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

Jerry Hausman* Guido Kuersteiner

MIT UC Davis

November 2007

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 formal 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.

JEL Codes: C12, C21, C23

Keywords: panel data, treatment effect, GLS, size corrected tests, Edgeworth expansion.

*Corresponding author: MIT, Department of Economics, 50 Memorial Drive, Cambridge, MA 02142; (617) 253-3644; email: jhausman@mit.edu.

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 variables, 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 states 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 stochastic disturbances are serially correlated. 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 an exogenous policy variable or is a dummy variable for an exogenous change in policy or regulation and z_{it} are exogenous covariates which are allowed to be constant across i to accomodate time fixed effects. The parameter θ is the parameter vector for the covariates z_{it}, 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} such that the matrix $\tilde{\Sigma}$ with elements

$\tilde{\sigma}_{t,s}$ of serial covariance parameters $\tilde{\sigma}_{t,s} = \mathrm{Cov}(\varepsilon_{it}, \varepsilon_{is})$ 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 $\tilde{\Sigma}$, as Bertrand et.al. (2004) 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 $\tilde{\Sigma}$ 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 $\tilde{\Sigma}$ is unknown and needs to be estimated, the usual estimator would be "feasible" GLS (FGLS) where a consistent estimate $\widehat{\tilde{\Sigma}}$ replaces $\tilde{\Sigma}$ in the GLS formula. Indeed, if the estimate of $\tilde{\Sigma}$ 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 $\widehat{\tilde{\Sigma}}$ 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 $\mathrm{plim}\left[\sqrt{n}\left(\hat{\delta}_{GLS} - \hat{\delta}_{FGLS}\right)\right] = 0$ where $\delta = (\theta', \gamma)'$ so long as $\mathrm{plim}\,\widehat{\tilde{\Sigma}} = \tilde{\Sigma}$. 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 $\tilde{\Sigma}$ is 55 compared to a sample size of 500. Thus, in this paper we use the second order Edgeworth approximation approach of Rothenberg (1988) that accounts for the uncertainty in estimating $\widehat{\tilde{\Sigma}}$. We derive explicit formulas for the size distortions of test statistics based on FGLS and use these formulas to construct size corrected tests which we denote by FGLS-SC. Without these size corrections 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

[1]Bertrand et. al. (2004) in their survey of the literature find only one paper out of nearly 100 papers that uses this approach where Σ is unrestricted.

adjustment for the estimation of $\hat{\tilde{\Sigma}}$ is needed.

Once we consider the effects of uncertainty in the estimate of $\hat{\tilde{\Sigma}}$, 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 $\tilde{\Sigma}$ increases at rate T^2, aggregation to reduce the dimension of $\tilde{\Sigma}$ 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 $\hat{\tilde{\Sigma}}$, we analyze the effects of temporal aggregation on size and power using the Edgeworth expansions of Rothenberg (1988). We demonstrate that for a commonly occurring situation where once the treatment begins in a state it continues thereafter, that small sample improvements in terms of local power from time aggregation can arise, but only under the somewhat unrealistic assumption that the degree of serial correlation is small and that this fact is known to the investigator.

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 $\tilde{\Sigma}$. 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. Heteroskedasticity robust inference for the OLS estimator of (1.1) has been considered by Arellano (1987), Kézdi (2004) and Bertrand et.al. (2004). We focus on the case of homoskedastic errors in this paper and leave the development of efficient, heteorskedasticity robust inference procedures for future research.

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. We focus on the single treatment date situation in our Monte Carlo experiments and empirical application.

Our 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 is less robust to misspecification and will tend to 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 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

$$y_{it} = T_{it}\gamma + z_{it}'\theta + \alpha_i + \varepsilon_{it} \tag{2.1}$$

where we define $\tilde{Y}_t = [y_{1t}, ..., y_{nt}]'$, $\tilde{Y} = \left[\tilde{Y}_1', ..., \tilde{Y}_T'\right]'$, $\tilde{Z}_t = [z_{1t}, ..., z_{nt}]'$, $\tilde{Z} = \left[\tilde{Z}_1', ..., \tilde{Z}_T'\right]'$, $\tilde{\varepsilon}_t = [\varepsilon_{1t}, ..., \varepsilon_{nt}]'$ and $\tilde{\varepsilon} = [\tilde{\varepsilon}_1', ..., \tilde{\varepsilon}_T']'$. 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 and that z_{it} is a $(k-1)\times 1$ vector where $k=\dim(\gamma,\theta')$. 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 note that V is $n\times(Tk)$ as opposed to $\tilde{Z}$ which is $nT\times(k-1)$. We use V to estimate $\tilde{\Sigma}$ free of fixed effects bias as we discuss in more detail below. We impose the following condition on the fixed effects.

Condition 1. *Let $M_V=I_n-V(V'V)^{-1}V'$ be the matrix projecting onto the orthogonal complement of V. Conditional on $\tilde{Z}$ and $\tilde{\Upsilon}$, the fixed effects α_i satisfy one of the two conditions below:*
i) The α_i are independent across i and independent of ε_{jt} for all i,j,t. Moreover, $E(\alpha' M_V\alpha|V)/n$ is bounded for all n, almost surely.
ii) The α_i are fixed parameters such that $\alpha' M_V\alpha/n$ is bounded for all n almost surely.

Remark 1. *Condition 1i) generalizes the specification of Mundlak (1978) in the sense that α_i is random but can depend on V_i in a possibly nonlinear way. .*

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[2]. 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 poor performance of the usual bias corrected estimators and a stationarity assumption is inconsistent with our assumption of an unrestricted $\tilde{\Sigma}$.

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

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 bias[3]. The final step consists of projecting out the common constant. For this purpose define the projection matrix $M_{\mathbf{1}_T} = I_T - T^{-1}\mathbf{1}_T\mathbf{1}_T'$ projecting onto the orthogonal complement of $\mathbf{1}_T$. We estimate the (t,s)-th element $\sigma_{t,s}$ of $\tilde{\Sigma}$ as

$$\hat{\sigma}_{t,s} = \frac{\tilde{Y}_t' M_V \tilde{Y}_s}{\operatorname{tr}(M_V)}. \tag{2.2}$$

We now form the $T \times T$ matrix $\tilde{S}$ consisting of the elements $\hat{\sigma}_{t,s}$. We then form the estimate $\widehat{\tilde{\Sigma}}$ of $M_{\mathbf{1}_T}\tilde{\Sigma}M_{\mathbf{1}_T}$ as $\widehat{\tilde{\Sigma}} = M_{\mathbf{1}_T}\tilde{S}M_{\mathbf{1}_T}$.

Note that $\widehat{\tilde{\Sigma}}$ is of rank $T-1$ due to a loss of degrees of freedom resulting from removing the fixed effects. A full rank matrix can be obtained by deleting the first column and row from $\widehat{\tilde{\Sigma}}$ 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 (2.1) can be achieved by applying the transformation

$$\left(BM_{\mathbf{1}_T} \otimes I_n\right)\tilde{Y} = \left(BM_{\mathbf{1}_T} \otimes I_n\right)\tilde{Z}\theta + \left(BM_{\mathbf{1}_T} \otimes I_n\right)\tilde{\Upsilon}\gamma + \left(BM_{\mathbf{1}_T} \otimes I_n\right)\tilde{\varepsilon}.$$

We define $Y = \left(BM_{\mathbf{1}_T} \otimes I_n\right)\tilde{Y}$. Similarly we define $\Upsilon = \left(BM_{\mathbf{1}_T} \otimes I_n\right)\tilde{\Upsilon}$, $Z = \left(BM_{\mathbf{1}_T} \otimes I_n\right)\tilde{Z}$ and $\varepsilon = \left(BM_{\mathbf{1}_T} \otimes I_n\right)\tilde{\varepsilon}$. Note that $E(\varepsilon) = 0$ and $E(\varepsilon\varepsilon') = \Omega = \Sigma \otimes I_n$ where $\Sigma = BM_{\mathbf{1}_T}\tilde{\Sigma}M_{\mathbf{1}_T}B'$ such that GLS can be implemented by using the estimator[4]

$$\hat{\Sigma} = BM_{\mathbf{1}_T}\tilde{S}M_{\mathbf{1}_T}B'. \tag{2.3}$$

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

---

[3]The bias of the estimated autocovariances $\hat{\sigma}_{t,s}$ defined in (2.2) is constant by construction and withouth any additional assumptions beyond Conditions 1 and 2. The reason is that the bias only depends on α which is constant over time and the regressor matrix V. The matrix V is the same for each time period in which the residuals are computed and thus time-invariant itself. The fact that V contains all leads and lags of z_{it} and T_{it} is thus critical for our unbiasedness result.

