate the total effect of the policy relative to the level specification.

References

Bertrand, M., E. Duflo, and S. Mullainathan (2002): "How Much Should We Trust Difference-in-Difference Estimates?," *mimeo*.

Hahn, J., J. Hausman, and G. M. Kuersteiner (2000): "Bias Corrected Instrumental Variables Estimation for Dynamic Panel Models with Fixed Effects," *unpublished manuscript*.

Hahn, J., and G. Kuersteiner (2002): "Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects," *Econometrica*, 70, 1639–1657.

Hansen, C. (2004): "Generalized Least Squares Estimation In Difference In Difference And Other Panel Models," *mimeo*.

Hausman, J. (1978): "Specification tests in econometrics," *Econometrica*, pp. 1251–71.

Hausman, J. (1995): "State Regulation of Cellular Prices," *Wireless Communications Forum*, 3.

——— (2002): "Mobile Telephone," in *Handbook of Telecommunications Economics*, ed. by M. Cave. North Holland.

Kiefer, N. M. (1980): "Estimation of Fixed Effect Models for Time Series of Cross-Sections with Arbitrary Intertemporal Covariance," *Journal of Econometrics*, 14, 195–202.

Magnus, J. R., and H. Neudecker (1979): "The Commutation Matrix: Some Properties and Applications," *Annals of Statistics*, 7(2), 381–394.

Magnus, J. R., and H. Neudecker (1988): *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley and Sons.

Moulton, B. R. (1986): "Random Group Effects and the Precision of Regression Estimates," *Journal of Econometrics*, pp. 385–397.

MUNDLAK, Y. (1978): "On the Pooling of Time Series and Cross Section Data," *Econometrica*, pp. 69–85.

NICKELL, S. (1981): "Biases In Dynamic Models with Fixed Effects," *Econometrica*, 49, 1417–1426.

ROTHENBERG, T. J. (1984a): "Approximate Normality of Generalized Least Squares Estimates," *Econometrica*, 52, 811–825.

——— (1984b): "Hypothesis Testing in Linear Models when the Error Covariance Matrix is Nonscalar," *Econometrica*, 52, 827–842.

——— (1988): "Approximate Power Functions for Some Robust Tests of Regression Coefficients," *Econometrica*, pp. 997–1019.

A. Appendix

A.1. Unbiased Weight Matrix Estimation

In this appendix we show that (2.1) is approximately unbiased. Let $\hat{\varepsilon}_{it}$ be the residual from a regression of y_{it} onto T_{is} and all elements of Z_{is}, $s = 1, ..., T$ and define $V_{it} = [T_{it}, Z'_{it}]$ with regressor matrix $V = [V'_1, ..., V'_n]'$ where $V_i = [V_{i1}, ..., V_{iT}]$ such that

$$\hat{\varepsilon}_{it} = \varepsilon_{it} - V_i \left(V'V\right)^{-1} V'\varepsilon_t + \alpha_i - V_i \left(V'V\right)^{-1} V'\alpha$$

and we note that

$$E\left[\hat{\varepsilon}_{it}|V\right] = 0.$$

Let $y_t = [y_{1t}, ..., y_{nT}]'$ and $y_t = \alpha + \tilde{Z}_t\theta_t + \Upsilon_t\gamma + \varepsilon_t$ where $\alpha, \theta_t, \varepsilon_t$ are defined in the obvious way. Now consider

$$\begin{aligned} E\left[\hat{\sigma}_{t,s}|\mathcal{I}_{t,s}\right] &= E\left[\frac{y'_t M_V y_s}{\operatorname{tr}(M_V)}|V\right] = \frac{E\left[\sum_{i=1}^n \hat{\varepsilon}_{it}\hat{\varepsilon}_{is}|V\right]}{\operatorname{tr}(M_V)} \\ &= \frac{\operatorname{tr} M_V E\left[\left(\alpha+\varepsilon_s\right)\left(\alpha+\varepsilon_t\right)'|V\right] M_V}{\operatorname{tr}(M_V)} \\ &= \frac{\operatorname{tr} M_V E\left[\alpha\alpha'|V\right] M_V}{\operatorname{tr}(M_V)} + \tilde{\sigma}_{t,s} \end{aligned}$$

where

$$E\left[\alpha\alpha'|V\right] = \sigma^2_\alpha I_n + E\left[\alpha|V\right] E\left[\alpha'|V\right] = \sigma^2_\alpha I_n + V m_\alpha m'_\alpha V'$$

by Condition (1) such that

$$\frac{\operatorname{tr} M_V E\left[\alpha\alpha'|V\right] M_V}{\operatorname{tr}(M_V)} = \sigma^2_\alpha.$$

It follows that $E\left[\hat{\sigma}_{t,s}|V\right] = \sigma^2_\alpha + \tilde{\sigma}_{t,s}$ such that

$$E\left[\hat{S}_\eta|V\right] = \tilde{\Sigma} + \sigma^2_\alpha \mathbf{1}_T\mathbf{1}'_T$$

and thus

$$E\left[\hat{\Sigma}|V\right] = M_{\mathbf{1}_T}\tilde{\Sigma}M_{\mathbf{1}_T}.$$

For the variance we consider

$$\operatorname{vec} M_{\mathbf{1}_T}\hat{\Sigma}M_{\mathbf{1}_T} = \left(M_{\mathbf{1}_T} \otimes M_{\mathbf{1}_T}\right) \operatorname{vec}\hat{\Sigma}$$

such that it is enough to look at

$$
\begin{aligned}
& E\left[\hat{\sigma}_{t,s}\hat{\sigma}_{q,r}|V\right]-E\left[\hat{\sigma}_{t,s}|V\right]E\left[\hat{\sigma}_{q,r}|V\right] \\
= & \frac{E\left[\left(\alpha+\varepsilon_t\right)' M_V\left(\alpha+\varepsilon_s\right)\left(\alpha+\varepsilon_q\right)' M_V\left(\alpha+\varepsilon_r\right)|V\right]}{\operatorname{tr}(M_V)^2}-E\left[\hat{\sigma}_{t,s}|V\right]E\left[\hat{\sigma}_{q,r}|V\right] \\
= & \frac{\operatorname{tr} M_V E\left[\left(\alpha+\varepsilon_r\right)\left(\alpha+\varepsilon_t\right)'|V\right] M_V E\left[\left(\alpha+\varepsilon_s\right)\left(\alpha+\varepsilon_q\right)'|V\right] M_V}{\operatorname{tr}(M_V)^2} \\
& +\frac{\operatorname{tr} M_V E\left[\left(\alpha+\varepsilon_q\right)\left(\alpha+\varepsilon_t\right)'|V\right] M_V E\left[\left(\alpha+\varepsilon_s\right)\left(\alpha+\varepsilon_r\right)'|V\right] M_V}{\operatorname{tr}(M_V)^2} \\
= & \frac{\left(\sigma_\alpha^2+\tilde{\sigma}_{r,t}\right)\left(\sigma_\alpha^2+\tilde{\sigma}_{s,q}\right)\operatorname{tr}(M_V)}{\operatorname{tr}(M_V)^2}+\frac{\left(\sigma_\alpha^2+\tilde{\sigma}_{q,t}\right)\left(\sigma_\alpha^2+\tilde{\sigma}_{s,r}\right)\operatorname{tr}(M_V)}{\operatorname{tr}(M_V)^2} \\
= & \frac{\left(\sigma_\alpha^2+\tilde{\sigma}_{r,t}\right)\left(\sigma_\alpha^2+\tilde{\sigma}_{s,q}\right)}{n}+\frac{\left(\sigma_\alpha^2+\tilde{\sigma}_{q,t}\right)\left(\sigma_\alpha^2+\tilde{\sigma}_{s,r}\right)}{n}+o(n^{-1}).
\end{aligned}
$$

It follows that

$$
nE\operatorname{vec}\left(\hat{S}-\tilde{\Sigma}\right)\operatorname{vec}\left(\hat{S}-\tilde{\Sigma}\right)'=\left(I_T+K_{TT}\right)\left(\left(\sigma_\alpha^2\mathbf{1}_T\mathbf{1}_T'+\tilde{\Sigma}\right)\otimes\left(\sigma_\alpha^2\mathbf{1}_T\mathbf{1}_T'+\tilde{\Sigma}\right)\right)+o(1)
$$

and

$$
\begin{aligned}
& nE\operatorname{vec}\left(\hat{\Sigma}-M_{\mathbf{1}_T}\tilde{\Sigma}M_{\mathbf{1}_T}\right)\operatorname{vec}\left(\hat{\Sigma}-M_{\mathbf{1}_T}\tilde{\Sigma}M_{\mathbf{1}_T}\right)' \\
= & \left(M_{\mathbf{1}_T}\otimes M_{\mathbf{1}_T}\right)\left(I_T+K_{TT}\right)\left(\left(\sigma_\alpha^2\mathbf{1}_T\mathbf{1}_T'+\tilde{\Sigma}\right)\otimes\left(\sigma_\alpha^2\mathbf{1}_T\mathbf{1}_T'+\tilde{\Sigma}\right)\right)\left(M_{\mathbf{1}_T}\otimes M_{\mathbf{1}_T}\right)+o(1) \\
= & \left(M_{\mathbf{1}_T}\otimes M_{\mathbf{1}_T}\right)\left(I_T+K_{TT}\right)\left(\tilde{\Sigma}M_{\mathbf{1}_T}\otimes\tilde{\Sigma}M_{\mathbf{1}_T}\right)+o(1) \\
= & \left(I_T+K_{TT}\right)\left(M_{\mathbf{1}_T}\tilde{\Sigma}M_{\mathbf{1}_T}\otimes M_{\mathbf{1}_T}\tilde{\Sigma}M_{\mathbf{1}_T}\right)+o(1)
\end{aligned}
$$

where the last line follows from Magnus and Neudecker (1988, p.47). Note that similar results hold when $M_{\mathbf{1}_T}$ is replaced by B^{Δ}. We then write

$$
nE\operatorname{vec}\left(C'B\hat{\Sigma}B'C-\Sigma\right)\operatorname{vec}\left(C'B\hat{\Sigma}B'C-\Sigma\right)'=\left(I_T+K_{TT}\right)\left(\Sigma\otimes\Sigma\right)+o(1)
$$

and define $V_\Sigma=\left(I_T+K_{TT}\right)\left(\Sigma\otimes\Sigma\right)$. Using Magnus and Neudecker (1988, Theorem 10, p.47) we write

$$
\operatorname{vec}\Omega=\left(I_r\otimes K_{nr}\otimes I_n\right)\left(\operatorname{vec}\Sigma\otimes\operatorname{vec}I_n\right)
$$

such that the asymptotic variance V_Ω of $n^{1/2}\left(\operatorname{vec}\left(\hat{\Omega}-\Omega\right)\right)$ is

$$
V_\Omega=\left(I_r\otimes K_{nr}\otimes I_n\right)\left(V_\Sigma\otimes\operatorname{vec}I_n\left(\operatorname{vec}I_n\right)'\right)\left(I_T\otimes K_{nr}'\otimes I_n\right).
$$

B. Derivations of (2.2) and (2.3)

We start by verifying the assumptions of Rothenberg (1984a). This then implies that the expansions of Rothenberg (1988) are valid and the remainder of our work can be limited to finding explicit algebraic expressions of the terms in the expansions. For this purpose consider the transformed model $Y=Z\theta+\Upsilon\gamma+\varepsilon$. Let $X=[Z,\Upsilon]$ and $\beta=\left[\theta',\gamma\right]'$. If $\tilde{\varepsilon}$ is jointly normal as in Condition (2) then $\varepsilon\sim N\left(0,\Sigma\otimes I_n\right)$. Let $M=I_T\otimes M_V$ such that

$\hat{\varepsilon} = M_V \tilde{\varepsilon}$. Then stack $\hat{E} = [\hat{\varepsilon}_1, ..., \hat{\varepsilon}_T]'$ such that $\hat{\Sigma} = M_{\mathbf{1}_T} \frac{\hat{E}\hat{E}'}{\operatorname{tr}(M_V)} M_{\mathbf{1}_T}$. It follows that $\hat{\Omega}$ is a quadratic form of $\tilde{\varepsilon}$ and does not depend on β. Thus Assumption A of Rothenberg (1984a) is satisfied. Assumption B follows from Condition (2) except that here we only require approximations of order $o(n^{-1})$. Also note that Σ is linear in the elements $\sigma_{t,s}$ such that all higher order derivatives of Ω with respect to $\sigma_{t,s}$ are zero. Thus, Assumptions 1-3 of Rothenberg (1984a) follow easily from Condition (2). Finally, Assumption 4 of Rothenberg (1984a) is automatically satisfied because $\hat{\Sigma}$ is a quadratic form of Gaussian random variables and thus has bounded moments of all orders.

We first note that

$$R = \sqrt{n} \frac{\left(\Upsilon' \hat{\Omega}_z \Upsilon\right)^{-1} \Upsilon' \hat{\Omega}_z Y - (\Upsilon' \Omega_z \Upsilon)^{-1} \Upsilon' \Omega_z Y}{(\Upsilon' \Omega_z \Upsilon)^{-1/2}} = \frac{\left(\Upsilon' \hat{\Omega}_z \Upsilon\right)^{-1} \Upsilon' \hat{\Omega}_z \varepsilon - (\Upsilon' \Omega_z \Upsilon)^{-1} \Upsilon' \Omega_z \varepsilon}{(\Upsilon' \Omega_z \Upsilon)^{-1/2}}.$$

Consider the total derivative

$$\begin{aligned} d\left((\Upsilon'\Omega_z\Upsilon)^{-1}\Upsilon'\Omega_z\varepsilon\right) &= -(\Upsilon'\Omega_z\Upsilon)^{-1}\Upsilon' d\Omega_z \Upsilon (\Upsilon'\Omega_z\Upsilon)^{-1}\Upsilon'\Omega_z\varepsilon \\ &\quad + (\Upsilon'\Omega_z\Upsilon)^{-1}\Upsilon' d\Omega_z \varepsilon \end{aligned}$$

with

$$\begin{aligned} d\Omega_z &= -\Omega^{-1} d\Omega \Omega^{-1} + \Omega^{-1} d\Omega \Omega^{-1} Z (Z'\Omega^{-1}Z)^{-1} Z'\Omega^{-1} \\ &\quad -\Omega^{-1} Z (Z'\Omega^{-1}Z)^{-1} (Z'\Omega^{-1} d\Omega \Omega^{-1} Z)(Z'\Omega^{-1}Z)^{-1} Z'\Omega^{-1} \\ &\quad +\Omega^{-1} Z (Z'\Omega^{-1}Z)^{-1} Z'\Omega^{-1} d\Omega \Omega^{-1} \end{aligned}$$

such that

$$\operatorname{vec} d\Omega_z = -(\Omega_z \otimes \Omega_z) \operatorname{vec} d\Omega.$$

Now,

$$d \operatorname{vec}\left(\left(\Upsilon'\Omega_z\Upsilon\right)^{-1}\Upsilon'\Omega_z\varepsilon\right) = -\left(\varepsilon'(\Omega_z - \Omega_z\Upsilon(\Upsilon'\Omega_z\Upsilon)^{-1}\Upsilon'\Omega_z) \otimes (\Upsilon'\Omega_z\Upsilon)^{-1}\Upsilon'\Omega_z\right) \operatorname{vec} d\Omega$$

such that

$$R = \frac{(\varepsilon' H \otimes h')}{(h'\Omega h)^{-1/2}} \sqrt{n} \left(\Omega - \hat{\Omega}\right) + o_p(1)$$

by the delta method. In the same way,

$$\begin{aligned} d(\Upsilon'\Omega_z\Upsilon)^{-1} &= -(\Upsilon'\Omega_z\Upsilon)^{-1}\Upsilon' d\Omega_z \Upsilon \left(\Upsilon'\Omega_z\Upsilon\right)^{-1} \\ &= -h' d\Omega h \end{aligned}$$

and

$$\begin{aligned} d^2(\Upsilon'\Omega_z\Upsilon)^{-1} &= 2(\Upsilon'\Omega_z\Upsilon)^{-1}\Upsilon' d\Omega_z\Upsilon(\Upsilon'\Omega_z\Upsilon)^{-1}\Upsilon' d\Omega_z \Upsilon (\Upsilon'\Omega_z\Upsilon)^{-1} \\ &\quad -(\Upsilon'\Omega_z\Upsilon)^{-1}\Upsilon' d^2\Omega_z \Upsilon (\Upsilon'\Omega_z\Upsilon)^{-1} \end{aligned}$$

such that

$$\begin{aligned} d^2\Omega_z &= d\Omega_z d\Omega \Omega_z + \Omega_z d^2\Omega \Omega_z + \Omega_z d\Omega d\Omega_z \\ &= 2\Omega_z d\Omega \Omega_z d\Omega \Omega_z + \Omega_z d^2\Omega \Omega_z. \end{aligned}$$

Then

$$\begin{aligned}
d^2\left(\Upsilon'\Omega_z\Upsilon\right)^{-1} &= 2\left(\Upsilon'\Omega_z\Upsilon\right)^{-1}\Upsilon'\Omega_z d\Omega\Omega_z\Upsilon\left(\Upsilon'\Omega_z\Upsilon\right)^{-1}\Upsilon'\Omega_z d\Omega\Omega_z\Upsilon\left(\Upsilon'\Omega_z\Upsilon\right)^{-1} \\
&\quad -2\left(\Upsilon'\Omega_z\Upsilon\right)^{-1}\Upsilon'\Omega_z d\Omega\Omega_z d\Omega\Omega_z\Upsilon\left(\Upsilon'\Omega_z\Upsilon\right)^{-1} + \left(\Upsilon'\Omega_z\Upsilon\right)^{-1}\Upsilon'\Omega_z d^2\Omega\Omega_z\Upsilon\left(\Upsilon'\Omega_z\Upsilon\right)^{-1} \\
&= 2\operatorname{tr} hh'd\Omega H d\Omega + \operatorname{tr} d^2\Omega hh'
\end{aligned}$$

and

$$\operatorname{vec} d^2\left(\Upsilon'\Omega_z\Upsilon\right)^{-1} = 2\left(\operatorname{vec} d\Omega\right)'\left(H\otimes hh'\right)\operatorname{vec} d\Omega + \operatorname{tr} d^2\Omega hh'.$$

In the Taylor expansion the term $d^2\Omega = 0$ because all the elements are linear functions of the parameters. We thus have

$$S = \frac{(h'\otimes h')}{(h'\Omega h)}\sqrt{n}\operatorname{vec}(\Omega-\hat{\Omega}) + \frac{\sqrt{n}\operatorname{tr}\left[\operatorname{vec}(\Omega-\hat{\Omega})\operatorname{vec}(\Omega-\hat{\Omega})'\left(H\otimes hh'\right)\right]}{(h'\Omega h)}.$$

Note that $E\operatorname{vec}(\Omega-\hat{\Omega}) = 0$ and $\operatorname{Var}\left(\sqrt{n}\operatorname{vec}(\Omega-\hat{\Omega})\right) = V_\Omega + O(n^{-1/2})$. Then,

$$ES = n^{-1/2}\frac{\operatorname{tr}\left[V_\Omega\left(H\otimes hh'\right)\right]}{(h'\Omega h)} + O(n^{-1}),$$

$$\operatorname{Var} S = \frac{\operatorname{tr}\left[V_\Omega\left(hh'\otimes hh'\right)\right]}{(h'\Omega h)^2} + O(n^{-1/2}).$$

For R note that under normality all the third moments are zero such that $ER = 0$ and

$$\begin{aligned}
(h'\Omega h)\operatorname{Var} R &= nE\left[\operatorname{tr}\operatorname{vec}\left(\Omega-\hat{\Omega}\right)\operatorname{vec}\left(\Omega-\hat{\Omega}\right)'\left(H\varepsilon\varepsilon' H\otimes hh'\right)\right] \\
&= \operatorname{tr}\left[V_\Omega\left(H\Omega H\otimes hh'\right)\right] + O(n^{-1})
\end{aligned}$$

where remaining terms are zero due to normality. Since $H\Omega H = H(I - Z(Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}) = H$ it follows that

$$\operatorname{Var} R = \frac{\operatorname{tr}\left[V_\Omega\left(H\otimes hh'\right)\right]}{(h'\Omega h)} + O(n^{-1}).$$

B.1. Derivation of (2.4) and (2.5)

Note that $I_n = \sum_{i=1}^n b_i b_i'$ and $\operatorname{vec} b_i b_i' = b_i \otimes b_i$. Then $\operatorname{vec} I_n\left(\operatorname{vec} I_n\right)' = \sum_{i,j=1}^n b_i b_j' \otimes b_i b_j'$. Next, consider

$$\begin{aligned}
\left(I_r\otimes K_{nr}\otimes I_n\right)\left(\Sigma\otimes\Sigma\otimes\operatorname{vec} I_n\left(\operatorname{vec} I_n\right)'\right)\left(I_T\otimes K_{nr}'\otimes I_n\right) &= \textstyle\sum_{i,j=1}^n\left(\Sigma\otimes K_{nr}\left(\Sigma\otimes b_i b_j'\right)K_{nr}'\otimes b_i b_j'\right) \\
&= \textstyle\sum_{i,j=1}^n\left(\Sigma\otimes b_i b_j'\otimes\Sigma\otimes b_i b_j'\right).
\end{aligned}$$

We then find that

$$\operatorname{tr}\sum_{i,j=1}^n\left(\Sigma\otimes b_i b_j'\otimes\Sigma\otimes b_i b_j'\right)\left(hh'\otimes hh'\right) = \textstyle\sum_{i,j=1}^n\left(h'\left(\Sigma\otimes b_i b_j'\right)h\right)^2.$$

Next consider

$$\begin{aligned}
&\textstyle\sum_{i,j=1}^n\left(I_r\otimes K_{nr}\otimes I_n\right)\left(K_{rr}\left(\Sigma\otimes\Sigma\right)\otimes b_i b_j'\otimes b_i b_j'\right)\left(I_T\otimes K_{nr}'\otimes I_n\right) \\
=\ &\textstyle\sum_{l,m=1}^r\sum_{i,j=1}^n\left(I_r\otimes K_{nr}\otimes I_n\right)\left(e_l e_m'\Sigma\otimes e_m e_l'\Sigma\otimes b_i b_j'\otimes b_i b_j'\right)\left(I_T\otimes K_{nr}'\otimes I_n\right) \\
=\ &\textstyle\sum_{l,m=1}^r\sum_{i,j=1}^n\left(e_l e_m'\Sigma\otimes b_i b_j'\otimes e_m e_l'\Sigma\otimes b_i b_j'\right)
\end{aligned}$$

such that

$$\begin{aligned} & \operatorname{tr} \textstyle\sum_{l,m=1}^{r} \sum_{i,j=1}^{n} \left(e_l e_m' \Sigma \otimes b_i b_j' \otimes e_m e_l' \Sigma \otimes b_i b_j'\right) \left(hh' \otimes hh'\right) \\ = & \textstyle\sum_{l,m=1}^{r} \sum_{i,j=1}^{n} \left(h' \left(e_l e_m' \Sigma \otimes b_i b_j'\right) h\right) \left(h' \left(e_m e_l' \Sigma \otimes b_i b_j'\right) h\right) \end{aligned}$$

and

$$\begin{aligned} \operatorname{tr}\left(V_\Omega \left(H \otimes hh'\right)\right) &= \textstyle\sum_{i,j=1}^{n} \left(h' \left(\Sigma \otimes b_i b_j'\right) h\right)^2 \\ &\quad + \textstyle\sum_{l,m=1}^{r} \sum_{i,j=1}^{n} \left(h' \left(e_l e_m' \Sigma \otimes b_i b_j'\right) h\right) \left(h' \left(e_m e_l' \Sigma \otimes b_i b_j'\right) h\right) \end{aligned}$$

It now also follows immediately that

$$\begin{aligned} \operatorname{tr}\left(V_\Omega \left(H \otimes hh'\right)\right) &= \textstyle\sum_{i,j=1}^{n} \operatorname{tr}\left(\left(\Sigma \otimes b_i b_j'\right) H\right) \left(\left(h' \left(\Sigma \otimes b_i b_j'\right) h\right)\right) \\ &\quad + \textstyle\sum_{l,m=1}^{r} \sum_{i,j=1}^{n} \operatorname{tr}\left(\left(e_l e_m' \Sigma \otimes b_i b_j'\right) H\right) \left(h' \left(e_m e_l' \Sigma \otimes b_i b_j'\right) h\right) \end{aligned}$$

B.2. Derivation of Approximation to $\mathbf{T}_2$

Define $M = I_T \otimes M_V$, $\tilde{x} = \sqrt{n} M_{\tilde{z}} \tilde{\Upsilon} \left(\tilde{\Upsilon}' M_{\tilde{z}} \tilde{\Upsilon}\right)^{-1}$ and $\tilde{v} = M \tilde{\Omega} \tilde{x} / \sqrt{\tilde{x}' \tilde{\Omega} \tilde{x}}$ where $\tilde{\Omega} = \left(\tilde{\Sigma} \otimes I\right)$. Then let $\bar{T}_2 - b_2 \equiv \mathcal{Y}$,

$$\hat{\varepsilon} = M \tilde{\varepsilon} = \frac{M \Omega \tilde{x}}{\tilde{x}' \Omega \tilde{x}} \tilde{x}' \tilde{\varepsilon} + M \left(I - \frac{\Omega \tilde{x} \tilde{x}'}{\tilde{x}' \Omega \tilde{x}}\right) \tilde{\varepsilon} \equiv \tilde{v} \mathcal{Y} + \tilde{u}.$$

The estimator for $C' B M_{\mathbf{1}_T} \tilde{\Sigma} M_{\mathbf{1}_T} B' C$ with typical element t, s denoted by $\sigma_{t,s}$ can be written by defining the t-th $n \times 1$ block of $\hat{\varepsilon}$ as $\hat{\varepsilon}_t$ and in the same way we define the $n \times 1$ vectors $\tilde{v}_t$ and $\tilde{u}_t$. Then stack $\hat{E} = \left[\hat{\varepsilon}_1, ..., \hat{\varepsilon}_T\right]'$, $\tilde{\Psi} = \left[\tilde{v}_1, ..., \tilde{v}_T\right]'$ and $\tilde{U} = \left[\tilde{u}_1, ..., \tilde{u}_T\right]'$ such that

$$\begin{aligned} C' B \hat{\Sigma} B' C &= C' B M_{\mathbf{1}_T} \frac{\hat{E} \hat{E}'}{\operatorname{tr}\left(M_V\right)} M_{\mathbf{1}_T} B' C \\ &= C' B M_{\mathbf{1}_T} \left[\left(\mathcal{Y}\right)^2 \frac{\tilde{\Psi} \tilde{\Psi}'}{\operatorname{tr}\left(M_V\right)} + \mathcal{Y} \left(\frac{\tilde{\Psi} \tilde{U}'}{\operatorname{tr}\left(M_V\right)} + \frac{\tilde{U} \tilde{\Psi}'}{\operatorname{tr}\left(M_V\right)}\right) + \frac{\tilde{U} \tilde{U}'}{\operatorname{tr}\left(M_V\right)}\right] M_{\mathbf{1}_T} B' C. \end{aligned}$$