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

Condition 2. *All asymptotic arguments are for T fixed and $n \to \infty$. Let $X = [Z, \Upsilon]$ be a $n(T-1) \times k$ matrix and similarly define $\tilde{X} = \left[\tilde{Z}, \tilde{\Upsilon}\right]$. Assume that conditional on $\tilde{X}$, $\tilde{\varepsilon}$ is jointly normal with $\tilde{\varepsilon} \sim N(0, \tilde{\Sigma} \otimes I_n)$. Assume that X has full column rank k almost surely (a.s.). Assume that Σ is of full rank. Let $\vartheta_0 = \operatorname{vech} \Sigma$. Assume that ϑ_0 is in the interior of some parameter space $\Phi \subset \mathbb{R}^{T(T-1)/2}$. Use the notation $\Omega(\vartheta)^{-1}$ to emphasize the dependence of Ω^{-1} on ϑ and for the j-th element of ϑ, denoted by ϑ_j, let $\dot{\Omega}_j^{-1} = \partial \Omega(\vartheta_0)^{-1} / \partial \vartheta_j$, $\dot{\Omega}_{jl}^{-1} = \partial^2 \Omega(\vartheta_0)^{-1} / (\partial \vartheta_j \partial \vartheta_l)$ and $\dot{\Omega}_{jlm}^{-1} = \partial^3 \Omega(\vartheta_0)^{-1} / (\partial \vartheta_j \partial \vartheta_l \partial \vartheta_m)$ be $n(T-1) \times n(T-1)$ dimensional matrices of partial derivatives. Assume that the limits as $n \to \infty$ of $X'\Omega^{-1}X/n$, $X'\dot{\Omega}_j^{-1}X/n$, $X'\dot{\Omega}_{jl}^{-1}X/n$ and $X'\dot{\Omega}_{jlm}^{-1}X/n$ exist and are bounded a.s. for all j, l, m. Let $Q = I_{n(T-1)} - X\left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}$. Assume that the limits of $X'\dot{\Omega}_j^{-1} Q\Omega Q \dot{\Omega}_j^{-1} X/n$, $X'\dot{\Omega}_{jl}^{-1} Q\Omega Q \dot{\Omega}_{jl}^{-1} X/n$ and $X'\dot{\Omega}_{jlm}^{-1} Q\Omega Q \dot{\Omega}_{jlm}^{-1} X/n$ as $n \to \infty$ exist and are bounded a.s. for all j, l, m. Assume that uniformly in a neighborhood of ϑ_0, $X'\left(\dot{\Omega}_{ijlm}^{-1}\right)^2 X/n$ is bounded a.s. as $n \to \infty$ where $\dot{\Omega}_{ijlm}^{-1} = \partial^4 \Omega(\vartheta_0)^{-1} / (\partial \vartheta_i \partial \vartheta_j \partial \vartheta_l \partial \vartheta_m)$ is an $n(T-1) \times n(T-1)$ dimensional matrix of partial derivatives.*

Remark 2. *We regard the analysis as being conditional on a particular draw of regressors X. If the regressors are fixed the qualifier 'almost surely' can be omitted in Condition 2.*

Remark 3. *Note that Σ is of full rank implies that $\Omega = \Sigma \otimes I_n$ is full rank for all n. Together with the rank condition on X this implies that $X'X/n > 0$, $X'\Omega X/n > 0$ and $X'\Omega^{-1}X/n > 0$ a.s for all n where the inequality denotes that the matrix on the l.h.s is positive definite.*

The following theorem establishes the unbiasedness of the covariance matrix estimator $\hat{\Sigma}$ for $B'M_{\mathbf{1}_T}\tilde{\Sigma}M_{\mathbf{1}_T}B'$ and establishes its asymptotic variance.

Theorem 2.1. *Assume that Conditions 1 and 2 hold and that y_{it} is generated by (2.1). Let $\hat{\Sigma}$ be defined as in (2.3). Then $E\left(\hat{\Sigma}|X\right) = \Sigma$ a.s. Let $\tilde{T} = T-1$ and let the $K_{\tilde{T}\tilde{T}}$ be the $(T-1)^2 \times (T-1)^2$ commutation matrix of Magnus and Neudecker (1979) defined as $K_{\tilde{T}\tilde{T}} = \sum_{i,j=1}^{(T-1)} a_i a_j' \otimes a_j a_i'$ where a_i is the i-th unit vector of dimension $T-1$. It then follows that*

$$n \operatorname{Var}\left(\operatorname{vec}\left(\hat{\Sigma}\right)|X\right) = V_\Omega + O\left(n^{-1}\right) \text{ a.s.}$$

where

$$V_\Omega = (I_{T-1} \otimes K_{n\tilde{T}} \otimes I_n) \left(V_\Sigma \otimes \operatorname{vec} I_n (\operatorname{vec} I_n)'\right) \left(I_{T-1} \otimes K'_{n\tilde{T}} \otimes I_n\right)$$

with $K_{n\tilde{T}} = \sum_{i,j=1} E_{ij} \otimes E'_{ij}$ *where* E_{ij} *is an* $n \times (T-1)$ *matrix with* i, j *th position equal to one and all the other elements are equal to zero,* $V_\Sigma = (I_{T-1} \otimes I_{T-1} + K_{\tilde{T}\tilde{T}}) (\Sigma \otimes \Sigma)$ *and* $\Sigma = BM_{\mathbf{1}_T} \tilde{\Sigma} M_{\mathbf{1}_T} B'$.

Remark 4. *Note that* V_Σ *is singular because of repeated elements in* Σ.

For the time being the autocorrelation structure of $\tilde{\Sigma}$ is assumed unrestricted. Thus, OLS, GLS and FGLS are all unbiased estimators. Estimation of the parameter γ can be done using OLS or GLS[5]. The model 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

$$\bar{T}_1 = \frac{(\Upsilon'\Omega_z\Upsilon)^{-1} \Upsilon'\Omega_z Y - \gamma_0}{(\Upsilon'\Omega_z\Upsilon)^{-1/2}} \tag{2.4}$$

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

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

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

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

[5] If the dimension of T is relatively large compared to the dimension of n, it has been argued in the literature (see for example Bertrand et. al. (2004) or Moulton (1986)) 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. The working paper version of this paper provides details on how to handle time averaging and more general grouping of time periods but to preserve space and notation we omit these results in what follows.

and

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

Note, that in the same way as Rothenberg (1988, p. 1001) we assume that b_1 and b_2 stay fixed as n tends to infinity. This is equivalent to assuming Pitman type local alternatives to analyze power of the tests. 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 robust OLS. For first order asymptotics, because of orthogonality between $\hat{\gamma}$ 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 $\hat{\Omega} = BM_{\mathbf{1}_T} \tilde{S} M_{\mathbf{1}_T} B' \otimes I_n$ with $\tilde{S}$ defined as in (2.3). We use the Edgeworth expansions of Rothenberg (1988) to obtain more precise statements about the finite sample behavior of the feasible test statistic T_1 where

$$T_1 = \frac{\left(\Upsilon' \hat{\Omega}_z \Upsilon\right)^{-1} \Upsilon' \hat{\Omega}_z Y - \gamma_0}{\left(\Upsilon' \hat{\Omega}_z \Upsilon\right)^{-1/2}} \tag{2.5}$$

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 as a function of the infeasible test $\bar{T}_1$ where $\bar{T}_1$ is defined in (2.4). As noted in Rothenberg, the statistic T_1 can be written as

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

where

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

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

$$U = \frac{(h' \otimes h')}{h' \Omega h} \sqrt{n} \operatorname{vec}\left(\Omega - \hat{\Omega}\right) - \sqrt{n} \frac{\operatorname{tr}\left[\operatorname{vec}(\Omega - \hat{\Omega}) \operatorname{vec}(\Omega - \hat{\Omega})' \left(H \otimes hh'\right)\right]}{h' \Omega h} \tag{2.6}$$

and

$$R = \frac{(\varepsilon' H \otimes h')}{(h' \Omega h)^{1/2}} \sqrt{n} \operatorname{vec}\left(\Omega - \hat{\Omega}\right). \tag{2.7}$$

We state the following result which will be needed for the distributional approximation of our test statistics.

Theorem 2.2. *Assume that Conditions 1 and 2 hold and that y_{it} is generated by (2.1). Then,* $\tilde{U} - U = O_p\left(n^{-1}\right)$ *and* $\tilde{R} - R = O_p\left(n^{-1}\right)$ *and* $E\left(R|X\right) = 0$ *a.s.,* $\text{cov}(R, U|X) = O(n^{-1})$ *a.s.,*

$$E\left(U|X\right) = -n^{-1/2}\frac{\text{tr}\, V_\Omega\left(H \otimes hh'\right)}{h'\Omega h} + O(n^{-1}) \text{ a.s.}$$

$$\text{var}\left(U|X\right) = \frac{\text{tr}\, V_\Omega\left(hh' \otimes hh'\right)}{\left(h'\Omega h\right)^2} + O(n^{-1}) \text{ a.s.}$$

and

$$\text{var}\left(R|X\right) = \frac{\text{tr}\, V_\Omega\left(H \otimes hh'\right)}{h'\Omega h} + O(n^{-1}) \text{ a.s.}$$

We also define the feasible Wald test based on the OLS estimator using robust standard errors,

$$T_2 = \frac{\left(\Upsilon' M_z \Upsilon\right)^{-1} \Upsilon' M_z Y - \gamma_0}{\left(\left(\Upsilon' M_z \Upsilon\right)^{-1} \Upsilon' M_z \hat{\Omega}_z M_z \Upsilon \left(\Upsilon' M_z \Upsilon\right)^{-1}\right)^{1/2}}, \tag{2.8}$$

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.

We now state the first main result of the paper which gives explicit formulas for the higher order size correction terms of the tests T_1 and T_2 by providing formal Edgeworth approximations to the distribution of T_1 and T_2.

Theorem 2.3. *Assume that Conditions 1 and 2 hold and that y_{it} is generated by (2.1). Then*

$$\Pr(T_1 \leq t|X) = \Phi\left[t\left(1 - \frac{A_1\left(t, \Omega\right)}{2n}\right) - b_1\left(1 - \frac{B_1\left(t, \Omega\right)}{2n}\right)\right] + o(n^{-1}) \text{ a.s.}$$

and

$$\Pr\left(T_2 \leq t|X\right) = \Phi\left[t\left(1 - \frac{A_2\left(t, \Omega\right)}{2n}\right) - b_2\left(1 - \frac{B_2\left(t, \Omega\right)}{2n}\right)\right] + o(n^{-1}) \text{ a.s.}$$

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

$$A_1\left(t\right) = \frac{1}{4}\left(1 + t^2\right)\frac{\text{tr}\, V_\Omega\left(hh' \otimes hh'\right)}{\left(h'\Omega h\right)^2} + 2\frac{\text{tr}\, V_\Omega\left(H \otimes hh'\right)}{h'\Omega h}, \; A_2\left(t, \Omega\right) = \frac{1}{4}\left(1 + t^2\right)\frac{\text{tr}\, V_\Omega\left(xx' \otimes xx'\right)}{\left(x'\Omega x\right)^2}$$

and

$$B_1(t) = \frac{1}{4}t^2 \frac{\operatorname{tr} V_\Omega (hh' \otimes hh')}{(h'\Omega h)^2} + \frac{\operatorname{tr} V_\Omega (H \otimes hh')}{h'\Omega h}, \quad B_2(t,\Omega) = \frac{1}{4}t^2 \frac{\operatorname{tr} V_\Omega (xx' \otimes xx')}{(x'\Omega x)^2}.$$

Note that 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 $A_1(t,\Omega)$ and $A_2(t,\Omega)$ if biased estimators were used.