Now define the t-th row of the $r \times n$ matrix $C' B M_{\mathbf{1}_T} \tilde{\Psi}$ by v_t' and correspondingly for the rows of $C' B M_{\mathbf{1}_T} \tilde{U}$ we use the notation u_t'. We also define $x = \sqrt{n} M_z \Upsilon \left(\Upsilon' M_z \Upsilon\right)^{-1}$ with t-th $n \times 1$ block denoted by x_t. Then, consider $T_2 = \bar{T}_2 / \left(1 + W/\sqrt{n}\right)^{1/2}$ with

$$W = \sqrt{n} \frac{\left(\Upsilon' M_z \Upsilon\right)^{-1} \Upsilon' M_z \left(\hat{\Omega} - \Omega\right) M_z \Upsilon \left(\Upsilon' M_z \Upsilon\right)^{-1}}{\left(\Upsilon' M_z \Upsilon\right)^{-1} \Upsilon' M_z \Omega M_z \Upsilon \left(\Upsilon' M_z \Upsilon\right)^{-1}} = \sqrt{n} \frac{x' \left(\hat{\Omega} - \Omega\right) x}{x' \Omega x}$$

such that

$$W = \sqrt{n} \frac{\sum_{t,s}^{r} x_t' x_s \left(\frac{u_t' u_s}{\operatorname{tr} M_V} - \sigma_{t,s}\right)}{\sum_{t,s}^{r} x_t' x_s \sigma_{t,s}} + \sqrt{n} \frac{\sum_{t,s}^{r} x_t' x_s \frac{v_t' v_s}{\operatorname{tr} M_V}}{\sum_{t,s}^{r} x_t' x_s \sigma_{t,s}} \mathcal{Y}^2 + \sqrt{n} \mathcal{Y} \frac{\sum_{t,s}^{r} x_t' x_s \frac{u_t' v_s}{\operatorname{tr} M_V}}{\sum_{t,s}^{r} x_t' x_s \sigma_{t,s}}$$

where, under normality assumptions, u is normally distributed and uncorrelated with $\mathcal{Y}$. Note that $x_t' x_s = O(1)$ and $v_t' v_s = \left(x' \Omega x\right)^{-1} \sum_{p,q=1}^{r} \sigma_{t,p} \sigma_{q,s} \operatorname{tr} x_p x_q' M_V = O(1)$ for all t, s. Next consider

$$\begin{aligned} E \tilde{u} \tilde{u}' &= M \left(I - \frac{\Omega \tilde{x} \tilde{x}'}{\tilde{x}' \Omega \tilde{x}}\right) \Omega \left(I - \frac{\Omega \tilde{x} \tilde{x}'}{\tilde{x}' \Omega \tilde{x}}\right)' M \\ &= M \Omega M - \frac{M \Omega \tilde{x} \tilde{x}' \Omega M}{\tilde{x}' \Omega \tilde{x}} \end{aligned}$$

where

$$M\Omega M=\begin{bmatrix}\tilde{\sigma}_{11}M_V & \cdots & \tilde{\sigma}_{1T}M_V\\ & \ddots & \vdots\\ & & \tilde{\sigma}_{TT}M_V\end{bmatrix}=\tilde{\Sigma}_\eta\otimes M_V$$

and

$$\frac{M\Omega\tilde{x}\tilde{x}'\Omega M}{\tilde{x}'\Omega\tilde{x}}=\frac{1}{\tilde{x}'\Omega\tilde{x}}\begin{bmatrix}M_V\sum_{ij}\tilde{\sigma}_{1i}\tilde{\sigma}_{j1}\tilde{x}_i\tilde{x}_j'M_V & \cdots & M_V\sum_{ij}\tilde{\sigma}_{1i}\tilde{\sigma}_{jT}\tilde{x}_i\tilde{x}_j'M_V\\ & \ddots & \vdots\\ & & M_V\sum_{ij}\tilde{\sigma}_{Ti}\tilde{\sigma}_{jT}\tilde{x}_i\tilde{x}_j'M_V\end{bmatrix}$$

such that

$$\begin{aligned}Eu_t'u_s &= \operatorname{tr}\tilde{\sigma}_{ts}M_V-\frac{1}{\tilde{x}'\Omega\tilde{x}}\sum_{ij}\tilde{\sigma}_{ti}\tilde{\sigma}_{js}\operatorname{tr}M_V\tilde{x}_i\tilde{x}_j'\\ &= (\operatorname{tr}M_V)\,\tilde{\sigma}_{ts}-\frac{1}{\tilde{x}'\Omega\tilde{x}}\sum_{ij}\tilde{\sigma}_{ti}\tilde{\sigma}_{js}\operatorname{tr}M_V\tilde{x}_i\tilde{x}_j'.\end{aligned}$$

By the same arguments as in Rothenberg (1988) it then follows that

$$\sqrt{n}E\left[W|\bar{T}_2\right]=-\frac{\sum_{t,s}^r x_t'x_s\sum_{ij}\tilde{\sigma}_{ti}\tilde{\sigma}_{js}\frac{\operatorname{tr}M_V\tilde{x}_i\tilde{x}_j'}{\operatorname{tr}M_V}}{\left(\sum_{t,s}^r x_t'x_s\tilde{\sigma}_{t,s}\right)^2}+\frac{\sum_{t,s}^r x_t'x_s\frac{v_t'v_s}{\operatorname{tr}M_V}}{\sum_{t,s}^r x_t'x_s\tilde{\sigma}_{t,s}}\mathcal{Y}^2+O(n^{-1/2})$$

and

$$\operatorname{var}\left(W|\bar{T}_2\right)=\frac{\operatorname{tr}V_\Omega\left(xx'\otimes xx'\right)}{\left(x'\Omega x\right)^2}.$$

Note that

$$\operatorname{tr}M_V\tilde{x}_i\tilde{x}_j'=\tilde{x}_j'M_V\tilde{x}_i=\left(\tilde{\Upsilon}_i-\tilde{Z}_i\left(\tilde{Z}'\tilde{Z}\right)^{-1}Z'\tilde{\Upsilon}\right)'M_V\left(\tilde{\Upsilon}_j-Z_j\left(\tilde{Z}'\tilde{Z}\right)^{-1}Z'\tilde{\Upsilon}\right)\left(\tilde{\Upsilon}'M_{\tilde{z}}\tilde{\Upsilon}\right)^{-2}$$

with

$$M_V\tilde{\Upsilon}_j=0\text{ and }M_VZ_j=0\text{ for all }j=1,...,T$$

such that $\operatorname{tr}M_V\tilde{x}_i\tilde{x}_j'=0$. Next, consider

$$\tilde{v}=M\tilde{\Omega}\tilde{x}/\sqrt{\tilde{x}'\tilde{\Omega}\tilde{x}}=\left(\tilde{\Sigma}\otimes M_V\right)\left(\tilde{\Upsilon}-Z\left(\tilde{Z}'\tilde{Z}\right)^{-1}Z'\tilde{\Upsilon}\right)/\sqrt{\tilde{x}'\tilde{\Omega}\tilde{x}}=0$$

by the same argument as before.

B.3. Derivations for Results in Section 3

B.3.1. Level Specification

The results follow from specializing the expressions for $\operatorname{var}(S)$ and $\operatorname{var}(R)$ to this case. First consider

$$Z'\Omega^{-1}Z=n\Sigma^{-1},$$

$$\Omega^{-1}Z\left(Z'\Omega^{-1}Z\right)^{-1}Z'\Omega^{-1}=\Sigma^{-1}\otimes\frac{\mathbf{1}_n\mathbf{1}_n'}{n}$$

and

$$\Omega_z^{-1} = \Sigma^{-1} \otimes M_{\mathbf{1}_n}.$$

Next, we express $\Upsilon = \sum_{t=\tau}^{T} C'BM_{\mathbf{1}_T} a_{t,T} \otimes \tilde{\Upsilon}_\tau^\Delta$. Then

$$n^{-1}\Upsilon'\Omega_z\Upsilon = n^{-1}\tilde{X}_\tau' M_{\mathbf{1}_n}\tilde{X}_\tau \sum_{t,s=\tau}^{T}\left(a_{t,T}' M_{\mathbf{1}_T} B'C\Sigma^{-1}C'BM_{\mathbf{1}_T} a_{s,T}\right)$$

since $n^{-1}\tilde{\Upsilon}_\tau^{\Delta\prime} M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta \xrightarrow{p} \operatorname{Var}\Delta T_{i\tau} = p(1-p)$. Also let $\Delta_T = p\,(1-p)\sum_{t,s=\tau}^{T}\left(a_{t,T}' M_{\mathbf{1}_T} B'C\Sigma^{-1}C'BM_{\mathbf{1}_T} a_{s,T}\right)$. Then $n^{-1}\Upsilon'\Omega_z\Upsilon \xrightarrow{p} \Delta_T$ as $n \to \infty$.

Next let $\delta_x = n\left(\Upsilon'\Omega_z\Upsilon\right)^{-1}$, $\gamma(t_1) = \Sigma^{-1}C'BM_{\mathbf{1}_T} a_{t_1,T}\delta_x$and $\Gamma\left(t_1,t_2\right) = \gamma(t_1)\gamma(t_2)'$. Then

$$h = n^{-1}\delta_x\Omega_z\Upsilon = n^{-1}\delta_x \sum_{t=1}^{T}\left(\Sigma^{-1}C'BM_{\mathbf{1}_T} a_{t_1,T} \otimes M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta\right) = n^{-1}\sum_{t=1}^{T}\left(\gamma(t) \otimes M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta\right).$$

Consider

$$\begin{aligned}
n^4\left(hh' \otimes hh'\right) &= \delta_x^4\Omega_z\Upsilon\Upsilon'\Omega_z \otimes \Omega_z\Upsilon\Upsilon'\Omega_z \\
&= \textstyle\sum_{t_1,\ldots,t_4=1}^{T}\left[\Gamma(t_1,t_2) \otimes M_{\mathbf{1}_T}\tilde{\Upsilon}_\tau^\Delta\tilde{\Upsilon}_\tau^{\Delta\prime} M_{\mathbf{1}_T} \otimes \Gamma(t_3,t_4) \otimes M_{\mathbf{1}_n}'\tilde{\Upsilon}_\tau^\Delta\tilde{\Upsilon}_\tau^{\Delta\prime} M_{\mathbf{1}_n}\right] \\
&= \textstyle\sum_{t_1,\ldots,t_4=1}^{T}\left(I_r \otimes K_{nr} \otimes I_n\right)\left[\Gamma(t_1,t_2) \otimes \Gamma(t_3,t_4) \otimes M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta\tilde{\Upsilon}_\tau^{\Delta\prime} M_{\mathbf{1}_n} \otimes M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta\tilde{\Upsilon}_\tau^{\Delta\prime} M_{\mathbf{1}_n}\right]\left(I_r \otimes K_{nr}' \otimes I_n\right)
\end{aligned}$$

such that

$$\begin{aligned}
& n^4 \operatorname{tr} V_{\Omega_\eta}\left(hh' \otimes hh'\right) \\
={}& \operatorname{tr}\textstyle\sum_{t_1,\ldots,t_4=\tau}^{T} V_{\Omega_\eta}\left(I_r \otimes K_{nr} \otimes I_n\right)\left[\Gamma(t_1,t_2) \otimes \Gamma(t_3,t_4) \otimes M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta\tilde{\Upsilon}_\tau^{\Delta\prime} M_{\mathbf{1}_n} \otimes M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta\tilde{\Upsilon}_\tau^{\Delta\prime} M_{\mathbf{1}_n}\right]\left(I_r \otimes K_{nr}' \otimes I_n\right) \\
={}& \textstyle\sum_{t_1,\ldots,t_4=\tau}^{T}\left[\gamma(t_2)'\Sigma\gamma(t_1) \otimes \gamma(t_4)'\Sigma\gamma(t_3) \otimes \left(\tilde{\Upsilon}_\tau^{\Delta\prime} M_{\mathbf{1}_n} \otimes \tilde{\Upsilon}_\tau^{\Delta\prime} M_{\mathbf{1}_n}\right)\operatorname{vec} I_n\left(\operatorname{vec} I_n\right)'\left(M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta \otimes M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta\right)\right] \\
& + \textstyle\sum_{t_1,\ldots,t_4=\tau}^{T}\left[\left(\gamma(t_2)' \otimes \gamma(t_4)'\right)K_{rr}\left(\Sigma \otimes \Sigma\right)\left(\gamma(t_1) \otimes \gamma(t_3)\right)\right. \\
& \left.\otimes\left(\tilde{\Upsilon}_\tau^{\Delta\prime} M_{\mathbf{1}_n} \otimes \tilde{\Upsilon}_\tau^\Delta M_{\mathbf{1}_n}\right)\operatorname{vec} I_n\left(\operatorname{vec} I_n\right)'\left(M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta \otimes M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta\right)\right]
\end{aligned}$$

where $\left(\tilde{\Upsilon}_\tau^{\Delta\prime} M_{\mathbf{1}_n} \otimes \tilde{\Upsilon}_\tau^{\Delta\prime} M_{\mathbf{1}_n}\right)\operatorname{vec} I_n\left(\operatorname{vec} I_n\right)'\left(M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta \otimes M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta\right) = \left(\sum_{i=1}^{n}\left(\Delta T_{i\tau} - \bar{T}_\tau^\Delta\right)^2\right)^2$ and $\bar{T}_\tau^\Delta = n^{-1}\sum_{i=1}^{n}\Delta T_{i\tau}$. Therefore

$$\begin{aligned}
& n^2 \operatorname{tr} V_\Omega\left(hh' \otimes hh'\right) \\
={}& \left(\textstyle\sum_{t_1,t_2=\tau}^{T}\gamma(t_1)'\Sigma\gamma(t_2)\right)^2 n^{-2}\left(\textstyle\sum_{i=1}^{n}\left(\Delta T_{i\tau} - \bar{T}_\tau^\Delta\right)^2\right)^2 \\
& + \textstyle\sum_{t_1,\ldots,t_4=\tau}^{T}\left[\left(\gamma(t_2)' \otimes \gamma(t_4)'\right)K_{rr}\left(\Sigma \otimes \Sigma\right)\left(\gamma(t_1) \otimes \gamma(t_3)\right)\right] n^{-2}\left(\textstyle\sum_{i=1}^{n}\left(\Delta T_{i\tau} - \bar{T}_\tau^\Delta\right)^2\right)^2 \\
& \xrightarrow{p} \left(p\left(1-p\right)\right)^2\Delta_T^{-4}\left(\textstyle\sum_{t_1,t_2=\tau}^{T} a_{t_1,T}' M_{\mathbf{1}_T} B'C\Sigma^{-1}C'BM_{\mathbf{1}_T} a_{t_2,T}\right)^2 \\
& + \left(p\left(1-p\right)\right)^2\Delta_T^{-4}\left(\textstyle\sum_{t_1,t_2=\tau}^{T} a_{t_1,T}' M_{\mathbf{1}_T} B'C\Sigma^{-1}C'BM_{\mathbf{1}_T} a_{t_2,T}\right)^2 \\
={}& 2\Delta_T^{-2}.
\end{aligned}$$

where the last equality follows from the fact that for two $r \times 1$ vectors a and b, $K_{rr}\left(a \otimes b\right) = \sum_{i,j=1}^{r}\left(e_i e_j' a \otimes e_j e_i' b\right) = \left(b \otimes a\right)$. Also note that $n^{-1}h'\Omega h = n^{-1}\delta_x^{-2}\Upsilon'\Omega\Upsilon \xrightarrow{p} \Delta_T^{-1}$. It thus follows that

$$\frac{\operatorname{tr} V_\Omega\left(hh' \otimes hh'\right)}{\left(h'\Omega h\right)^2} \xrightarrow{p} 2.$$