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

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

where t_α is the critical value satisfying $\Phi(t_\alpha) = 1 - \alpha$. In principle similar size corrected tests could be achieved for robust OLS. However, our analysis of a special case of particular interest in the next section reveals, that robust OLS seems to be far less sensitive to the dimension of the unknown covariance matrix Σ. Monte Carlo evidence indicates that tests based on robust OLS have approximately correct size even without the correction but that their good size properties come at the cost of lower power compared with GLS based tests.

In general the constant $A_1(t_\alpha, \Omega)$ 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. An estimator for $A_1(t_\alpha, \Omega)$ is obtained from $\hat{V}_\Sigma = (I_{T-1} \otimes I_{T-1} + K_{\tilde{T}\tilde{T}})\left(\hat{\Sigma} \otimes \hat{\Sigma}\right)$, replacing V_Σ with $\hat{V}_\Sigma$ in V_Ω to obtain $\hat{V}_\Omega$ and by replacing Ω_z with $\hat{\Omega}_z$ in H and h to obtain $\hat{H}$ and $\hat{h}$. Since the analysis is conditional on X, the only parametric estimate needed is $\hat{\Omega}$ and we denote the estimate of $A_1(t_\alpha, \Omega)$ by $A_1\left(t_\alpha, \hat{\Omega}\right)$. The following Theorem states that $A_1\left(t_\alpha, \hat{\Omega}\right)$ can be used instead of $A_1(t_\alpha, \Omega)$ without affecting the size of the corrected test up to order $o(n^{-1})$.

Theorem 2.4. *Assume that Conditions 1 and 2 hold and that y_{it} is generated by (2.1). Then,*

$$\Pr\left(T_1 \le t_\alpha\left(1+(2n)^{-1}A_1(t_\alpha,\Omega)\right)|X\right) - \Pr\left(T_1 \le t_\alpha\left(1+(2n)^{-1}A_1\left(t_\alpha,\hat{\Omega}\right)\right)|X\right) = o\left(n^{-1}\right).$$

Computation of $A_1(t_\alpha,\Omega)$ requires us to evaluate $\operatorname{tr} V_\Omega(hh'\otimes hh')$ and $\operatorname{tr} V_\Omega(H\otimes hh')$. Because the dimensions of $V_\Omega(hh'\otimes hh')$ and $V_\Omega(H\otimes hh')$ can be very large it is more convenient for computational purposes to use the expression

$$\begin{aligned}
(2.9)\qquad & \operatorname{tr} V_\Omega\left(hh'\otimes hh'\right) \\
= & \textstyle\sum_{i,j=1}^{n}\left(h'\left(\Sigma\otimes e_ie_j'\right)h\right)^2 \\
& +\textstyle\sum_{l,m=1}^{(T-1)}\sum_{i,j=1}^{n}\left(h'\left(a_la_m'\Sigma\otimes e_ie_j'\right)h\right)\left(h'\left(a_ma_l'\Sigma\otimes e_ie_j'\right)h\right)
\end{aligned}$$

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

$$\begin{aligned}
(2.10)\qquad & \operatorname{tr} V_\Omega\left(H\otimes hh'\right) \\
= & \textstyle\sum_{i,j=1}^{n}\operatorname{tr}\left[\left(\Sigma\otimes e_ie_j'\right)H\right]\left(h'\left(\Sigma\otimes e_ie_j'\right)h\right) \\
& +\textstyle\sum_{l,m=1}^{(T-1)}\sum_{i,j=1}^{n}\operatorname{tr}\left(\left(a_la_m'\Sigma\otimes e_ie_j'\right)H\right)\left(h'\left(a_ma_l'\Sigma\otimes e_ie_j'\right)h\right).
\end{aligned}$$

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 low power when the model is misspecified in terms of the timing of policy effects[6]. Despite these potential problems we turn to models formulated in first

[6] Consider the model where $y_{it+1} = T_{it}\gamma + \varepsilon_{it}$ and T_{it} is binary. The effect of the policy T_{it} occurs with one period delay. If the estimated model is misspecified such that Δy_{it} is regressed on ΔT_{it} rather than on ΔT_{it-1} then $E\left(\Delta y_{it}\Delta T_{it}\right) = \Delta T_{it-1}\Delta T_{it} = 0$. Thus the OLS estimator of the misspecified first difference model is zero in large samples while the estimator for the misspecified model in levels will be asymptotically biased with a bias of $O\left(T^{-1}\right)$ as $n\to\infty$.

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 $\Delta y_{it} = y_{it} - y_{it-1}$ is the first difference of y_{it} and is used to remove fixed effects. The $k-1$ dimensional exogenous regressor Δz_{it} contains time dummies as well as other covariates.

We stack the observations as $\Delta Y = [\Delta y_{12}, ..., \Delta y_{n2}, ..., \Delta y_{1T}, ..., \Delta y_{nT}]'$, $\Upsilon_t^\Delta = [\Delta T_{1t}, ..., \Delta T_{nt}]'$, $\Upsilon^\Delta = \left[\Upsilon_2^{\Delta\prime}, ..., \Upsilon_T^{\Delta\prime}\right]'$, $Z_t^\Delta = [\Delta z_{1t}, ..., \Delta z_{nt}]'$ and $Z^\Delta = \left[Z_2^{\Delta\prime}, ..., Z_T^{\Delta\prime}\right]'$ with $\Delta\tilde{\varepsilon}$ being the corresponding vector of error terms. The model then can be written as

$$\Delta Y = Z^\Delta\theta + \Upsilon^\Delta\gamma + \Delta\varepsilon \tag{2.11}$$

with $E(\Delta\varepsilon) = 0$ and $E(\Delta\varepsilon\Delta\varepsilon') = \Sigma_{\Delta,T} \otimes I_n \equiv \Omega_\Delta$. If we define the $T-1 \times T$ matrix

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

then (2.11) can be obtained from the level specification by noting that $\Delta Y = \left(B^\Delta \otimes I_n\right)\tilde{Y}$, $Z^\Delta = \left(B^\Delta \otimes I_n\right)\tilde{Z}$, $\Upsilon^\Delta = \left(B^\Delta \otimes I_n\right)\tilde{\Upsilon}$ and $\Delta\varepsilon = \left(B^\Delta \otimes I_n\right)\tilde{\varepsilon}$. It thus follows that $\Omega_\Delta = B^\Delta\tilde{\Sigma}B^{\Delta\prime} \otimes I_n$. Moreover, because $B^\Delta\mathbf{1}_T = 0$ it also follows that $E\left(B^\Delta\tilde{S}B^\Delta\right) = B^\Delta\tilde{\Sigma}B^{\Delta\prime}$. 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.

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 $h_\Delta = \Omega_z^\Delta\Upsilon^\Delta\left(\Upsilon^{\Delta\prime}\Omega_z^\Delta\Upsilon^\Delta\right)^{-1}$ and $H_\Delta = \Omega_z^\Delta - \Omega_z^\Delta\Upsilon^\Delta\left(\Upsilon^{\Delta\prime}\Omega_z^\Delta\Upsilon^\Delta\right)^{-1}\Upsilon^{\Delta\prime}\Omega_z^\Delta$ where $\Omega_z^\Delta = \Omega_\Delta^{-1} - \Omega_\Delta^{-1}Z_\Delta(Z^{\Delta\prime}\Omega_\Delta^{-1}Z^\Delta)^{-1}Z^{\Delta\prime}\Omega_\Delta^{-1}$. Using these results we conclude from Rothenberg (1988) and our previous analysis for the level case that

$$\Pr(T_1 \le t) = \Phi\left[t\left(1 - \frac{A_1(t, \Omega_\Delta)}{2n}\right) - b_1\left(1 - \frac{B_1(t, \Omega_\Delta)}{2n}\right)\right] + o(n^{-1})$$

where

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

and

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

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 $M_{z\Delta}$ is the matrix projecting onto the orthogonal complement of Z^Δ. Then the Edgeworth approximation for T_2 is given by

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

where

$$A_2(t,\Omega_\Delta) = \frac{1}{4}\left(1+t^2\right)\frac{\operatorname{tr} V_{\Omega_\Delta}\left(x_\Delta x'_\Delta \otimes x_\Delta x'_\Delta\right)}{\left(x'_\Delta \Omega x_\Delta\right)^2} \tag{2.12}$$

and

$$B_2(t,\Omega_\Delta) = \frac{1}{4}t^2\frac{\operatorname{tr} V_{\Omega_\Delta}\left(x_\Delta x'_\Delta \otimes x_\Delta x'_\Delta\right)}{\left(x'_\Delta \Omega x_\Delta\right)^2}. \tag{2.13}$$

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 $T_{it}=1$ and that $T_{it}=0$ implies $T_{i\tau}=0$ for all $t>\tau$. The fixed effects α_i are assumed to satisfy Condition 1, ε_{it} satisfies Condition 2 and V_i is defined in the same way as before, but using the covariates of (3.1). We thus consider the transformed model

$$(BM_{\mathbf{1}_T}\otimes I_n)\,\tilde{Y}=(BM_{\mathbf{1}_T}\otimes \mathbf{1}_n)\,\tilde{\beta}+(BM_{\mathbf{1}_T}\otimes I_n)\,\tilde{\Upsilon}\gamma+(BM_{\mathbf{1}_T}\otimes I_n)\,\tilde{\varepsilon}$$

where $\tilde{\beta}=[\beta_1,...,\beta_T]'$. For the transformed model we let $Y=(BM_{\mathbf{1}_T}\otimes I_n)\,\tilde{Y}$, $Z=(I_{T-1}\otimes \mathbf{1}_n)$, $\Upsilon=(BM_{\mathbf{1}_T}\otimes I_n)\,\tilde{\Upsilon}$, $\varepsilon=(BM_{\mathbf{1}_T}\otimes I_n)\,\tilde{\varepsilon}$ and $\beta=\left[\beta_2-\bar{\beta},...,\beta_T-\bar{\beta}\right]'$ a $(T-1)\times 1$ vector where $\bar{\beta}=T^{-1}\sum_{t=1}^{T}\beta_t$. The transformed model then takes the form

$$Y=(I_{T-1}\otimes \mathbf{1}_n)\,\beta+\Upsilon\gamma+\varepsilon.$$

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 with the help of (2.9) and (2.10) and summarize the results in the following Theorem.