For $H=\Omega_z-\Omega_z\Upsilon\left(\Upsilon'\Omega_z\Upsilon\right)^{-1}\Upsilon'\Omega_z$ use $n\left(H_l-\Omega_z\right)=\Delta_T^{-1}\Omega_z\Upsilon\Upsilon'\Omega_z+o_p(1)$ such that we can replace H by $\Omega_z-n^{-1}\Delta_T^{-1}\Omega_z\Upsilon\Upsilon'\Omega_z$. Now

$$\Omega_z-n^{-1}\Delta_T^{-1}\Omega_z\Upsilon\Upsilon'\Omega_z=\Sigma^{-1}\otimes M_{\mathbf{1}_n}-n^{-1}\Delta_T^{-1}\textstyle\sum_{t_1,t_2=\tau}^{T}\Sigma^{-1}C'BM_{\mathbf{1}_T}a_{t_1,T}a_{t_2,T}'M_{\mathbf{1}_T}B'C\Sigma^{-1}\otimes M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}$$

such that

$$\begin{aligned}
& n^2\left(\Omega^{-1}-n^{-1}\Delta_{T,l}^{-1}\Omega_{\eta,z}\Upsilon\Upsilon'\Omega_{\eta,z}\right)\otimes h_lh_l' \\
= \;& \textstyle\sum_{t_1,t_2=\tau}^{T}\left[\Sigma_\eta^{-1}\otimes M_{\mathbf{1}_n}\otimes\Gamma_l(t_1,t_2)\otimes M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}\right] \\
& -n^{-1}\Delta_T^{-1}\textstyle\sum_{t_1,\ldots,t_4=\tau}^{T}\left[\Sigma_\eta^{-1}C'BM_{\mathbf{1}_T}a_{t_1,T}a_{t_2,T}'M_{\mathbf{1}_T}B'C\Sigma_\eta^{-1}\otimes M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}\otimes\Gamma_l(t_3,t_4)\otimes M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}\right] \\
= \;& \textstyle\sum_{t_1,t_2=\tau}^{T}\left(I_r\otimes K_{nr}\otimes I_n\right)\left[\Sigma^{-1}\otimes\Gamma(t_1,t_2)\otimes M_{\mathbf{1}_n}\otimes M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}\right]\left(I_r\otimes K_{nr}'\otimes I_n\right) \\
& -n^{-1}\Delta_T^{-1}\textstyle\sum_{t_1,\ldots,t_4=\tau}^{T}\left(I_r\otimes K_{nr}\otimes I_n\right)\left[\Sigma^{-1}C'BM_{\mathbf{1}_T}a_{t_1,T}a_{t_2,T}'M_{\mathbf{1}_T}B'C\Sigma^{-1}\otimes\Gamma(t_3,t_4)\right. \\
& \left.\otimes M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}\otimes M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}\right]\left(I_r\otimes K_{nr}'\otimes I_n\right).
\end{aligned}$$

It then follows that

$$\begin{aligned}
& n\operatorname{tr}V_\Omega\left(H\otimes hh'\right) \\
= \;& n^{-1}\textstyle\sum_{t_1,t_2=1}^{T}\operatorname{tr}\left[\Sigma^{-1/2}\otimes\gamma(t_2)'\otimes M_{\mathbf{1}_n}\otimes\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}\right]\left[\Sigma\otimes\Sigma\otimes\operatorname{vec}I_n\left(\operatorname{vec}I_n\right)'\right]\left[\Sigma^{-1/2}\otimes\gamma(t_1)\otimes M_{\mathbf{1}_n}\otimes M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta\right] \\
& -\delta_x^{-2}n^{-2}\Delta_T^{-1}\frac{T}{p}\textstyle\sum_{t_1,\ldots,t_4=1}^{T}\left[\gamma(t_2)'\otimes\gamma(t_4)'\otimes\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}\otimes\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}\right]\left[\Sigma\otimes\Sigma\otimes\operatorname{vec}I_n\left(\operatorname{vec}I_n\right)'\right] \\
& \times\left[\gamma(t_1)\otimes\gamma(t_3)\otimes M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta\otimes M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta\right] \\
= \;& n^{-1}\textstyle\sum_{t_1,t_2=1}^{T}\operatorname{tr}\left[I_r\otimes\gamma(t_2)'\Sigma\gamma(t_1)\otimes\left(M_{\mathbf{1}_n}\otimes\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}\right)\operatorname{vec}I_n\left(\operatorname{vec}I_n\right)'\left(M_{\mathbf{1}_n}\otimes M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^{\Delta\prime}\right)\right] \\
& -\delta_x^{-2}n^{-2}\Delta_T^{-1}\textstyle\sum_{t_1,\ldots,t_4=1}^{T}\operatorname{tr}\left[\gamma(t_2)'\Sigma\gamma(t_1)\otimes\gamma(t_4)'\Sigma\gamma(t_3)\otimes\left(\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}\otimes\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}\right)\operatorname{vec}I_n\left(\operatorname{vec}I_n\right)'\left(M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta\otimes M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta\right)\right]
\end{aligned}$$

where $\left(\operatorname{vec}I_n\right)'\left(M_{\mathbf{1}_n}\otimes M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta\right)=\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}$. From $\operatorname{tr}\left(A\otimes B\right)=\operatorname{tr}A\operatorname{tr}B$ it follows that

$$n^{-1}\textstyle\sum_{t_1,t_2=\tau}^{T}\operatorname{tr}\left[I_r\otimes\gamma(t_2)'\Sigma\gamma(t_1)\otimes M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}\right]=r\textstyle\sum_{t_1,t_2=\tau}^{T}\gamma(t_1)'\Sigma\gamma(t_2)\dfrac{\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta}{n}\xrightarrow{p}r\Delta_T^{-1}$$

and

$$\begin{aligned}
& \delta_x^{-2}n^{-2}\Delta_T^{-1}\textstyle\sum_{t_1,\ldots,t_4=\tau}^{T}\operatorname{tr}\left[\gamma(t_2)'\Sigma\gamma(t_1)\otimes\gamma(t_4)'\Sigma\gamma(t_3)\otimes\left(\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}\otimes\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}\right)\operatorname{vec}I_n\left(\operatorname{vec}I_n\right)'\left(M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta\otimes M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta\right)\right] \\
= \;& \delta_x^{-2}n^{-2}\Delta_T^{-1}\left(\textstyle\sum_{t_1,t_2=\tau}^{T}\gamma(t_1)'\Sigma\gamma(t_2)\right)^2\left(\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta\right)^2. \\
& \xrightarrow{p}\Delta_T^{-1}.
\end{aligned}$$

It thus follows that

$$\frac{\operatorname{tr}V_\Omega\left(H\otimes hh'\right)}{\left(h'\Omega h\right)}\xrightarrow{p}r-1$$

which establishes the results for T_1.

For the test T_2 note that in this case $V_{it} = T_{it}$, $V_i = [(V_{i1}, ..., V_{iT}), 1]$ and $V = [V_1', ..., V_n']'$. Define $M_V = I_n - V(V'V)^{-1}V', M = I_r \otimes M_V,\ x = \sqrt{n}\left(I_r \otimes M_{\mathbf{1}_n}\right) \Upsilon \left(\Upsilon' \left(I_r \otimes M_{\mathbf{1}_n}\right) \Upsilon\right)^{-1}$ and $v = M\Omega_\eta x/\sqrt{x'\Omega_\eta x} = 0$ because $M_V M_{\mathbf{1}_n} \tilde{\Upsilon}_\tau^\Delta = 0$ such that

$$n^{-1}\Upsilon' \left(I_r \otimes M_{\mathbf{1}_n}\right) \Omega \left(I_r \otimes M_{\mathbf{1}_n}\right) \Upsilon = n^{-1}\tilde{\Upsilon}_\tau^{\Delta\prime} M_{\mathbf{1}_n} \tilde{\Upsilon}_\tau^\Delta \left(\sum_{t_1,t_2=\tau}^T a_{t_1,T}' M_{\mathbf{1}_T} B' C \Sigma C' B M_{\mathbf{1}_T} a_{t_2,T}\right) \xrightarrow{p} D_\tau$$

where $D_\tau = \left(\sum_{t_1,t_2=\tau}^T a_{t_1,T}' M_{\mathbf{1}_T} B' C \Sigma_\eta C' B M_{\mathbf{1}_T} a_{t_2,T}\right) p(1-p)$ and

$$\Upsilon' \left(I_r \otimes M_{\mathbf{1}_n}\right) \Upsilon = n^{-1}\tilde{\Upsilon}_\tau^{\Delta\prime} M_{\mathbf{1}_n} \tilde{\Upsilon}_\tau^\Delta \left(\sum_{t_1,t_2=\tau}^T a_{t_1,T}' M_{\mathbf{1}_T} B' C C' B M_{\mathbf{1}_T} a_{t_2,T}\right) \xrightarrow{p} D_\tau^\circ$$

such that $x'\Omega x \xrightarrow{p} D_\tau / \left(D_\tau^\circ\right)^2$. Then let

$$\bar{T}_{2,\eta} = \frac{\left(\Upsilon' M_z \Upsilon\right)^{-1} \Upsilon' M_z Y - \gamma_0}{\left(\left(\Upsilon' M_z \Upsilon\right)^{-1} \Upsilon' M_z \Omega_\eta M_z \Upsilon \left(\Upsilon' M_z \Upsilon\right)^{-1}\right)^{1/2}}$$

and set $\bar{T}_2 - b_2 \equiv \mathcal{Y}$ such that

$$\hat{\varepsilon} = M\varepsilon = \frac{M\Omega x}{x'\Omega x} x'\varepsilon + M\left(I - \frac{\Omega x x'}{x'\Omega x}\right)\varepsilon \equiv e.$$

Then, consider $T_2 = \bar{T}_2 / \left(1 + W/\sqrt{n}\right)^{1/2}$ with

$$W = \sqrt{n}\frac{\Upsilon' M_z \left(\hat{\Omega} - \Omega\right) M_z \Upsilon}{\Upsilon' M_z \Omega M_z \Upsilon}.$$

Specializing previous arguments it then follows that

$$\sqrt{n} E\left[W|\bar{T}_2\right] = O(n^{-1/2})$$

and

$$\begin{aligned}
\operatorname{var}\left(W|\bar{T}_2\right) &= \frac{n^{-2}\left(\Upsilon' M_z \otimes \Upsilon' M_z\right) V_\Omega \left(M_z \Upsilon \otimes M_z \Upsilon\right)}{\left(D_\tau\right)^2} + O_p(n^{-1/2}) \\
&= 2\frac{\left(\sum_{t_1,t_2=\tau}^T a_{t_1,T}' M_{\mathbf{1}_T} B' C \Sigma_\eta C' B M_{\mathbf{1}_T} a_{t_2,T}\right)^2 \left(n^{-1}\tilde{\Upsilon}_\tau^{\Delta\prime} M_{\mathbf{1}_n} \tilde{\Upsilon}_\tau^\Delta\right)^2}{\left(D_\tau\right)^2} + O_p\left(n^{-1/2}\right) \\
&= 2 + O_p(n^{-1/2}).
\end{aligned}$$

B.3.2. First Difference Specification

We first simplify the expression for the GLS estimator

$$\gamma_{GLS} = \left(\Upsilon_\Delta' \left(\Omega^{-1} - \Omega^{-1} Z_\Delta \left(Z_\Delta' \Omega^{-1} Z_\Delta\right)^{-1} Z_\Delta' \Omega^{-1}\right) \Upsilon_\Delta\right)^{-1} \left(\Upsilon_\Delta' \left(\Omega^{-1} - \Omega^{-1} Z_\Delta \left(Z_\Delta' \Omega^{-1} Z_\Delta\right)^{-1} Z_\Delta' \Omega^{-1}\right) Y_\Delta\right)$$

where

$$\Omega^{-1}Z_\Delta\left(Z_\Delta'\Omega^{-1}Z_\Delta\right)^{-1}Z_\Delta'\Omega^{-1} = \sum\nolimits_{t_1,t_2}\left(\Sigma^{-1}C'a_{t,T}a_{t,T}'C\otimes\mathbf{1}_n\right)\frac{\Sigma}{n}\left(C'a_{t,T}a_{t,T}'C\Sigma^{-1}\otimes\mathbf{1}_n'\right)$$
$$= \Sigma^{-1}\otimes\frac{\mathbf{1}_n\mathbf{1}_n'}{n}.$$

Also let $\sigma^{t,s} = a_{t,T}'C\Sigma^{-1}C'a_{s,T}$, $\sigma_{t,s} = a_{t,T}'C\Sigma C'a_{s,T}$ and $A_{t,s} = a_{t,T}'CC'a_{s,T}$ with corresponding expressions for $\hat{\sigma}^{t,s}$ and $\hat{\sigma}_{t,s}$ by replacing Σ by $\hat{\Sigma}$. Then,

$$\gamma_{GLS} = \left(\sigma^{\tau\tau}\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^{\Delta}\right)^{-1}\left(\sum\nolimits_t\sigma^{\tau t}\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}\tilde{Y}_t\right)$$

with $M_{\mathbf{1}_n} = I_n - \frac{\mathbf{1}_n\mathbf{1}_n'}{n}$. For the variance of GLS we note that

$$\begin{aligned}
&\operatorname{Var}\left[\left(\sum\nolimits_t\sigma^{\tau t}\tilde{X}_\tau'M_{\mathbf{1}_n}\tilde{\varepsilon}_t\right)|X_\tau\right]\\
=\;&\sum\nolimits_{t_1t_2}a_{\tau,T}'C\Sigma^{-1}C'a_{t_1,T}a_{t_2,T}'C\Sigma^{-1}C'a_{\tau,T}\tilde{X}_\tau'M_{\mathbf{1}_n}E\tilde{\varepsilon}_{t_1}\tilde{\varepsilon}_{t_2}'M_{\mathbf{1}_n}\tilde{X}_\tau\\
=\;&\sum\nolimits_{t_1t_2}a_{\tau,T}'C\Sigma^{-1}C'a_{t_1,T}\tilde{\Sigma}_Ta_{t_2,T}'C\Sigma^{-1}C'a_{\tau,T}\tilde{X}_\tau'M_{\mathbf{1}_n}\tilde{X}_\tau\\
=\;&\sigma^{\tau\tau}\tilde{X}_\tau'M_{\mathbf{1}_n}\tilde{X}_\tau.
\end{aligned}$$