Theorem 3.1. *Assume Conditions 1 and 2 hold with y_{it} generated by (3.1). Let T_1 be defined as in (2.5) with $Z=(I_{T-1}\otimes \mathbf{1}_n)$ and let T_2 be defined as in (2.8) with $Z=(I_{T-1}\otimes \mathbf{1}_n)$. Then*

$$\Pr(T_1\le t|X)=\Phi\left[t\left(1-\frac{A_1(t)}{2n}\right)-b_1\left(1-\frac{B_1(t)}{2n}\right)\right]+o(n^{-1})\text{ a.s.}$$

and

$$\Pr(T_2\le t|X)=\Phi\left[t\left(1-\frac{A_2(t)}{2n}\right)-b_1\left(1-\frac{B_2(t)}{2n}\right)\right]+o(n^{-1})\text{ a.s.}$$

where

$$A_1(t)=\frac{1}{2}\left(1+t^2\right)+2\,(T-2)\,,\ A_2(t)=\frac{1}{2}\left(1+t^2\right) \tag{3.2}$$

and

$$B_1(t)=\frac{1}{2}t^2+T-2,\ B_2(t)=\frac{1}{2}t^2 \tag{3.3}$$

These results turn out to be the same as for the first difference version of the test. Further, it turns out that robust OLS is unaffected by the dimension T as far as the second order terms are concerned. Using the result for B_1 it now follows that the power of the test T_1 can be approximated by

$$\frac{\gamma-\gamma_0}{\left(\sigma^{\tau\tau}\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^{\Delta}\right)^{-1/2}}\left(1-\frac{1}{4n}t^2-\frac{T-2}{2n}\right) \tag{3.4}$$

where $\tilde{\Upsilon}_\tau^{\Delta}=[T_{1\tau}-T_{1\tau-1},...,T_{n\tau}-T_{n\tau-1}]'$ is the vector of changes in treatment status at time τ for each state and

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

where $\xi_\tau=\left[\mathbf{0}_{\tau-1}',\mathbf{1}_{T-\tau+1}'\right]'$ and $\mathbf{0}_{\tau-1}$ is a $(\tau-1)\times 1$ vector of zeros. Note that $\Sigma=BM_{\mathbf{1}_T}\tilde{\Sigma}M_{\mathbf{1}_T}B'$ and $\Upsilon'\Omega_z\Upsilon=\sigma^{\tau\tau}\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^{\Delta}$ as shown in (A.4) in the Appendix.

We can exploit the simple structure of (3.4) to analyze the question of optimal temporal aggregation. For this purpose we introduce the averaging matrix C of dimension $(T-1)\times r$ such that $C'C=I_r$. We define $Y_C=(C'BM_{\mathbf{1}_T}\otimes I_n)\tilde{Y}$. Similarly we define $\Upsilon_C=(C'BM_{\mathbf{1}_T}\otimes I_n)\tilde{\Upsilon}$, $Z_C=(C'BM_{\mathbf{1}_T}\otimes I_n)\tilde{Z}$ and $\varepsilon_C=(C'BM_{\mathbf{1}_T}\otimes I_n)\tilde{\varepsilon}$. Note that $E(\varepsilon_C)=0$ and $E(\varepsilon_C\varepsilon_C')=\Omega_C=\Sigma_C\otimes I_n$ where $\Sigma_C=C'BM_{\mathbf{1}_T}\tilde{\Sigma}M_{\mathbf{1}_T}B'C$.

It can be shown[7] that for general C of dimension $(T-1)\times r$, $A_1(t)=\frac{1}{2}(1+t^2)+2(r-1)$. This result shows that the size distortion can be reduced by choosing r small. This is likely to occur at the cost of lower power. In fact, as a function of C, (3.4) is

$$\frac{(\gamma-\gamma_0)}{\left(\sigma^{\tau\tau}(C)\,\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^{\Delta}\right)^{-1/2}}\left(1-\frac{1}{4n}t^2-\frac{r-1}{2n}\right)$$

with $\sigma^{\tau\tau}(C)=\xi_\tau' M_{\mathbf{1}_T}B'C\Sigma_C^{-1}C'BM_{\mathbf{1}_T}\xi_\tau'$. Maximizing (3.4) thus amounts to choosing C optimally. 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 $\sigma^{\tau\tau}(C)$. Estimation errors of the elements in the optimal weight matrix affect power to order n^{-1} through the term $(r-1)/2n$. An algorithm for maximizing (3.4) 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

$$C_r^*=\underset{C\text{ s.t. }C'C=I_r}{\arg\max}\;\sigma^{\tau\tau}(C)=\underset{C\text{ s.t. }C'C=I_r}{\arg\max}\;\xi_\tau' M_{\mathbf{1}_T}B'C\Sigma_C^{-1}C'BM_{\mathbf{1}_T}\xi_\tau'. \tag{3.5}$$

[7] Derivations are available on request from the authors.

Theorem 3.2. *Let C_r^* be as defined in (3.5). Then, for $\sigma^{\tau\tau}(C) = \xi_\tau' M_{\mathbf{1}_T} B' C \Sigma_C^{-1} C' B M_{\mathbf{1}_T} \xi_\tau'$ it follows that*

$$\max_{r=1,..T-1} \sigma^{\tau\tau}(C_r^*) = \sigma^{\tau\tau}(C_1^*)$$

with

$$C_1^* = (BM_{\mathbf{1}_T}\tilde{\Sigma}M_{\mathbf{1}_T}B')^{-1}BM_{\mathbf{1}_T}\xi_\tau.$$

and approximate power defined (3.4) is maximized for $C = C_1^*$.

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 $C^* = -[\mathbf{0}_{\tau-2}', \mathbf{1}_{T-\tau+1}']'$ which implies that

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

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. To put it differently, addressing the size problem by using the difference in difference approach is likely to result in lower power, especially when $\tilde{\Sigma}$ differs from I_T.

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.

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 $\tilde{X}$ for this particular case takes on the form

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

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

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

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

The OLS estimator for this case is

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

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

$$T_{1,\Delta} = \frac{\left(\hat{\sigma}_\Delta^{\tau\tau}\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta\right)^{-1}\left(\sum_t \hat{\sigma}_\Delta^{\tau t}\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}\tilde{Y}_t^\Delta\right) - \gamma_0}{\left(\hat{\sigma}_\Delta^{\tau\tau}\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta\right)^{-1/2}} \tag{3.6}$$

and

$$T_{2,\Delta} = \frac{\left(\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta\right)^{-1}\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}\tilde{Y}_\tau^\Delta - \gamma_0}{\hat{\sigma}_{\Delta,\tau\tau}^{1/2}\left(\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta\right)^{-1/2}}. \tag{3.7}$$

Theorem 3.3. *Assume Conditions 1 and 2 hold with y_{it} generated by (3.1). Let T_1 be defined as in (3.6). Then* $\Pr(T_1 \le t|X) = \Phi\left[t\left(1 - \frac{A_1(t)}{2n}\right) - b_1\left(1 - \frac{B_1(t)}{2n}\right)\right] + o(n^{-1})$ *a.s. and* $\Pr(T_2 \le t|X) = \Phi\left[t\left(1 - \frac{A_2(t)}{2n}\right) - b_1\left(1 - \frac{B_2(t)}{2n}\right)\right] + o(n^{-1})$ *a.s. where $A_1(t)$, $A_2(t)$, $B_1(t)$ and $B_2(t)$ are the same as in (3.2) and (3.3).*

As before it can be shown that the previous result generalizes to the case where the temporal averaging matrix C is used. As argued for the level case, conventional pre and post treatment averaging does improve the size properties of T_1 but usually comes at the cost of reduced power.

For the OLS based test T_2 we note as before that size is not affected by the number of time periods which suggest that temporal aggregation is not needed for the robust OLS based tests as far as achieving correct size is concerned. The power function can be approximated by

$$\frac{\gamma - \gamma_0}{\sigma_{\Delta;\tau\tau}^{1/2}\left(\tilde{\Upsilon}_\tau^{\Delta\prime} M_{\mathbf{1}_n} \tilde{\Upsilon}_\tau^{\Delta}\right)^{-1/2}} \left(1 - \frac{1}{4n} t^2\right)$$

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

$$\begin{aligned} \varepsilon_{it} &= \rho \varepsilon_{it-1} + u_{it} \\ &\varepsilon_{i0} \sim N\left(0, \frac{1}{1-\rho^2}\right) \end{aligned} \tag{4.1}$$

where $u_{it} \sim N(0,1)$ is iid both across i and t. We generate $\varepsilon_{it}$for $t = 1, ...T + 500$ for each i and discard the first 500 realizations to enforce stationarity. We assume that $\alpha_i \sim N(0,1)$, $\beta_t \sim N(0,1)$ and generate

$$y_{it} = \alpha_i + \beta_t + T_{it}\gamma + \varepsilon_{it} \tag{4.2}$$

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 50,000 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 50,000 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[8], in levels does well in terms of size for T equal 5,10 and 15. 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 0.081 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% of 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% of 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 (ROLS)[9]. 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 that exceeds 0.25 when T = 10, 15, and 20 although the nominal size is only 0.05. Thus, as the previous literature found, OLS cannot be used without a correction to the estimated

[8] In the tables we use the acronyms GLS and GLS-SC to denote the feasible GLS based tests without and with size corrections to save space.

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

variance matrix of the estimates or severe size distortions may result[10].

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[11]. 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 when ρ is large. Thus, we do not recommend their use. The traditional DID approach loses too much power to solve the problem of a consistent estimate of 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 outcome 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[12]. 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 FGLS, 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

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

[11] For the pre-treatment average we drop the first time period of the sample. This ensures comparability with the full sample tests where the first time period is dropped as well.

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

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.

In Table 4 we explore the robustness of our results to changes in the cross-sectional sample size. Keeping $T = 10$ fixed and only considering the case of $\rho = .8$ we find that FGLS-SC performs well for cross-sectional sample sizes as small as $n = 25$ while the uncorrected FGLS procedure shows rapidly deteriorating size properties as n becomes smaller. For $n = 15$ FGLS-SC no longer performs well but the three period averaging version of FGLS-SC still works well. Based on these limited experiments we recommend temporal aggregation combined with FGLS-SC when the ratio of observations to estimated covariance parameters falls below 5.

In Table 5 we investigate how important the assumption of normally distributed innovations is for our procedure. We draw u_{it} from a t-distribution with 4 degrees of freedom and generate ε_{it} according to (4.1). Stationarity is enforced by discarding the first 500 draws of ε_{it} for each cross-sectional unit. The t-distribution with low degrees of freedom has higher kurtosis than the normal distribution and provides a robustness check to outliers. The results in Table 5 show that FGLS-SC continues to have approximately correct size for all sample sizes considered. We repeat the same exercise for u_{it} drawn from a demeaned chi-square distribution with 4 degrees of freedom. This distribution is skewed. The results again show that our procedure is not sensitive to departure from normality. The drop in power for these tests is a result of the higher variance of the t and chi-square distributions. We also note that the size of the uncorrected FGLS test only deteriorates marginally relative to the corresponding results for normal innovations in Table 1c. This is further evidence that the shape of the innovation distribution is not the main factor driving our results.

Recently Hansen (2007) proposed to fit a parametric model, in his case an AR(p) model, to the serial correlation process of ε_{it}. Based on Hahn and Kuersteiner's (2002) method of bias correction Hansen (2007) uses a formula of Nickell (1981) to correct for fixed effects bias in the estimated AR(p) model. 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 $\tilde{X} = \left[\mathbf{1}_T \otimes I_n, \tilde{Z}, \tilde{\Upsilon}\right]$. 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[13]. We then construct $\hat{\Sigma} = BM_{\mathbf{1}_T}\tilde{\Sigma}_{\hat{\rho}}M_{\mathbf{1}_T}B'$ or $\hat{\Sigma} = B^{\Delta}\tilde{\Sigma}_{\hat{\rho}}B^{\Delta\prime}$ depending on whether the model is estimated in first differences or in levels. Note that we estimate $\hat{\rho}$ on the full sample even for the cases where we consider averages over 2 or 3 subperiods because it was shown by Hahn and Kuersteiner (2002) that the performance of the bias correction improves with larger T.

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 calibrate the variance covariance matrix of our simulated innovations ε_{it} to the estimated covariance matrix of the residuals from that model. Denote the calibrated variance covariance matrix by $\hat{\Omega}$. Table 7 contains the entries of

[13] In Tables 1,2 and 4 we compute GLS-AR assuming that the variance of u_{it} is known. In Table 3 we compute GLS-AR based on the estimated variance of u_{it}, assuming that the variance is the same for all i and t.

the estimated correlation matrix corresponding to $\hat{\Omega}$. The sample size is $T = 11$. We then draw $\varepsilon_i \sim N(0, \hat{\Omega})$ 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 (4.2) where we again vary the size of the treatment effect measured by $\gamma = [0, .1, .6, 1]$.

Results for 50,000 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 9%. These size distortions are of the same order of magnitude as the size distortions of the uncorrected test based on feasible GLS. This contrasts with the unrestricted covariance matrix estimator used in FGLS with a size correction. This procedure continues to have approximately correct size as it did previously in stationary designs. The parametric GLS-AR method however has two additional disadvantages: because the weight matrix for the GLS estimator is biased when the data are not generated by an AR(1), the resulting test has less power. The results in Table 3 show that GLS-AR has lower power for $\gamma = [.6, 1]$ than FGLS-SC, despite the fact that it does not have correct size. The second disadvantage is that unlike for the uncorrected FGLS test, the size distortion of GLS-AR does not disappear for the 3 and 2 period based tests. The size distortion of GLS-AR worsens in these versions of the test as compared to the full sample. 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.

In Table 3 we also report the performance of the bootstrap procedure proposed by Bertrand et.al. (2004). Because the bootstrap is more time consuming to compute we run the simulations with 5000 instead of 50,000 replications. As suggested by Bertrand et.al. we use 200 bootstrap replications to compute critical values for each sample in our 5000 simulation draws. The bootstrap based tests, GLS-B, are somewhat undersized for the level specification and approximately correctly sized for the first difference specification in the full sample. The problem of undersized tests becomes more pronounced for the 3 and 2 period versions of the tests both for the level and first difference specifications. As a result, power also suffers for these cases when $\gamma = [.1, .6]$. Based on our results we can not fully recommend the bootstrap because it loses power due to a tendency to undersize the tests.

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.[14] In an interesting natural experiment 26 states regulated cellular prices, while the other 25 did not. In Table 6 we list monthly service prices in 1994 for the least expensive plan for average usage of 160 minutes per month (80% peak)[15] for up to a 1-year contract in the 10 largest MSAs, which are the metropolitan areas where cellular licenses were granted.[16]

Table 6 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 population, average commuting time, average MSA income, and an index of constructions costs.[17] 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.[18] The estimated coefficient of the price regulation

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

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

[16] 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.

[17] See Hausman (1995)

[18] 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.

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 ran 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".[19] 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.

A complicating factor arose because cellular prices decreased significantly in 1995-96 both because of new PCS entry and because of deregulation.[20] 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

[19] 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.

[20] 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.

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. Table 7 shows the serial correlations between regression residuals from OLS estimation of Equation (1.1). Apart from the magnitude of the serial correlation coefficients which reaches up to .6 in absolute value a striking feature of this correlation matrix is the non-stationarity of the residuals as evidenced in the changing magnitude of the first order serial correlation coefficient for different years in the sample.

We now use FGLS on the entire sample. We allow for an unrestricted covariance matrix and estimate it using an unbiased estimator. The FGLS 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 FGLS 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 0.004 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 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 conventional 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 FGLS 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. OLS with a robustly estimated covariance matrix, 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 is of accurate size and significantly more powerful than 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 data 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 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.

A. Appendix

A.1. Unbiased Weight Matrix Estimation

Proof of Theorem 2.1 . 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

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

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. First assume that Condition 1i) holds. Now consider

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

where $\operatorname{tr} E\left(\alpha' M_V \alpha|V\right)/\operatorname{tr}(M_V)$ is bounded by Condition (1). Since $E\left(\alpha' M_V \alpha|V\right)/\operatorname{tr}(M_V)$ does not depend on t or s it follows that

$$E\left(\tilde{S}|V\right) = \tilde{\Sigma} + \frac{E\left(\alpha' M_V \alpha|V\right)}{\operatorname{tr}(M_V)}\mathbf{1}_T\mathbf{1}'_T$$

and thus

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

For the case of Condition 1ii) note that $E\left(\hat{\sigma}_{t,s}|V\right) = \alpha' M_V \alpha/\operatorname{tr}(M_V) + \tilde{\sigma}_{t,s}$ and $E\left(\hat{\tilde{\Sigma}}|V\right) = M_{\mathbf{1}_T}\tilde{\Sigma}M_{\mathbf{1}_T}$ as before.