Let $\bar{x}_{i\tau} = x_{i\tau}-\mu_x$ and $\bar{y}_{it} = y_{it}-\mu_{yt}$. The infeasible tests can be written as

$$\bar{T}_1 = \frac{\left(\sigma^{\tau\tau}\sum_{i=1}^n\bar{x}_{i\tau}^2\right)^{-1}\left(\sum_t\sigma^{\tau t}\sum_{i=1}^n\bar{x}_{i\tau}\bar{y}_{it}\right)-\gamma_0}{\left(\sigma^{\tau\tau}\sum_{i=1}^n\bar{x}_{i\tau}^2\right)^{-1/2}}$$

and

$$\bar{T}_2 = \frac{\left(A_{\tau\tau}\sum_{i=1}^n\bar{x}_{i\tau}^2\right)^{-1}\sum_tA_{\tau t}\sum_{i=1}^n\bar{x}_{i\tau}\bar{y}_{it}-\gamma_0}{\sigma_{\tau\tau}^{1/2}\left(A_{\tau\tau}^2\sum_{i=1}^n\bar{x}_{i\tau}^2\right)^{-1/2}}.$$

As before one can write

$$T_1 = \frac{\bar{T}_1-n^{-1/2}Z}{\left(1+n^{-1/2}S\right)^{1/2}}$$

with

$$Z = \sqrt{n}\frac{\left(\hat{\sigma}^{\tau\tau}\sum_{i=1}^n\bar{x}_{i\tau}^2\right)^{-1}\left(\sum_t\hat{\sigma}^{\tau t}\sum_{i=1}^n\bar{x}_{i\tau}\bar{y}_{it}\right)-\left(\sigma^{\tau\tau}\sum_{i=1}^n\bar{x}_{i\tau}^2\right)^{-1}\left(\sum_t\sigma^{\tau t}\sum_{i=1}^n\bar{x}_{i\tau}\bar{y}_{it}\right)}{\left(\sigma^{\tau\tau}\sum_{i=1}^n\bar{x}_{i\tau}^2\right)^{-1/2}}+O_p(n^{-1/2})$$

and

$$S = \sqrt{n}\frac{\left(\left(\hat{\sigma}^{\tau\tau}\right)^{-1}-\left(\sigma^{\tau\tau}\right)^{-1}\right)\left(\sum_{i=1}^n\bar{x}_{i\tau}^2\right)^{-1}}{\left(\sigma^{\tau\tau}\sum_{i=1}^n\bar{x}_{i\tau}^2\right)^{-1}}+O_p(n^{-1/2}).$$

Define $\tilde{h} = \Sigma^{-1}C'a_\tau$, $\tilde{H} = \Sigma^{-1}-\tilde{h}\tilde{h}'/\sigma^{\tau\tau}$ and note that

$$S = \left[\frac{\tilde{h}\otimes\tilde{h}}{\sigma^{\tau\tau}}\operatorname{vec}\left(\hat{\Sigma}-\Sigma\right)+\frac{\operatorname{tr}\left(\operatorname{vec}\left(\hat{\Sigma}-\Sigma\right)\operatorname{vec}\left(\hat{\Sigma}-\Sigma\right)'\left(\tilde{H}\otimes\tilde{h}\tilde{h}'\right)\right)}{\sigma^{\tau\tau}}\right]+O_p(n^{-1/2}).$$

The last result follows from an expansion of $(\sigma^{\tau\tau})^{-1}$. Taking total derivatives we find

$$\begin{aligned}
d\left(\sigma^{\tau\tau}\right)^{-1} &= -\left(\sigma^{\tau\tau}\right)^{-2} d\sigma^{\tau\tau} \\
d^2\left(\sigma^{\tau\tau}\right)^{-1} &= 2\left(\sigma^{\tau\tau}\right)^{-3}\left(d\sigma^{\tau\tau}\right)^2 - \left(\sigma^{\tau\tau}\right)^{-2} d^2\sigma^{\tau\tau} \\
d\sigma^{\tau\tau} &= -\left(a'_{\tau,T}C\Sigma^{-1} \otimes a'_{\tau,T}C\Sigma^{-1}\right)\operatorname{vec} d\Sigma \\
d^2\sigma^{\tau\tau} &= 2\left(\operatorname{vec} d\Sigma\right)'\left(\Sigma^{-1}C'a_{\tau,T}a'_{\tau,T}C\Sigma^{-1} \otimes \Sigma^{-1}C'a_{\tau,T}a'_{\tau,T}C\Sigma^{-1}\right)\operatorname{vec} d\Sigma \\
&\quad - \left(a'_{\tau,T}C\Sigma^{-1} \otimes a'_{\tau,T}C\Sigma^{-1}\right)\operatorname{vec} d^2\Sigma
\end{aligned}$$

Substituting $d\Sigma = -\left(\Sigma - \hat{\Sigma}\right)$ and $d^2\Sigma = 0$ then leads to the expansion for S by use of the delta method.

By the same arguments as before it follows that

$$ES = n^{-1/2}\frac{\operatorname{tr}\left(V_\Sigma\left(\tilde{H} \otimes \tilde{h}\tilde{h}'\right)\right)}{\sigma^{\tau\tau}} + O(n^{-1}),$$

$$\operatorname{Var} Z = \frac{\operatorname{tr}\left(V_\Sigma\left(\tilde{H} \otimes \tilde{h}\tilde{h}'\right)\right)}{\sigma^{\tau\tau}}$$

and

$$\operatorname{Var}(S) = \frac{\operatorname{tr}\left(V_\Sigma\left(\tilde{h}\tilde{h}' \otimes \tilde{h}\tilde{h}'\right)\right)}{\left(\sigma^{\tau\tau}\right)^2}$$

such that

$$A_1(t) = \frac{1}{4}\left(1+t^2\right)\frac{\operatorname{tr}\left(V_\Sigma\left(\tilde{h}\tilde{h}' \otimes \tilde{h}\tilde{h}'\right)\right)}{\left(\sigma^{\tau\tau}\right)^2} + 2\frac{\operatorname{tr}\left(V_\Sigma\left(\tilde{H} \otimes \tilde{h}\tilde{h}'\right)\right)}{\sigma^{\tau\tau}}$$

and

$$B_1(t) = \frac{1}{4}t^2\frac{\operatorname{tr}\left(V_\Sigma\left(\tilde{h}\tilde{h}' \otimes \tilde{h}\tilde{h}'\right)\right)}{\left(\sigma^{\tau\tau}\right)^2} + \frac{\operatorname{tr}\left(V_\Sigma\left(\tilde{H} \otimes \tilde{h}\tilde{h}'\right)\right)}{\sigma^{\tau\tau}} = \frac{1}{2}t^2\frac{\left(\tilde{h}'\Sigma\tilde{h}\right)^2}{\sigma^{\tau\tau}} + r - 1 = \frac{1}{4}t^2.$$

Note that

$$\frac{\operatorname{tr}\left(V_\Sigma\left(\tilde{H} \otimes \tilde{h}\tilde{h}'\right)\right)}{\sigma^{\tau\tau}} = \frac{\operatorname{tr}\left(V_\Sigma\left(\Sigma^{-1} \otimes \tilde{h}\tilde{h}'\right)\right)}{\sigma^{\tau\tau}} - \frac{\operatorname{tr}\left(V_\Sigma\left(\tilde{h}\tilde{h}' \otimes \tilde{h}\tilde{h}'\right)\right)}{\left(\sigma^{\tau\tau}\right)^2} = r - 1$$

such that

$$A_1(t) = \frac{1}{2}\left(1+t^2\right) + (r-1).$$

For the test T_2 we see that

$$W = \frac{(\hat{\sigma}_{\tau\tau} - \sigma_{\tau\tau})(X'_\tau M_{\mathbf{1}_n} X_\tau)^{-1}}{\sigma_{\tau\tau}\left(X'_\tau M_{\mathbf{1}_n} X_\tau\right)^{-1}} = \frac{(\hat{\sigma}_{\tau\tau} - \sigma_{\tau\tau})}{\sigma_{\tau\tau}} = \frac{a'_{\tau,T}C\left(\hat{\Sigma} - \Sigma\right)C'a_{\tau,T}}{a'_{\tau,T}C\Sigma C'a_{\tau,T}}.$$

If we denote by α_t the t-th element of the vector $C'a_{\tau,T}$ then it can be seen that the same formulas as in (2.7) and (2.8) can be used to obtain the constants A_2 and B_2. In particular $x = \sqrt{n}\left(\sum_{t=1}^r a_t \otimes A_{\tau t} M_{\mathbf{1}_n}\tilde{X}_\tau\right) / \left(A_{\tau\tau}\tilde{X}'_\tau M_{\mathbf{1}_n}\tilde{X}_\tau\right)^{1/2}$ and $x_i = \sqrt{n}c'_i a_{\tau,T} M_{\mathbf{1}_n}\tilde{X}_\tau / \left(A_{\tau\tau}\tilde{X}'_\tau M_{\mathbf{1}_n}\tilde{X}_\tau\right)^{1/2}$. Also, let $Z_t = \mathbf{1}_n,\ M_{z,t} = M_{\mathbf{1}_n},\ X_t = c'_t a_{\tau,T}\ \tilde{X}_\tau$ and

$$M_{x,t} = \begin{cases} I_n - \tilde{X}_\tau\left(\tilde{X}'_\tau\tilde{X}_\tau\right)^{-1}\tilde{X}'_\tau & \text{if } X_t \neq 0 \\ I_n & \text{if } X_t = 0 \end{cases}$$

such that

$$M_t = \begin{cases} I_n - \mathbf{1}_n \left(\mathbf{1}_n M_{x,t} \mathbf{1}_n\right)^{-1} \mathbf{1}_n M_{x,t} - \tilde{X}_\tau \left(\tilde{X}_\tau' M_{\mathbf{1}_n} \tilde{X}_\tau\right)^{-1} \tilde{X}_\tau' M_{\mathbf{1}_n} & \text{if } X_t \neq 0 \\ M_{\mathbf{1}_n} & \text{if } X_t = 0 \end{cases}.$$

From $a_{\tau,T}' C \left(\hat{\Sigma} - \Sigma\right) C' a_{\tau,T} = \left(a_{\tau,T}' C \otimes a_{\tau,T}' C\right) \text{vec}\left(\hat{\Sigma} - \Sigma\right)$ it follows that

$$\text{var}\left(W|\bar{T}_2\right) = \frac{\text{tr}\left[V_\Sigma \left(C' a_{\tau,T} a_{\tau,T}' C \otimes C' a_{\tau,T} a_{\tau,T}' C\right)\right]}{\left(a_{\tau,T}' C \Sigma C' a_{\tau,T}\right)^2} = 2.$$

It then follows that

$$A_2(t, \Omega) = \frac{1}{2}\left(1 + t^2\right)$$

and

$$B_2(t, \Omega, b_2) = \frac{1}{2} t^2.$$

The results obtained here by considering approximations to unconditional versions of formulas (2.4) and (2.5) remain correct even for the conditional versions directly. To see this, note that

$$h = \frac{\left(\Sigma^{-1} C' a_{\tau,T} \otimes M_{\mathbf{1}_n} X_\tau\right)}{a_{\tau,T}' C \Sigma^{-1} C' a_{\tau,T} X_\tau' M_{\mathbf{1}_n} X_\tau}$$

such that

$$h' \left(\Sigma \otimes b_i b_j\right) h = \frac{a_{\tau,T}' C \Sigma^{-1} C' a_{\tau,T} X_\tau' M_{\mathbf{1}_n} b_i b_j' M_{\mathbf{1}_n} X_\tau}{\left(a_{\tau,T}' C \Sigma^{-1} C' a_{\tau,T} X_\tau' M_{\mathbf{1}_n} X_\tau\right)^2}$$

and

$$\textstyle\sum_{i,j=1}^n \left(h' \left(\Sigma \otimes b_i b_j\right) h\right)^2 = \frac{\sum_{i,j=1}^n X_\tau' M_{\mathbf{1}_n} b_i b_i' M_{\mathbf{1}_n} X_\tau X_\tau' M_{\mathbf{1}_n} b_j b_j M_{\mathbf{1}_n} X_\tau}{\left(a_{\tau,T}' C \Sigma^{-1} C' a_{\tau,T}\right)^2 \left(X_\tau' M_{\mathbf{1}_n} X_\tau\right)^4} = \frac{1}{\left(a_{\tau,T}' C \Sigma^{-1} C' a_{\tau,T}\right)^2 \left(X_\tau' M_{\mathbf{1}_n} X_\tau\right)^2}.$$

Also,

$$h' \Omega h = \frac{1}{a_{\tau,T}' C \Sigma^{-1} C' a_{\tau,T} X_\tau' M_{\mathbf{1}_n} X_\tau}.$$

Moreover note that

$$\begin{aligned} \text{tr}\left[\left(\Sigma \otimes b_i b_j'\right) H\right] &= \text{tr}\left(I_r \otimes b_i b_j' M_{\mathbf{1}_n}\right) - \frac{\text{tr}\left[\left(\Sigma \otimes b_i b_j'\right)\left(\Sigma^{-1} C' a_{\tau,T} a_{\tau,T}' C \Sigma^{-1} \otimes M_{\mathbf{1}_n} X_\tau X_\tau' M_{\mathbf{1}_n}\right)\right]}{a_{\tau,T}' C \Sigma^{-1} C' a_{\tau,T} X_\tau' M_{\mathbf{1}_n} X_\tau} \\ &= r b_j' M_{\mathbf{1}_n} b_i - \frac{b_j' M_{\mathbf{1}_n} X_\tau X_\tau' M_{\mathbf{1}_n} b_i}{X_\tau' M_{\mathbf{1}_n} X_\tau} \end{aligned}$$

and

$$\textstyle\sum_{i,j=1}^n r b_j' M_{\mathbf{1}_n} b_i h \left(\Sigma \otimes b_i b_j\right) h = \frac{\sum_{i,j=1}^n r X_\tau' M_{\mathbf{1}_n} b_i b_i' M_{\mathbf{1}_n} b_j b_j' M_{\mathbf{1}_n} X_\tau}{a_{\tau,T}' C \Sigma^{-1} C' a_{\tau,T} \left(X_\tau' M_{\mathbf{1}_n} X_\tau\right)^2} = \frac{r}{a_{\tau,T}' C \Sigma^{-1} C' a_{\tau,T} X_\tau' M_{\mathbf{1}_n} X_\tau},$$

$$\begin{aligned} \textstyle\sum_{i,j=1}^n \frac{b_j' M_{\mathbf{1}_n} X_\tau X_\tau' M_{\mathbf{1}_n} b_i}{X_\tau' M_{\mathbf{1}_n} X_\tau} h' \left(\Sigma \otimes b_i b_j\right) h &= \textstyle\sum_{i,j=1}^n \frac{b_j' M_{\mathbf{1}_n} X_\tau X_\tau' M_{\mathbf{1}_n} b_i X_\tau' M_{\mathbf{1}_n} b_i b_j' M_{\mathbf{1}_n} X_\tau}{a_{\tau,T}' C \Sigma^{-1} C' a_{\tau,T} \left(X_\tau' M_{\mathbf{1}_n} X_\tau\right)^3} \\ &= \frac{1}{a_{\tau,T}' C \Sigma^{-1} C' a_{\tau,T} X_\tau' M_{\mathbf{1}_n} X_\tau} \end{aligned}$$

Next consider

$$\left(h'\left(e_l e_m'\Sigma\otimes b_i b_j'\right)h\right)=\frac{a_{\tau,T}'C\Sigma^{-1}e_l e_m'C'a_{\tau,T}X_\tau' M_{\mathbf{1}_n}b_i b_j' M_{\mathbf{1}_n}X_\tau}{\left(a_{\tau,T}'C\Sigma^{-1}C'a_{\tau,T}X_\tau' M_{\mathbf{1}_n}X_\tau\right)^2},$$

$$\begin{aligned}
&\textstyle\sum_{l,m=1}^r\sum_{i,j=1}^n\left(h'\left(e_l e_m'\Sigma\otimes b_i b_j'\right)h\right)\left(h'\left(e_m e_l'\Sigma\otimes b_i b_j'\right)h\right)\\
=\;&\frac{\sum_{l,m=1}^r a_{\tau,T}'C\Sigma^{-1}e_l e_m'C'a_{\tau,T}a_{\tau,T}'C\Sigma^{-1}e_m e_l'C'a_{\tau,T}}{\left(a_{\tau,T}'C\Sigma^{-1}C'a_{\tau,T}\right)^4\left(X_\tau' M_{\mathbf{1}_n}X_\tau\right)^2}\\
=\;&\frac{\sum_{l,m=1}^r a_{\tau,T}'C\Sigma^{-1}e_l e_l'C'a_{\tau,T}a_{\tau,T}'C\Sigma^{-1}e_m e_m'C'a_{\tau,T}}{\left(a_{\tau,T}'C\Sigma^{-1}C'a_{\tau,T}\right)^4\left(X_\tau' M_{\mathbf{1}_n}X_\tau\right)^2}=\frac{1}{\left(a_{\tau,T}'C\Sigma^{-1}C'a_{\tau,T}\right)^2\left(X_\tau' M_{\mathbf{1}_n}X_\tau\right)^2}.
\end{aligned}$$

and

$$\operatorname{tr}\left(\left(e_l e_m'\Sigma\otimes b_i b_j'\right)\left(\Sigma^{-1}\otimes M_{\mathbf{1}_n}\right)\right)=\left(e_m' e_l b_j' M_{\mathbf{1}_n}b_i\right)$$

with

$$\begin{aligned}
&\textstyle\sum_{l,m=1}^r\sum_{i,j=1}^n\operatorname{tr}\left(\left(e_l e_m'\Sigma\otimes b_i b_j'\right)\left(\Sigma^{-1}\otimes M_{\mathbf{1}_n}\right)\right)\left(h'\left(e_m e_l'\Sigma\otimes b_i b_j'\right)h\right)\\
=\;&\textstyle\sum_{m=1}^r\sum_{i,j=1}^n b_j' M_{\mathbf{1}_n}b_i\left(h'\left(e_m e_m'\Sigma\otimes b_i b_j'\right)h\right)\\
=\;&\textstyle\sum_{i,j=1}^n b_j' M_{\mathbf{1}_n}b_i\left(h'\left(\Sigma\otimes b_i b_j'\right)h\right)=\frac{\sum_{i,j=1}^n X_\tau' M_{\mathbf{1}_n}b_i b_i' M_{\mathbf{1}_n}b_j b_j' M_{\mathbf{1}_n}X_\tau}{a_{\tau,T}'C\Sigma^{-1}C'a_{\tau,T}\left(X_\tau' M_{\mathbf{1}_n}X_\tau\right)^2}=\frac{1}{a_{\tau,T}'C\Sigma^{-1}C'a_{\tau,T}X_\tau' M_{\mathbf{1}_n}X_\tau}.
\end{aligned}$$

Also,

$$\begin{aligned}
&\frac{\operatorname{tr}\left[\left(e_l e_m'\Sigma\otimes b_i b_j'\right)\left(\Sigma^{-1}C'a_{\tau,T}a_{\tau,T}'C\Sigma^{-1}\otimes M_{\mathbf{1}_n}X_\tau X_\tau' M_{\mathbf{1}_n}\right)\right]}{a_{\tau,T}'C\Sigma^{-1}C'a_{\tau,T}X_\tau' M_{\mathbf{1}_n}X_\tau}\\
=\;&\frac{a_{\tau,T}'C\Sigma^{-1}e_l e_m'C'a_{\tau,T}b_j' M_{\mathbf{1}_n}X_\tau X_\tau' M_{\mathbf{1}_n}b_i}{a_{\tau,T}'C\Sigma^{-1}C'a_{\tau,T}X_\tau' M_{\mathbf{1}_n}X_\tau}
\end{aligned}$$

such that

$$\begin{aligned}
&\textstyle\sum_{l,m=1}^r\sum_{i,j=1}^n\frac{\operatorname{tr}\left[\left(e_m e_l'\Sigma\otimes b_i b_j'\right)\left(\Sigma^{-1}C'a_{\tau,T}a_{\tau,T}'C\Sigma^{-1}\otimes M_{\mathbf{1}_n}X_\tau X_\tau' M_{\mathbf{1}_n}\right)\right]}{a_{\tau,T}'C\Sigma^{-1}C'a_{\tau,T}X_\tau' M_{\mathbf{1}_n}X_\tau}\left(h'\left(e_l e_m'\Sigma\otimes b_i b_j'\right)h\right)\\
=\;&\frac{\sum_{l,m=1}^r\sum_{i,j=1}^n a_{\tau,T}'C\Sigma^{-1}e_l e_m'C'a_{\tau,T}a_{\tau,T}'C\Sigma^{-1}e_l e_m'C'a_{\tau,T}X_\tau' M_{\mathbf{1}_n}b_i b_j' M_{\mathbf{1}_n}X_\tau b_j' M_{\mathbf{1}_n}X_\tau X_\tau' M_{\mathbf{1}_n}b_i}{\left(a_{\tau,T}'C\Sigma^{-1}C'a_{\tau,T}X_\tau' M_{\mathbf{1}_n}X_\tau\right)^3}\\
=\;&\frac{\sum_{l,m=1}^r a_{\tau,T}'C\Sigma^{-1}e_m e_m'C'a_{\tau,T}a_{\tau,T}'C\Sigma^{-1}e_l e_l'C'a_{\tau,T}}{\left(a_{\tau,T}'C\Sigma^{-1}C'a_{\tau,T}\right)^3X_\tau' M_{\mathbf{1}_n}X_\tau}=\frac{1}{a_{\tau,T}'C\Sigma^{-1}C'a_{\tau,T}X_\tau' M_{\mathbf{1}_n}X_\tau}
\end{aligned}$$

and therefore

$$\textstyle\sum_{l,m=1}^r\sum_{i,j=1}^n\operatorname{tr}\left(\left(e_l e_m'\Sigma\otimes b_i b_j'\right)H\right)\left(h'\left(e_m e_l'\Sigma\otimes b_i b_j'\right)h\right)=0.$$

Collecting these results and substitution in (2.4) and (2.5) leads to

$$\frac{\operatorname{tr}V_\Omega\left(hh'\otimes hh'\right)}{\left(h'\Omega h\right)^2}=2$$

and

$$\frac{\operatorname{tr}V_\Omega\left(H\otimes hh'\right)}{h'\Omega h}=r-1.$$

Table 1a: Results for the Levels Specification

Full Sample			Estimated Covariance				Known Covariance	
			GLS-SC	GLS	ROLS	OLS	GLS	ROLS
gamma	rho	T						
0	**0**	**5**	0.046	0.064	0.05	0.052	0.056	0.048
0.1			0.056	0.082	0.072	0.07	0.082	0.07
0.6			0.612	0.692	0.574	0.596	0.64	0.6
1			0.964	0.978	0.948	0.95	0.974	0.95
0	**0**	**10**	0.036	0.074	0.048	0.05	0.048	0.044
0.1			0.07	0.134	0.072	0.076	0.074	0.078
0.6			0.788	0.876	0.844	0.85	0.866	0.854
1			0.994	0.998	0.996	0.998	0.998	0.998
0	**0**	**15**	0.058	0.152	0.048	0.042	0.04	0.038
0.1			0.096	0.226	0.09	0.088	0.096	0.086
0.6			0.874	0.936	0.934	0.938	0.95	0.942
1			1	1	1	1	1	1
0	**0**	**20**	0.058	0.204	0.04	0.046	0.048	0.046
0.1			0.118	0.24	0.11	0.106	0.114	0.114
0.6			0.954	0.994	0.988	0.996	0.998	0.996
1			1	1	1	1	1	1

3 Periods			Estimated Covariance				Known Covariance	
			GLS-SC	GLS	ROLS	OLS	GLS	ROLS
gamma	rho	T						
0	**0**	**5**	0.042	0.064	0.05	0.032	0.056	0.048
0.1			0.064	0.078	0.072	0.046	0.082	0.07
0.6			0.626	0.68	0.574	0.504	0.64	0.6
1			0.966	0.98	0.948	0.92	0.974	0.95
0	**0**	**10**	0.036	0.046	0.042	0.01	0.048	0.052
0.1			0.07	0.086	0.078	0.024	0.074	0.08
0.6			0.836	0.87	0.826	0.682	0.866	0.822
1			0.996	0.996	1	0.982	0.998	1
0	**0**	**15**	0.038	0.048	0.05	0.016	0.04	0.056
0.1			0.084	0.1	0.078	0.024	0.096	0.08
0.6			0.922	0.938	0.872	0.768	0.95	0.892
1			1	1	1	0.998	1	1
0	**0**	**20**	0.022	0.04	0.048	0.008	0.048	0.056
0.1			0.1	0.122	0.098	0.03	0.114	0.096
0.6			0.994	0.996	0.936	0.824	0.998	0.946
1			1	1	1	1	1	1

2 Periods			Estimated Covariance				Known Covariance	
			GLS-SC	GLS	ROLS	OLS	GLS	ROLS
gamma	rho	T						
0	**0**	**5**	0.046	0.05	0.05	0.056	0.048	0.048
0.1			0.066	0.072	0.072	0.076	0.07	0.07
0.6			0.564	0.574	0.574	0.586	0.6	0.6
1			0.944	0.948	0.948	0.946	0.95	0.95
0	**0**	**10**	0.044	0.048	0.048	0.052	0.044	0.044
0.1			0.068	0.072	0.072	0.076	0.078	0.078
0.6			0.828	0.844	0.844	0.85	0.854	0.854
1			0.996	0.996	0.996	0.996	0.998	0.998
0	**0**	**15**	0.044	0.048	0.048	0.05	0.038	0.038
0.1			0.08	0.09	0.09	0.094	0.086	0.086
0.6			0.93	0.934	0.934	0.94	0.942	0.942
1			1	1	1	1	1	1
0	**0**	**20**	0.03	0.04	0.04	0.046	0.046	0.046
0.1			0.104	0.11	0.11	0.126	0.114	0.114
0.6			0.988	0.988	0.988	0.992	0.996	0.996
1			1	1	1	1	1	1

GLS-SC is the feasible GLS estimator with the size correction of Section 3. GLS is the feasible GLS estimator without size correction.
ROLS is the standard OLS estimator with robust standard errors. The covariance matrix is computed in the same way as for GLS
OLS is the usual OLS estimator with the standard variance estimator.