For the variance first assume that Condition 1i) holds. Consider

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

such that it is enough to look at

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

where second equality uses the fact that ε_t is Gaussian and where

$$\sigma_\alpha^2 \equiv \frac{\alpha' M_V \alpha}{\operatorname{tr}(M_V)}.$$

It follows that

$$
\begin{aligned}
& nE\left(\operatorname{vec}\left(\tilde{S}-\tilde{\Sigma}\right)\operatorname{vec}\left(\tilde{S}-\tilde{\Sigma}\right)'|V,\alpha\right) \\
= \; & \left(I_T\otimes I_T+K_{TT}\right)\left(\left(\sigma_\alpha^2\mathbf{1}_T\mathbf{1}_T'+\tilde{\Sigma}\right)\otimes\left(\sigma_\alpha^2\mathbf{1}_T\mathbf{1}_T'+\tilde{\Sigma}\right)-\left(\sigma_\alpha^2\mathbf{1}_T\mathbf{1}_T'\otimes\sigma_\alpha^2\mathbf{1}_T\mathbf{1}_T'\right)\right)+O(n^{-1})
\end{aligned}
$$

where K_{TT} is the $T^2\times T^2$ commutation matrix of Magnus and Neudecker (1979) and

$$
\begin{aligned}
\text{(A.1)}\qquad & nE\left(\operatorname{vec}\left(\hat{\Sigma}-B'M_{\mathbf{1}_T}\tilde{\Sigma}M_{\mathbf{1}_T}\right)\operatorname{vec}\left(\hat{\Sigma}-BM_{\mathbf{1}_T}\tilde{\Sigma}M_{\mathbf{1}_T}B'\right)'|V,\alpha\right) \\
= \; & \left(BM_{\mathbf{1}_T}\otimes BM_{\mathbf{1}_T}\right)\left(I_T\otimes I_T+K_{TT}\right)\left(\left(\sigma_\alpha^2\mathbf{1}_T\mathbf{1}_T'+\tilde{\Sigma}\right)\otimes\left(\sigma_\alpha^2\mathbf{1}_T\mathbf{1}_T'+\tilde{\Sigma}\right)\right)\left(M_{\mathbf{1}_T}B'\otimes M_{\mathbf{1}_T}B'\right) \\
& -\left(BM_{\mathbf{1}_T}\otimes BM_{\mathbf{1}_T}\right)\left(I_T\otimes I_T+K_{TT}\right)\left(\sigma_\alpha^2\mathbf{1}_T\mathbf{1}_T'\otimes\sigma_\alpha^2\mathbf{1}_T\mathbf{1}_T'\right)\left(M_{\mathbf{1}_T}B'\otimes M_{\mathbf{1}_T}B'\right)+O(n^{-1}) \\
= \; & \left(BM_{\mathbf{1}_T}\otimes BM_{\mathbf{1}_T}\right)\left(I_T\otimes I_T+K_{TT}\right)\left(\tilde{\Sigma}M_{\mathbf{1}_T}B'\otimes\tilde{\Sigma}M_{\mathbf{1}_T}B'\right)+O(n^{-1}) \\
= \; & \left(I_{T-1}\otimes I_{T-1}+K_{\tilde{T}\tilde{T}}\right)\left(BM_{\mathbf{1}_T}\tilde{\Sigma}M_{\mathbf{1}_T}B'\otimes BM_{\mathbf{1}_T}\tilde{\Sigma}M_{\mathbf{1}_T}B'\right)+O(n^{-1})
\end{aligned}
$$

where the last line follows from Magnus and Neudecker (1988, p.47) and $\tilde{T}=T-1$. Note that similar results hold when $M_{\mathbf{1}_T}$ is replaced by B^{Δ}. Since the conditional expectation in (A.1) does not depend on α it follows that

$$
nE\left(\operatorname{vec}\left(\hat{\Sigma}-\Sigma\right)\operatorname{vec}\left(\hat{\Sigma}-\Sigma\right)'|V\right)=\left(I_{T-1}\otimes I_{T-1}+K_{\tilde{T}\tilde{T}}\right)\left(\Sigma\otimes\Sigma\right)+O(n^{-1})
$$

and define $V_\Sigma=\left(I_{T-1}\otimes I_{T-1}+K_{\tilde{T}\tilde{T}}\right)\left(\Sigma\otimes\Sigma\right)$. Using Magnus and Neudecker (1988, Theorem 10, p.47) we write

$$
\operatorname{vec}\Omega=\left(I_{T-1}\otimes K_{n\tilde{T}}\otimes I_n\right)\left(\operatorname{vec}\Sigma\otimes\operatorname{vec}I_n\right)
$$

such that $\operatorname{Var}\left(n^{1/2}\left(\operatorname{vec}\left(\hat{\Omega}-\Omega\right)\right)|X\right)=V_\Omega+O\left(n^{-1}\right)$ where

$$
V_\Omega=\left(I_{T-1}\otimes K_{n\tilde{T}}\otimes I_n\right)\left(V_\Sigma\otimes\operatorname{vec}I_n\left(\operatorname{vec}I_n\right)'\right)\left(I_{T-1}\otimes K_{n\tilde{T}}'\otimes I_n\right).
$$

Since (A.1) holds conditional on α the case of Condition 1ii) follows by the same argument. ■

A.2. Proofs for Theorems in Section 2.1

Proof of Theorem 2.2. We first note that

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

Consider the total derivative

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

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

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

Now,

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

such that

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

by the delta method. In the same way,

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

and

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

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

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

or, using alternative notation,

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

In the Taylor expansion the term $d^2\Omega = 0$ because all the elements are linear functions of the parameters. We thus have $\tilde{U} - U = O_p(n^{-1})$ by the delta method where

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

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

$$E\,(U) = -n^{-1/2} \frac{\mathrm{tr}\,[V_\Omega\,(H \otimes hh')]}{(h'\Omega h)} + O(n^{-1}),$$

$$\mathrm{Var}\,(U) = \frac{\mathrm{tr}\,[V_\Omega\,(hh' \otimes hh')]}{(h'\Omega h)^2} + O(n^{-1}).$$

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

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

Since $\Omega_z(I - Z(Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}) = \Omega_z$ it follows that

$$H\Omega H = H(I - Z(Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}) \left(I - \Upsilon\,(\Upsilon'\Omega_z\Upsilon)^{-1}\,\Upsilon'\Omega_z\right) = H$$

which implies

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

and $\mathrm{Var}\,(R) = -n^{1/2}E\,(U) + O(n^{-1/2})$. ■

Proof of Theorem 2.3. From Theorem 2.1 it follows that $\sqrt{n}\, \mathrm{vec}(\Omega - \hat{\Omega}) = O_p\,(1)$ and from Theorem 2.2 it follows that $\tilde{R} = O_p\,(1)$ and $\tilde{U} = O_p(1)$. Using a result from Rothenberg (1988, p.1017) this implies that

$$T_1 = \bar{T}_1 + n^{-1/2}\left(\tilde{R} - \frac{1}{2}\bar{T}_1\tilde{U}\right) + n^{-1}\left(\frac{3}{8}\bar{T}_1\tilde{U} - \frac{1}{2}\tilde{U}\tilde{R}\right) + O_p\left(n^{-3/2}\right)$$

Next we verify 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 $\delta = \left[\theta', \gamma\right]'$. 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 $\hat{\varepsilon} = M\tilde{Y} = M\left(\tilde{\varepsilon} + (\mathbf{1}_T \otimes I_n)\,\alpha\right)$ with $\alpha = (\alpha_1, ..., \alpha_n)'$. Let $\hat{\varepsilon}_t = (\hat{\varepsilon}_{1t}, ..., \hat{\varepsilon}_{nt})'$ be the cross-section of estimated residuals for period t and note that $\hat{\varepsilon}_t = M_V\left(\tilde{\varepsilon}_t + \alpha\right)$. Then stack $\hat{E} = [\hat{\varepsilon}_1, ..., \hat{\varepsilon}_T]'$ such that $\hat{\Sigma} = M_{\mathbf{1}_T} \frac{\hat{E}\hat{E}'}{\mathrm{tr}(M_V)} M_{\mathbf{1}_T}$.

Consider a typical element

$$\frac{y_t' M_V y_s}{\mathrm{tr}(M_V)} = \frac{\sqrt{n}}{n - kT}\left(\tilde{\varepsilon}_t' M_V \tilde{\varepsilon}_s + \alpha' M_V \tilde{\varepsilon}_s + \tilde{\varepsilon}_t' M_V \alpha + \alpha' M_V \alpha\right)$$

such that for $\tilde{E}=(\tilde{\varepsilon}_1,...,\tilde{\varepsilon}_T)'$

$$\begin{aligned} M_{\mathbf{1}_T}\frac{\hat{E}\hat{E}'}{\operatorname{tr}(M_V)}M_{\mathbf{1}_T} &= \frac{1}{n-kT}M_{\mathbf{1}_T}\left(\tilde{E}M_V\tilde{E}'+\mathbf{1}_T\alpha' M_V\tilde{E}'+\tilde{E}M_V\alpha\mathbf{1}_T'+\alpha' M_V\alpha\mathbf{1}_T\mathbf{1}_T'\right)M_{\mathbf{1}_T} \\ &= \frac{1}{n-kT}M_{\mathbf{1}_T}\tilde{E}M_V\tilde{E}'M_{\mathbf{1}_T}. \end{aligned}$$

It follows that $\hat{\Omega}$ is an even function of $\tilde{\varepsilon}$ and does not depend on δ. Thus Assumption A of Rothenberg (1984a) is satisfied. Assumption B of Rothenberg (1984a) follows by verifying Assumptions 1-5 of Rothenberg (1984a, p.817) except that here we only require approximations of order $o(n^{-1})$. Assumptions 1-3 of Rothenberg (1984a) follow directly from Condition (2). Next, consider

$$\sqrt{n}\left(\frac{1}{n-kT}M_{\mathbf{1}_T}\tilde{E}M_V\tilde{E}'M_{\mathbf{1}_T}-\Sigma\right)=M_{\mathbf{1}_T}\sqrt{n}\left(\frac{1}{n-kT}\tilde{E}M_V\tilde{E}'-\tilde{\Sigma}\right)M_{\mathbf{1}_T}.$$

Since T is fixed as $n\to\infty$ it is enough to consider a typical element

$$\sqrt{n}\left(\frac{1}{n-kT}\tilde{E}M_V\tilde{E}'-\tilde{\Sigma}\right)_{t,s}=\sqrt{n}\left(\frac{1}{n-kT}\tilde{\varepsilon}_t'M_V\tilde{\varepsilon}_s-\tilde{\sigma}_{t,s}\right).$$

Let $kT=d$ such that for $n>d$, $\sqrt{n}\,(n-d)^{-1}=n^{-1/2}\left(1+d/n+(d/n)^2+...\right)=n^{-1/2}+d/n^{3/2}\,(1-d/n)^{-1}$. Then

$$\sqrt{n}\left(\frac{1}{n-kT}\tilde{\varepsilon}_t'M_V\tilde{\varepsilon}_s-\tilde{\sigma}_{t,s}\right)=\zeta_{n,ts}+\frac{\varsigma_{n,ts}}{n}$$

where

$$\varsigma_{n,ts}=\frac{d}{\sqrt{n}\,(1-d/n)}\sum_{i=1}^{n}(\varepsilon_{it}\varepsilon_{is}-\tilde{\sigma}_{t,s})$$

and

$$\zeta_{n,ts}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\varepsilon_{it}\varepsilon_{is}-\tilde{\sigma}_{t,s})-\frac{d\sqrt{n}}{n-d}\tilde{\sigma}_{t,s}+\frac{1}{n}\frac{1}{(1-d/n)}\sum_{i=1}^{n}\varepsilon_{is}V_i\left(n^{-1}V'V\right)^{-1}n^{-1/2}\sum_{i=1}^{n}V_i\varepsilon_{it}$$

where the stacked vector containing $\zeta_{n,ts}$ for all combinations of t and s is asymptotically

$$N\left(0,\left(I_T\otimes I_T+K_{TT}\right)(\Sigma\otimes\Sigma)\right),$$

the mean is $O\left(n^{-1/2}\right)$ and the covariance matrix is $(I_T\otimes I_T+K_{TT})(\Sigma\otimes\Sigma)+O(n^{-1})$. Furthermore, $\zeta_{n,ts}$ has bounded moments of all orders. For $\varsigma_{n,ts}$ we note that $(\varepsilon_{it}\varepsilon_{is}-\tilde{\sigma}_{t,s})$ are iid zero mean random variables with variance $\tilde{\sigma}_{t,t}\tilde{\sigma}_{s,s}+\tilde{\sigma}_{t,s}^2$. For any random variable V and constants $a,\lambda>0$ such that $E\left[e^{\lambda V}\right]$ exists it follows that $P(V\geq a)\leq e^{-\lambda a}E\left[e^{\lambda V}\right]$ and $P(V\leq -a)\leq e^{-\lambda a}E\left[e^{-\lambda V}\right]$ (see Chow and Teicher, 1997, p.109). For $\lambda=2$, $V=\varsigma_{n,ts}$ and $a=\log(n)$ one obtains, provided that n is large enough so that $E\left[e^{2\varsigma_{n,ts}}\right]$ and $E\left[e^{-2\varsigma_{n,ts}}\right]$ are bounded[21],

$$\Pr\left[\varsigma_{n,ts}\geq\log n\right]\leq e^{-2\log n}E\left[e^{2\varsigma_{n,ts}}\right]$$

[21] The moment generating function of $\varsigma_{n,ts}$ is given by $E\left[e^{\lambda\varsigma_{n,ts}}\right]=\left(\left(1-\tilde{\sigma}_{t,s}d_n n^{-1/2}\lambda\right)^2-\tilde{\sigma}_{t,t}\tilde{\sigma}_{s,s}d_n^2n^{-1}\lambda^2\right)^{-n/2}e^{-\lambda\sqrt{n}d_n\tilde{\sigma}_{t,s}}$ where $d_n=2d\,(1-d/n)^{-1}$.

and

$$\Pr\left[\varsigma_{n,ts} \leq -\log n\right] \leq e^{-2\log n} E\left[e^{-2\varsigma_{n,ts}}\right].$$

By the central limit theorem, $E\left[e^{2\varsigma_{n,ts}}\right]$ and $E\left[e^{2\varsigma_{n,ts}}\right]$ both converge to $e^{2\left(\tilde{\sigma}_{t,t}\tilde{\sigma}_{s,s}+\tilde{\sigma}_{t,s}^2\right)}$. It follows that

$$\Pr\left[|\varsigma_{n,ts}| > \log n\right] = n^{-2} e^{2\left(\tilde{\sigma}_{t,t}\tilde{\sigma}_{s,s}+\tilde{\sigma}_{t,s}^2\right)} + o\left(n^{-2}\right).$$

This establishes Assumption 4 of Rothenberg (1984a). For T_1 the result now follows from Theorem (2.2) and Rothenberg (1988).

For T_2 define $M = I_T \otimes M_V$, $x = \sqrt{n} M_z \Upsilon \left(\Upsilon' M_z \Upsilon\right)^{-1}$ where $'M_z$ is projecting onto the orthogonal complement of Z and $v = M\Omega x/\sqrt{x'\Omega x}$ where $\Omega = (\Sigma \otimes I)$. Then let $\bar{T}_2 - b_2 \equiv Y$ and note that $Y = x' M_z \varepsilon/\sqrt{x'\Omega x}$. Since Z and Υ are orthogonal to M and ε is Gaussian it follows that Y and $\hat{E}$ are independent. This in turn implies that $\hat{\Sigma} = BM_{\mathbf{1}_T}\hat{E}\hat{E}'M_{\mathbf{1}_T}B'/\operatorname{tr}(M_V)$ and Y are independent.

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

$$\sqrt{n}E\left[W|\bar{T}_2\right] = 0 \tag{A.2}$$

and

$$\operatorname{var}\left(W|\bar{T}_2\right) = \frac{\operatorname{tr} V_\Omega\left(xx' \otimes xx'\right)}{\left(x'\Omega x\right)^2} + O\left(n^{-1}\right) \tag{A.3}$$

by Theorem 2.1.

The formal Edgeworth expansion can be established in the same way as for T_1 and by noting, that as in Rothenberg (1988, p.1017), $T_2 = \bar{T}_2 - \left(2\sqrt{n}\right)^{-1}\bar{T}_2 W + 3\left(8n\right)^{-1}\bar{T}_2 W^2 + O_p\left(n^{-3/2}\right)$. Then use the expansion $W = \sqrt{n}x'\left(\hat{\Omega} - \Omega\right)x/\left(x'\Omega x\right)$ from before, together with the properties established earlier for $\hat{\Omega} - \Omega$ to conclude that the relevant terms in the expansion depend on the first two approximate moments of W derived in A.2 and A.3. ■

Proof of Theorem 2.4. Note that the result follows if $T_1 - t_\alpha\left(2n\right)^{-1}A_1\left(t_\alpha, \Omega\right)$ and $T_1 - t_\alpha\left(2n\right)^{-1}A_1\left(t_\alpha, \hat{\Omega}\right)$ have the same formal Edgeworth expansion up to order $o(n^{-1})$. The latter result follows if one can show that $A_1\left(t_\alpha, \Omega\right) - A_1\left(t_\alpha, \hat{\Omega}\right) = n^{-1/2}\Delta_A + R_A/n$ with $\Pr\left(|R_A| > \log n\right) = o\left(n^{-1}\right)$ and Δ_A is asymptotically normal with a mean which is $O\left(n^{-1/2}\right)$ and a variance which is $O\left(1\right)$. Using a two term Taylor expansion in ϑ around ϑ_0 and following the same arguments as in the proof of Theorem 2.3 is sufficient to establish these properties. ■

A.3. Derivation of (2.9) and (2.10)

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

$$\begin{aligned}
& \left(I_{(T-1)} \otimes K_{n(T-1)} \otimes I_n\right) \left(\Sigma \otimes \Sigma \otimes \operatorname{vec} I_n (\operatorname{vec} I_n)'\right) \left(I_T \otimes K_{n(T-1)}' \otimes I_n\right) \\
= & \sum_{i,j=1}^n \left(\Sigma \otimes K_{n(T-1)} \left(\Sigma \otimes e_i e_j'\right) K_{n(T-1)}' \otimes e_i e_j'\right) \\
= & \sum_{i,j=1}^n \left(\Sigma \otimes e_i e_j' \otimes \Sigma \otimes e_i e_j'\right).
\end{aligned}$$

We then find that

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

Next consider

$$\begin{aligned}
& \sum_{i,j=1}^n \left(I_{(T-1)} \otimes K_{n(T-1)} \otimes I_n\right) \left(K_{\tilde{T}\tilde{T}} (\Sigma \otimes \Sigma) \otimes e_i e_j' \otimes e_i e_j'\right) \left(I_T \otimes K_{n(T-1)}' \otimes I_n\right) \\
= & \sum_{l,m=1}^{(T-1)} \sum_{i,j=1}^n \left(I_{(T-1)} \otimes K_{n(T-1)} \otimes I_n\right) \left(a_l a_m' \Sigma \otimes a_m a_l' \Sigma \otimes e_i e_j' \otimes e_i e_j'\right) \left(I_T \otimes K_{n(T-1)}' \otimes I_n\right) \\
= & \sum_{l,m=1}^{(T-1)} \sum_{i,j=1}^n \left(a_l a_m' \Sigma \otimes e_i e_j' \otimes a_m a_l' \Sigma \otimes e_i e_j'\right)
\end{aligned}$$

such that

$$\begin{aligned}
& \operatorname{tr} \sum_{l,m=1}^{(T-1)} \sum_{i,j=1}^n \left(a_l a_m' \Sigma \otimes e_i e_j' \otimes a_m a_l' \Sigma \otimes e_i e_j'\right) (hh' \otimes hh') \\
= & \sum_{l,m=1}^{(T-1)} \sum_{i,j=1}^n \left(h' \left(a_l a_m' \Sigma \otimes e_i e_j'\right) h\right) \left(h' \left(a_m a_l' \Sigma \otimes e_i e_j'\right) h\right)
\end{aligned}$$

and

$$\begin{aligned}
\operatorname{tr}\left(V_\Omega (hh' \otimes hh')\right) = & \sum_{i,j=1}^n \left(h' \left(\Sigma \otimes e_i e_j'\right) h\right)^2 \\
& + \sum_{l,m=1}^{(T-1)} \sum_{i,j=1}^n \left(h' \left(a_l a_m' \Sigma \otimes e_i e_j'\right) h\right) \left(h' \left(a_m a_l' \Sigma \otimes e_i e_j'\right) h\right)
\end{aligned}$$