Table 1b: Results for the Levels Specification

Full Sample			Estimated Covariance				Known Covariance	
			GLS-SC	GLS	ROLS	OLS	GLS	ROLS
gamma	rho	T						
0	0.4	5	0.056	0.074	0.07	0.1	0.082	0.064
0.1			0.07	0.088	0.066	0.114	0.076	0.07
0.6			0.514	0.574	0.51	0.612	0.524	0.512
1			0.916	0.94	0.896	0.946	0.934	0.904
0	0.4	10	0.036	0.08	0.046	0.146	0.046	0.042
0.1			0.058	0.118	0.066	0.168	0.068	0.07
0.6			0.596	0.71	0.616	0.792	0.68	0.61
1			0.958	0.974	0.956	0.988	0.976	0.964
0	0.4	15	0.06	0.164	0.05	0.152	0.044	0.048
0.1			0.094	0.168	0.064	0.204	0.07	0.074
0.6			0.682	0.812	0.698	0.864	0.764	0.706
1			0.968	0.99	0.988	0.996	0.994	0.986
0	0.4	20	0.044	0.166	0.038	0.18	0.048	0.034
0.1			0.078	0.184	0.084	0.218	0.074	0.078
0.6			0.784	0.904	0.808	0.932	0.862	0.826
1			0.998	1	1	1	1	1

3 Periods			Estimated Covariance				Known Covariance	
			GLS-SC	GLS	ROLS	OLS	GLS	ROLS
gamma	rho	T						
0	0.4	5	0.05	0.066	0.07	0.072	0.058	0.064
0.1			0.068	0.08	0.066	0.082	0.072	0.07
0.6			0.494	0.54	0.51	0.526	0.516	0.512
1			0.92	0.93	0.896	0.904	0.912	0.904
0	0.4	10	0.03	0.046	0.042	0.026	0.046	0.042
0.1			0.064	0.076	0.076	0.064	0.074	0.074
0.6			0.592	0.662	0.628	0.606	0.656	0.638
1			0.95	0.96	0.962	0.958	0.966	0.964
0	0.4	15	0.048	0.064	0.056	0.038	0.048	0.052
0.1			0.054	0.088	0.08	0.056	0.078	0.08
0.6			0.678	0.712	0.7	0.64	0.722	0.712
1			0.986	0.99	0.978	0.968	0.99	0.982
0	0.4	20	0.028	0.048	0.046	0.024	0.042	0.046
0.1			0.074	0.088	0.086	0.034	0.08	0.082
0.6			0.794	0.818	0.786	0.692	0.828	0.776
1			1	1	1	0.99	1	1

2 Periods			Estimated Covariance				Known Covariance	
			GLS-SC	GLS	ROLS	OLS	GLS	ROLS
gamma	rho	T						
0	0.4	5	0.062	0.07	0.07	0.066	0.064	0.064
0.1			0.066	0.066	0.066	0.072	0.07	0.07
0.6			0.48	0.51	0.51	0.516	0.512	0.512
1			0.888	0.896	0.896	0.898	0.904	0.904
0	0.4	10	0.036	0.046	0.046	0.046	0.042	0.042
0.1			0.052	0.066	0.066	0.064	0.07	0.07
0.6			0.604	0.616	0.616	0.622	0.61	0.61
1			0.952	0.956	0.956	0.956	0.964	0.964
0	0.4	15	0.05	0.05	0.05	0.054	0.048	0.048
0.1			0.06	0.064	0.064	0.068	0.074	0.074
0.6			0.68	0.698	0.698	0.704	0.706	0.706
1			0.988	0.988	0.988	0.988	0.986	0.986
0	0.4	20	0.032	0.038	0.038	0.04	0.034	0.034
0.1			0.076	0.084	0.084	0.086	0.078	0.078
0.6			0.792	0.808	0.808	0.816	0.826	0.826
1			1	1	1	1	1	1

Table 1c: Results for the Levels Specification

Full Sample			Estimated Covariance				Known Covariance	
			GLS-SC	GLS	ROLS	OLS	GLS	ROLS
gamma	rho	T						
0	0.8	5	0.05	0.08	0.064	0.144	0.068	0.056
0.1			0.07	0.092	0.068	0.16	0.09	0.074
0.6			0.476	0.52	0.42	0.628	0.504	0.444
1			0.87	0.904	0.836	0.94	0.91	0.852
0	0.8	10	0.044	0.082	0.038	0.226	0.04	0.04
0.1			0.052	0.104	0.042	0.252	0.06	0.046
0.6			0.492	0.61	0.352	0.66	0.56	0.35
1			0.89	0.934	0.724	0.928	0.938	0.724
0	0.8	15	0.078	0.152	0.058	0.33	0.056	0.044
0.1			0.07	0.192	0.062	0.366	0.062	0.078
0.6			0.478	0.674	0.328	0.694	0.58	0.318
1			0.874	0.928	0.67	0.92	0.936	0.692
0	0.8	20	0.066	0.18	0.052	0.364	0.062	0.05
0.1			0.08	0.184	0.06	0.378	0.066	0.046
0.6			0.526	0.714	0.294	0.712	0.6	0.286
1			0.896	0.964	0.676	0.926	0.962	0.68

3 Periods			Estimated Covariance				Known Covariance	
			GLS-SC	GLS	ROLS	OLS	GLS	ROLS
gamma	rho	T						
0	0.8	5	0.056	0.08	0.064	0.084	0.066	0.056
0.1			0.068	0.084	0.068	0.114	0.076	0.074
0.6			0.444	0.472	0.42	0.508	0.474	0.444
1			0.848	0.876	0.836	0.89	0.872	0.852
0	0.8	10	0.024	0.04	0.034	0.066	0.038	0.04
0.1			0.036	0.046	0.042	0.082	0.046	0.044
0.6			0.352	0.406	0.364	0.48	0.39	0.382
1			0.736	0.746	0.748	0.81	0.752	0.738
0	0.8	15	0.05	0.06	0.054	0.092	0.046	0.044
0.1			0.056	0.084	0.074	0.114	0.072	0.076
0.6			0.334	0.37	0.348	0.434	0.352	0.334
1			0.682	0.718	0.702	0.78	0.72	0.72
0	0.8	20	0.038	0.062	0.046	0.076	0.054	0.054
0.1			0.042	0.058	0.054	0.088	0.046	0.046
0.6			0.286	0.35	0.342	0.406	0.324	0.326
1			0.674	0.714	0.698	0.754	0.696	0.696

2 Periods			Estimated Covariance				Known Covariance	
			GLS-SC	GLS	ROLS	OLS	GLS	ROLS
gamma	rho	T						
0	0.8	5	0.054	0.064	0.064	0.06	0.056	0.056
0.1			0.064	0.068	0.068	0.076	0.074	0.074
0.6			0.404	0.42	0.42	0.424	0.444	0.444
1			0.826	0.836	0.836	0.84	0.852	0.852
0	0.8	10	0.032	0.038	0.038	0.038	0.04	0.04
0.1			0.038	0.042	0.042	0.046	0.046	0.046
0.6			0.332	0.352	0.352	0.356	0.35	0.35
1			0.712	0.724	0.724	0.728	0.724	0.724
0	0.8	15	0.054	0.058	0.058	0.058	0.044	0.044
0.1			0.054	0.062	0.062	0.066	0.078	0.078
0.6			0.314	0.328	0.328	0.326	0.318	0.318
1			0.654	0.67	0.67	0.67	0.692	0.692
0	0.8	20	0.044	0.052	0.052	0.052	0.05	0.05
0.1			0.054	0.06	0.06	0.062	0.046	0.046
0.6			0.282	0.294	0.294	0.296	0.286	0.286
1			0.654	0.676	0.676	0.674	0.68	0.68

Table 1d: Results for the Levels Specification

Full Sample			Estimated Covariance				Known Covariance	
			GLS-SC	GLS	ROLS	OLS	GLS	ROLS
gamma	rho	T						
0	0.9	5	0.048	0.072	0.056	0.144	0.064	0.058
0.1			0.07	0.096	0.076	0.184	0.084	0.06
0.6			0.476	0.53	0.406	0.61	0.514	0.428
1			0.882	0.916	0.82	0.932	0.918	0.834
0	0.9	10	0.044	0.094	0.024	0.276	0.042	0.028
0.1			0.058	0.108	0.04	0.272	0.064	0.034
0.6			0.478	0.59	0.306	0.632	0.558	0.282
1			0.866	0.928	0.618	0.878	0.94	0.626
0	0.9	15	0.088	0.166	0.054	0.39	0.056	0.056
0.1			0.084	0.19	0.072	0.402	0.07	0.072
0.6			0.47	0.642	0.278	0.64	0.572	0.274
1			0.846	0.914	0.546	0.866	0.922	0.55
0	0.9	20	0.084	0.208	0.062	0.406	0.062	0.056
0.1			0.096	0.234	0.058	0.426	0.056	0.058
0.6			0.476	0.654	0.196	0.688	0.584	0.192
1			0.83	0.928	0.476	0.858	0.942	0.476

3 Periods			Estimated Covariance				Known Covariance	
			GLS-SC	GLS	ROLS	OLS	GLS	ROLS
gamma	rho	T						
0	0.9	5	0.052	0.072	0.056	0.092	0.066	0.058
0.1			0.07	0.078	0.076	0.114	0.074	0.06
0.6			0.444	0.484	0.406	0.522	0.488	0.428
1			0.834	0.868	0.82	0.882	0.874	0.834
0	0.9	10	0.032	0.04	0.026	0.074	0.038	0.024
0.1			0.038	0.046	0.038	0.074	0.044	0.038
0.6			0.296	0.334	0.308	0.42	0.328	0.3
1			0.664	0.706	0.66	0.774	0.72	0.656
0	0.9	15	0.046	0.062	0.05	0.108	0.042	0.056
0.1			0.054	0.068	0.07	0.132	0.054	0.062
0.6			0.294	0.334	0.28	0.396	0.312	0.292
1			0.566	0.62	0.57	0.684	0.602	0.572
0	0.9	20	0.044	0.054	0.054	0.094	0.046	0.044
0.1			0.044	0.06	0.052	0.098	0.046	0.052
0.6			0.192	0.232	0.216	0.322	0.21	0.212
1			0.528	0.588	0.542	0.65	0.554	0.532

2 Periods			Estimated Covariance				Known Covariance	
			GLS-SC	GLS	ROLS	OLS	GLS	ROLS
gamma	rho	T						
0	0.9	5	0.048	0.056	0.056	0.056	0.058	0.058
0.1			0.07	0.076	0.076	0.084	0.06	0.06
0.6			0.398	0.406	0.406	0.416	0.428	0.428
1			0.812	0.82	0.82	0.83	0.834	0.834
0	0.9	10	0.022	0.024	0.024	0.03	0.028	0.028
0.1			0.03	0.04	0.04	0.046	0.034	0.034
0.6			0.278	0.306	0.306	0.314	0.282	0.282
1			0.588	0.618	0.618	0.626	0.626	0.626
0	0.9	15	0.048	0.054	0.054	0.054	0.056	0.056
0.1			0.068	0.072	0.072	0.072	0.072	0.072
0.6			0.264	0.278	0.278	0.278	0.274	0.274
1			0.53	0.546	0.546	0.548	0.55	0.55
0	0.9	20	0.052	0.062	0.062	0.062	0.056	0.056
0.1			0.052	0.058	0.058	0.062	0.058	0.058
0.6			0.182	0.196	0.196	0.194	0.192	0.192
1			0.448	0.476	0.476	0.478	0.476	0.476

Table 2a: Results for the First Difference Specification

Full Sample			Estimated Covariance					Known Covariance	
			GLS-SC	GLS	GLS-AR	ROLS	OLS	GLS	ROLS
gamma	rho	T							
0	0	5	0.046	0.064	0.066	0.064	0.062	0.056	0.06
0.1			0.056	0.082	0.086	0.074	0.08	0.082	0.074
0.6			0.612	0.692	0.666	0.346	0.358	0.64	0.348
1			0.964	0.978	0.978	0.696	0.692	0.974	0.694
0	0	10	0.036	0.074	0.046	0.048	0.05	0.048	0.05
0.1			0.07	0.134	0.074	0.068	0.064	0.074	0.06
0.6			0.788	0.876	0.868	0.326	0.34	0.866	0.33
1			0.994	0.998	0.998	0.692	0.69	0.998	0.694
0	0	15	0.058	0.152	0.044	0.054	0.058	0.04	0.058
0.1			0.096	0.226	0.096	0.058	0.07	0.096	0.068
0.6			0.874	0.936	0.948	0.34	0.33	0.95	0.328
1			1	1	1	0.704	0.718	1	0.712
0	0	20	0.058	0.204	0.048	0.046	0.04	0.048	0.034
0.1			0.118	0.24	0.11	0.048	0.062	0.114	0.056
0.6			0.954	0.994	0.998	0.33	0.318	0.998	0.312
1			1	1	1	0.726	0.714	1	0.708

3 Periods			Estimated Covariance					Known Covariance	
			GLS-SC	GLS	GLS-AR	ROLS	OLS	GLS	ROLS
gamma	rho	T							
0	0	5	0.052	0.064	0.054	0.064	0.126	0.058	0.06
0.1			0.064	0.078	0.058	0.074	0.132	0.064	0.074
0.6			0.548	0.586	0.292	0.346	0.448	0.57	0.348
1			0.92	0.95	0.646	0.696	0.764	0.95	0.694
0	0	10	0.038	0.046	0.042	0.048	0.208	0.05	0.05
0.1			0.062	0.072	0.048	0.068	0.236	0.07	0.06
0.6			0.522	0.548	0.306	0.326	0.564	0.56	0.33
1			0.91	0.924	0.67	0.692	0.872	0.922	0.694
0	0	15	0.05	0.06	0.054	0.054	0.222	0.066	0.058
0.1			0.054	0.074	0.066	0.058	0.24	0.074	0.068
0.6			0.544	0.606	0.312	0.34	0.626	0.576	0.328
1			0.912	0.938	0.69	0.704	0.898	0.934	0.712
0	0	20	0.036	0.04	0.034	0.046	0.248	0.042	0.034
0.1			0.046	0.06	0.052	0.048	0.266	0.058	0.056
0.6			0.53	0.574	0.306	0.33	0.658	0.586	0.312
1			0.956	0.958	0.696	0.726	0.92	0.97	0.708

2 Periods			Estimated Covariance					Known Covariance	
			GLS-SC	GLS	GLS-AR	ROLS	OLS	GLS	ROLS
gamma	rho	T							
0	0	5	0.042	0.048	0.044	0.042	0.036	0.046	0.03
0.1			0.054	0.066	0.056	0.058	0.062	0.058	0.05
0.6			0.416	0.45	0.434	0.34	0.35	0.434	0.358
1			0.818	0.856	0.854	0.708	0.712	0.856	0.702
0	0	10	0.058	0.074	0.07	0.068	0.14	0.07	0.064
0.1			0.07	0.09	0.082	0.08	0.134	0.082	0.066
0.6			0.394	0.436	0.416	0.31	0.478	0.416	0.292
1			0.788	0.812	0.81	0.668	0.81	0.81	0.674
0	0	15	0.052	0.062	0.05	0.066	0.138	0.05	0.058
0.1			0.048	0.064	0.056	0.07	0.154	0.056	0.062
0.6			0.36	0.39	0.382	0.296	0.502	0.382	0.284
1			0.822	0.83	0.83	0.704	0.844	0.832	0.716
0	0	20	0.04	0.05	0.04	0.05	0.146	0.04	0.048
0.1			0.054	0.064	0.064	0.056	0.174	0.064	0.06
0.6			0.384	0.414	0.41	0.334	0.492	0.41	0.328
1			0.824	0.85	0.838	0.706	0.848	0.838	0.71

GLS-SC is the feasible GLS estimator with the size correction of Section 3. GLS is the feasible GLS estimator without size correction.
ROLS is the standard OLS estimator with robust standard errors. The covariance matrix is computed in the same way as for GLS
OLS is the usual OLS estimator with the standard variance estimator.

Table 2b: Results for the First Difference Specification

Full Sample			Estimated Covariance					Known Covariance	
			GLS-SC	GLS	GLS-AR	ROLS	OLS	GLS	ROLS
gamma	**rho**	**T**							
0	**0.4**	**5**	0.056	0.074	0.084	0.078	0.07	0.082	0.064
0.1			0.07	0.088	0.092	0.086	0.08	0.076	0.086
0.6			0.514	0.574	0.554	0.43	0.438	0.524	0.434
1			0.916	0.94	0.936	0.802	0.822	0.934	0.81
0	**0.4**	**10**	0.036	0.08	0.05	0.052	0.044	0.046	0.054
0.1			0.058	0.118	0.08	0.06	0.066	0.068	0.064
0.6			0.596	0.71	0.688	0.416	0.432	0.68	0.426
1			0.958	0.974	0.98	0.822	0.826	0.976	0.828
0	**0.4**	**15**	0.06	0.164	0.048	0.048	0.054	0.044	0.056
0.1			0.094	0.168	0.078	0.054	0.066	0.07	0.066
0.6			0.682	0.812	0.764	0.428	0.424	0.764	0.422
1			0.968	0.99	0.992	0.83	0.842	0.994	0.836
0	**0.4**	**20**	0.044	0.166	0.044	0.044	0.046	0.048	0.046
0.1			0.078	0.184	0.072	0.062	0.058	0.074	0.06
0.6			0.784	0.904	0.86	0.436	0.432	0.862	0.432
1			0.998	1	1	0.844	0.84	1	0.846

3 Periods			Estimated Covariance					Known Covariance	
			GLS-SC	GLS	GLS-AR	ROLS	OLS	GLS	ROLS
gamma	**rho**	**T**							
0	**0.4**	**5**	0.058	0.068	0.084	0.078	0.114	0.078	0.064
0.1			0.06	0.084	0.086	0.086	0.122	0.082	0.086
0.6			0.5	0.544	0.524	0.43	0.512	0.522	0.434
1			0.908	0.928	0.938	0.802	0.866	0.934	0.81
0	**0.4**	**10**	0.026	0.044	0.05	0.052	0.194	0.05	0.054
0.1			0.064	0.072	0.086	0.06	0.224	0.084	0.064
0.6			0.532	0.576	0.584	0.416	0.68	0.582	0.426
1			0.942	0.956	0.954	0.822	0.942	0.952	0.828
0	**0.4**	**15**	0.056	0.068	0.064	0.048	0.21	0.062	0.056
0.1			0.058	0.074	0.08	0.054	0.248	0.076	0.066
0.6			0.576	0.616	0.602	0.428	0.732	0.594	0.422
1			0.934	0.944	0.95	0.83	0.956	0.95	0.836
0	**0.4**	**20**	0.044	0.054	0.044	0.044	0.232	0.042	0.046
0.1			0.058	0.074	0.062	0.062	0.242	0.064	0.06
0.6			0.59	0.634	0.608	0.436	0.754	0.61	0.432
1			0.96	0.966	0.972	0.844	0.972	0.974	0.846

2 Periods			Estimated Covariance					Known Covariance	
			GLS-SC	GLS	GLS-AR	ROLS	OLS	GLS	ROLS
gamma	**rho**	**T**							
0	**0.4**	**5**	0.042	0.054	0.06	0.042	0.054	0.054	0.046
0.1			0.048	0.054	0.058	0.052	0.058	0.05	0.06
0.6			0.394	0.414	0.418	0.352	0.358	0.388	0.36
1			0.798	0.832	0.84	0.736	0.728	0.814	0.716
0	**0.4**	**10**	0.056	0.072	0.074	0.056	0.13	0.07	0.066
0.1			0.066	0.072	0.074	0.064	0.138	0.07	0.074
0.6			0.354	0.376	0.362	0.28	0.44	0.356	0.282
1			0.728	0.746	0.746	0.652	0.774	0.74	0.644
0	**0.4**	**15**	0.04	0.044	0.042	0.046	0.126	0.042	0.04
0.1			0.048	0.062	0.052	0.052	0.126	0.048	0.052
0.6			0.324	0.348	0.336	0.256	0.444	0.346	0.256
1			0.724	0.746	0.744	0.63	0.796	0.742	0.616
0	**0.4**	**20**	0.042	0.046	0.048	0.048	0.142	0.048	0.046
0.1			0.056	0.064	0.056	0.062	0.144	0.054	0.064
0.6			0.324	0.36	0.34	0.29	0.436	0.332	0.288
1			0.744	0.758	0.764	0.616	0.798	0.758	0.62

Table 2c: Results for the First Difference Specification

Full Sample			Estimated Covariance					Known Covariance	
			GLS-SC	GLS	GLS-AR	ROLS	OLS	GLS	ROLS
gamma	rho	T							
0	0.8	5	0.05	0.08	0.086	0.068	0.068	0.068	0.068
0.1			0.07	0.092	0.094	0.084	0.092	0.09	0.086
0.6			0.476	0.52	0.518	0.502	0.506	0.504	0.5
1			0.87	0.904	0.914	0.886	0.898	0.91	0.9
0	0.8	10	0.044	0.082	0.046	0.05	0.05	0.04	0.056
0.1			0.052	0.104	0.064	0.072	0.072	0.06	0.066
0.6			0.492	0.61	0.58	0.526	0.538	0.56	0.548
1			0.89	0.934	0.948	0.916	0.918	0.938	0.918
0	0.8	15	0.078	0.152	0.06	0.048	0.05	0.056	0.054
0.1			0.07	0.192	0.072	0.052	0.07	0.062	0.07
0.6			0.478	0.674	0.584	0.542	0.54	0.58	0.534
1			0.874	0.928	0.936	0.902	0.914	0.936	0.91
0	0.8	20	0.066	0.18	0.062	0.044	0.034	0.062	0.034
0.1			0.08	0.184	0.072	0.06	0.066	0.066	0.068
0.6			0.526	0.714	0.608	0.552	0.548	0.6	0.542
1			0.896	0.964	0.958	0.922	0.93	0.962	0.928

3 Periods			Estimated Covariance					Known Covariance	
			GLS-SC	GLS	GLS-AR	ROLS	OLS	GLS	ROLS
gamma	rho	T							
0	0.8	5	0.054	0.066	0.082	0.068	0.1	0.066	0.068
0.1			0.074	0.092	0.096	0.084	0.122	0.09	0.086
0.6			0.474	0.514	0.52	0.502	0.568	0.508	0.5
1			0.888	0.904	0.912	0.886	0.926	0.91	0.9
0	0.8	10	0.034	0.05	0.044	0.05	0.158	0.042	0.056
0.1			0.05	0.066	0.066	0.072	0.198	0.06	0.066
0.6			0.508	0.56	0.56	0.526	0.734	0.558	0.548
1			0.914	0.926	0.94	0.916	0.966	0.938	0.918
0	0.8	15	0.062	0.07	0.058	0.048	0.204	0.05	0.054
0.1			0.056	0.078	0.076	0.052	0.218	0.074	0.07
0.6			0.524	0.578	0.572	0.542	0.774	0.576	0.534
1			0.9	0.918	0.92	0.902	0.968	0.924	0.91
0	0.8	20	0.038	0.05	0.04	0.044	0.22	0.042	0.034
0.1			0.058	0.062	0.056	0.06	0.248	0.054	0.068
0.6			0.546	0.592	0.602	0.552	0.794	0.6	0.542
1			0.936	0.954	0.956	0.922	0.988	0.956	0.928

2 Periods			Estimated Covariance					Known Covariance	
			GLS-SC	GLS	GLS-AR	ROLS	OLS	GLS	ROLS
gamma	rho	T							
0	0.8	5	0.054	0.056	0.074	0.048	0.064	0.06	0.054
0.1			0.052	0.064	0.098	0.056	0.07	0.072	0.074
0.6			0.33	0.364	0.386	0.344	0.35	0.342	0.338
1			0.708	0.728	0.768	0.71	0.722	0.738	0.726
0	0.8	10	0.04	0.048	0.064	0.046	0.1	0.052	0.046
0.1			0.05	0.06	0.064	0.058	0.104	0.054	0.058
0.6			0.25	0.278	0.32	0.27	0.384	0.286	0.264
1			0.552	0.586	0.62	0.538	0.694	0.596	0.548
0	0.8	15	0.038	0.056	0.054	0.052	0.106	0.044	0.046
0.1			0.056	0.058	0.066	0.05	0.128	0.048	0.046
0.6			0.242	0.254	0.256	0.212	0.342	0.24	0.21
1			0.5	0.56	0.578	0.478	0.676	0.546	0.474
0	0.8	20	0.034	0.042	0.046	0.036	0.102	0.038	0.04
0.1			0.054	0.058	0.06	0.044	0.112	0.054	0.044
0.6			0.188	0.206	0.23	0.206	0.318	0.226	0.2
1			0.47	0.494	0.518	0.428	0.572	0.51	0.432

Table 2d: Results for the First Difference Specification

Full Sample			Estimated Covariance					Known Covariance	
			GLS-SC	GLS	GLS-AR	ROLS	OLS	GLS	ROLS
gamma	**rho**	**T**							
0	**0.9**	**5**	0.048	0.072	0.068	0.064	0.064	0.064	0.062
0.1			0.07	0.096	0.1	0.09	0.094	0.084	0.094
0.6			0.476	0.53	0.524	0.5	0.52	0.514	0.522
1			0.882	0.916	0.904	0.898	0.908	0.918	0.918
0	**0.9**	**10**	0.044	0.094	0.04	0.044	0.044	0.042	0.048
0.1			0.058	0.108	0.06	0.068	0.07	0.064	0.068
0.6			0.478	0.59	0.554	0.538	0.564	0.558	0.564
1			0.866	0.928	0.944	0.914	0.928	0.94	0.928
0	**0.9**	**15**	0.088	0.166	0.062	0.042	0.046	0.056	0.042
0.1			0.084	0.19	0.068	0.064	0.076	0.07	0.078
0.6			0.47	0.642	0.582	0.54	0.556	0.572	0.554
1			0.846	0.914	0.926	0.9	0.92	0.922	0.92
0	**0.9**	**20**	0.084	0.208	0.062	0.04	0.044	0.062	0.044
0.1			0.096	0.234	0.056	0.064	0.052	0.056	0.06
0.6			0.476	0.654	0.572	0.534	0.562	0.584	0.562
1			0.83	0.928	0.948	0.922	0.94	0.942	0.938

3 Periods			Estimated Covariance					Known Covariance	
			GLS-SC	GLS	GLS-AR	ROLS	OLS	GLS	ROLS
gamma	**rho**	**T**							
0	**0.9**	**5**	0.052	0.068	0.074	0.064	0.096	0.064	0.062
0.1			0.074	0.092	0.096	0.09	0.124	0.084	0.094
0.6			0.48	0.526	0.522	0.5	0.584	0.514	0.522
1			0.886	0.918	0.898	0.898	0.944	0.918	0.918
0	**0.9**	**10**	0.034	0.052	0.038	0.044	0.158	0.042	0.048
0.1			0.044	0.062	0.054	0.068	0.178	0.064	0.068
0.6			0.504	0.562	0.546	0.538	0.74	0.554	0.564
1			0.906	0.924	0.938	0.914	0.974	0.94	0.928
0	**0.9**	**15**	0.056	0.07	0.054	0.042	0.196	0.054	0.042
0.1			0.06	0.072	0.076	0.064	0.226	0.072	0.078
0.6			0.524	0.57	0.566	0.54	0.782	0.57	0.554
1			0.886	0.9	0.914	0.9	0.974	0.916	0.92
0	**0.9**	**20**	0.032	0.042	0.058	0.04	0.216	0.056	0.044
0.1			0.042	0.058	0.052	0.064	0.244	0.058	0.06
0.6			0.514	0.558	0.572	0.534	0.812	0.578	0.562
1			0.918	0.932	0.95	0.922	0.99	0.95	0.938

2 Periods			Estimated Covariance					Known Covariance	
			GLS-SC	GLS	GLS-AR	ROLS	OLS	GLS	ROLS
gamma	**rho**	**T**							
0	**0.9**	**5**	0.05	0.054	0.088	0.05	0.056	0.056	0.054
0.1			0.054	0.062	0.08	0.052	0.068	0.06	0.074
0.6			0.304	0.328	0.366	0.328	0.342	0.344	0.354
1			0.69	0.72	0.766	0.714	0.73	0.732	0.718
0	**0.9**	**10**	0.038	0.052	0.062	0.046	0.1	0.044	0.048
0.1			0.038	0.046	0.068	0.05	0.098	0.048	0.05
0.6			0.23	0.248	0.298	0.248	0.366	0.27	0.25
1			0.504	0.548	0.584	0.542	0.64	0.556	0.55
0	**0.9**	**15**	0.044	0.056	0.064	0.044	0.12	0.046	0.046
0.1			0.058	0.064	0.08	0.046	0.128	0.056	0.044
0.6			0.208	0.224	0.25	0.218	0.312	0.222	0.214
1			0.438	0.478	0.496	0.438	0.626	0.464	0.434
0	**0.9**	**20**	0.034	0.042	0.05	0.038	0.102	0.04	0.038
0.1			0.04	0.048	0.058	0.048	0.106	0.052	0.046
0.6			0.168	0.182	0.19	0.16	0.28	0.182	0.16
1			0.366	0.402	0.432	0.37	0.512	0.406	0.378

Table 3: Nonstationary Design with Empirical Variance Covariance Matrix

	Full Sample		3 Periods		2 Periods	
	GLS-SC	**GLS-AR**	**GLS-SC**	**GLS-AR**	**GLS-SC**	**GLS-AR**
gamma						
	Level Specification					
0	0.056	0.222	0.048	0.21	0.048	0.196
0.1	0.054	0.272	0.052	0.21	0.044	0.23
0.6	0.372	0.718	0.404	0.686	0.396	0.706
1	0.798	0.952	0.81	0.936	0.812	0.944
	First Difference Specification					
0	0.056	0.22	0.044	0.206	0.048	0.23
0.1	0.054	0.268	0.056	0.222	0.048	0.248
0.6	0.372	0.716	0.286	0.436	0.238	0.468
1	0.798	0.95	0.62	0.696	0.48	0.708

GLS-SC is the feasible GLS estimator with the size correction of Section 3. GLS-AR is feasible GLS where the weight matrix is computed from a parametric AR-1 model and bias corrected as described in Section 4.
The sample size for the simulations in this table is T=11 and n=50