It now also follows immediately that

$$\begin{aligned}
\operatorname{tr}\left(V_\Omega (H \otimes hh')\right) = & \sum_{i,j=1}^n \operatorname{tr}\left(\left(\Sigma \otimes e_i e_j'\right) H\right) \left(\left(h' \left(\Sigma \otimes e_i e_j'\right) h\right)\right) \\
& + \sum_{l,m=1}^{(T-1)} \sum_{i,j=1}^n \operatorname{tr}\left(\left(a_l a_m' \Sigma \otimes e_i e_j'\right) H\right) \left(h' \left(a_m a_l' \Sigma \otimes e_i e_j'\right) h\right)
\end{aligned}$$

A.4. Derivations for Results in Section 3

Proof of Theorem 3.1. The results follow from specializing the expressions for $\operatorname{var}(U)$ 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 $\Upsilon = BM_{\mathbf{1}_T}\xi_\tau \otimes \tilde{\Upsilon}_\tau^\Delta$. Then

$$n^{-1}\Upsilon'\Omega_z\Upsilon = n^{-1}\tilde{\Upsilon}_\tau^{\Delta\prime} M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta \textstyle\sum_{t,s=\tau}^{T}\left(a_t' M_{\mathbf{1}_T} B'\Sigma^{-1} B M_{\mathbf{1}_T} a_s\right) \tag{A.4}$$

which implies that for $\sigma^{\tau\tau} = \xi_\tau' M_{\mathbf{1}_T} B'\Sigma^{-1} B M_{\mathbf{1}_T}\xi_\tau$ we can write

$$h = \frac{\Sigma^{-1} B M_{\mathbf{1}_T}\xi_\tau \otimes M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta}{\sigma^{\tau\tau}\tilde{\Upsilon}_\tau^{\Delta\prime} M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta}$$

and

$$H = \Sigma^{-1} \otimes M_{\mathbf{1}_n} - \frac{\Sigma^{-1} B M_{\mathbf{1}_T}\xi_\tau\xi_\tau' M_{\mathbf{1}_T} B'\Sigma^{-1} \otimes M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta\tilde{\Upsilon}_\tau^{\Delta\prime} M_{\mathbf{1}_n}}{\sigma^{\tau\tau}\tilde{\Upsilon}_\tau^{\Delta\prime} M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta}.$$

Since the denominator of h cancels out in $A_1(t,\Omega)$ and $B_1(t,\Omega)$ we consider $h = \Sigma^{-1} B M_{\mathbf{1}_T}\xi_\tau \otimes M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta$ without loss of generality. Then,

$$\begin{aligned} h'(\Sigma \otimes e_i e_j) h &= \xi_\tau' M_{\mathbf{1}_T} B'\Sigma^{-1} B M_{\mathbf{1}_T}\xi_\tau \tilde{\Upsilon}_\tau^{\Delta\prime} M_{\mathbf{1}_n} e_i e_j' M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta \\ &= \sigma^{\tau\tau}\tilde{\Upsilon}_\tau^{\Delta\prime} M_{\mathbf{1}_n} e_i e_j' M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta \end{aligned}$$

and

$$\textstyle\sum_{i,j=1}^n \left(h'(\Sigma \otimes e_i e_j) h\right)^2 = (\sigma^{\tau\tau})^2 \sum_{i,j=1}^n \tilde{\Upsilon}_\tau^{\Delta\prime} M_{\mathbf{1}_n} e_i e_i' M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta\tilde{\Upsilon}_\tau^{\Delta\prime} M_{\mathbf{1}_n} e_j e_j M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta = (\sigma^{\tau\tau})^2\left(\tilde{\Upsilon}_\tau^{\Delta\prime} M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta\right)^2.$$

Also,

$$h'\Omega h = \sigma^{\tau\tau}\tilde{\Upsilon}_\tau^{\Delta\prime} M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta.$$

Moreover note that

$$\begin{aligned} \operatorname{tr}\left[\left(\Sigma \otimes e_i e_j'\right) H\right] &= \operatorname{tr}\left(I_{T-1} \otimes e_i e_j' M_{\mathbf{1}_n}\right) \\ &\quad - \frac{\operatorname{tr}\left[\left(\Sigma \otimes e_i e_j'\right)\left(\Sigma^{-1} B M_{\mathbf{1}_T}\xi_\tau\xi_\tau' M_{\mathbf{1}_T} B'\Sigma^{-1} \otimes M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta\tilde{\Upsilon}_\tau^{\Delta\prime} M_{\mathbf{1}_n}\right)\right]}{\sigma^{\tau\tau}\tilde{\Upsilon}_\tau^{\Delta\prime} M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta} \\ &= (T-1)\, e_j' M_{\mathbf{1}_n} e_i - \frac{e_j' M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta\tilde{\Upsilon}_\tau^{\Delta\prime} M_{\mathbf{1}_n} e_i}{\tilde{\Upsilon}_\tau^{\Delta\prime} M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta} \end{aligned}$$

and

$$\begin{aligned} \textstyle\sum_{i,j=1}^n (T-1)\, e_j' M_{\mathbf{1}_n} e_i h(\Sigma \otimes e_i e_j) h &= \sigma^{\tau\tau}\textstyle\sum_{i,j=1}^n (T-1)\,\tilde{\Upsilon}_\tau^{\Delta\prime} M_{\mathbf{1}_n} e_i e_i' M_{\mathbf{1}_n} e_j e_j' M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta \\ &= (T-1)\,\sigma^{\tau\tau}\tilde{\Upsilon}_\tau^{\Delta\prime} M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta, \end{aligned}$$

as well as

$$\begin{aligned} \textstyle\sum_{i,j=1}^n \frac{e_j' M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta\tilde{\Upsilon}_\tau^{\Delta\prime} M_{\mathbf{1}_n} e_i}{\tilde{\Upsilon}_\tau^{\Delta\prime} M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta} h'(\Sigma \otimes e_i e_j) h &= \sigma^{\tau\tau}\textstyle\sum_{i,j=1}^n \frac{e_j' M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta\tilde{\Upsilon}_\tau^{\Delta\prime} M_{\mathbf{1}_n} e_i \tilde{\Upsilon}_\tau^{\Delta\prime} M_{\mathbf{1}_n} e_i e_j' M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta}{\left(\tilde{\Upsilon}_\tau^{\Delta\prime} M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta\right)} \\ &= \sigma^{\tau\tau}\left(\tilde{\Upsilon}_\tau^{\Delta\prime} M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta\right). \end{aligned}$$

Next consider

$$\left(h'\left(a_l a_m'\Sigma\otimes e_i e_j'\right)h\right)=\xi_\tau' M_{\mathbf{1}_T}B'\Sigma^{-1}a_l a_m' B M_{\mathbf{1}_T}\xi_\tau\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}e_i e_j' M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^{\Delta},$$

$$\begin{aligned}
&\textstyle\sum_{l,m=1}^{T-1}\sum_{i,j=1}^{n}\left(h'\left(a_l a_m'\Sigma\otimes e_i e_j'\right)h\right)\left(h'\left(a_m a_l'\Sigma\otimes e_i e_j'\right)h\right)\\
=&\textstyle\sum_{l,m=1}^{T-1}\xi_\tau' M_{\mathbf{1}_T}B'\Sigma^{-1}a_l a_m' B M_{\mathbf{1}_T}\xi_\tau\xi_\tau' M_{\mathbf{1}_T}B'\Sigma^{-1}a_m a_l' B M_{\mathbf{1}_T}\xi_\tau\left(\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^{\Delta}\right)^2\\
=&\textstyle\sum_{l,m=1}^{T-1}\xi_\tau' M_{\mathbf{1}_T}B'\Sigma^{-1}a_l a_l' B M_{\mathbf{1}_T}\xi_\tau\xi_\tau' M_{\mathbf{1}_T}B'\Sigma^{-1}a_m a_m' B M_{\mathbf{1}_T}\xi_\tau\left(\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^{\Delta}\right)^2\\
=&\left(\xi_\tau' M_{\mathbf{1}_T}B'\Sigma^{-1}B M_{\mathbf{1}_T}\xi_\tau\right)^2\left(\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^{\Delta}\right)^2=\left(\sigma^{\tau\tau}\right)^2\left(\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^{\Delta}\right)^2.
\end{aligned}$$

and

$$\operatorname{tr}\left(\left(a_l a_m'\Sigma\otimes e_i e_j'\right)\left(\Sigma^{-1}\otimes M_{\mathbf{1}_n}\right)\right)=\left(a_m' a_l e_j' M_{\mathbf{1}_n}e_i\right)$$

with

$$\begin{aligned}
&\textstyle\sum_{l,m=1}^{T-1}\sum_{i,j=1}^{n}\operatorname{tr}\left(\left(a_l a_m'\Sigma\otimes e_i e_j'\right)\left(\Sigma^{-1}\otimes M_{\mathbf{1}_n}\right)\right)\left(h'\left(a_m a_l'\Sigma\otimes e_i e_j'\right)h\right)\\
=&\textstyle\sum_{m=1}^{T-1}\sum_{i,j=1}^{n}e_j' M_{\mathbf{1}_n}e_i\left(h'\left(a_m a_m'\Sigma\otimes e_i e_j'\right)h\right)\\
=&\textstyle\sum_{i,j=1}^{n}e_j' M_{\mathbf{1}_n}e_i\left(h'\left(\Sigma\otimes e_i e_j'\right)h\right)=\sigma^{\tau\tau}\sum_{i,j=1}^{n}\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}e_i e_i' M_{\mathbf{1}_n}e_j e_j' M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^{\Delta}\\
=&\sigma^{\tau\tau}\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^{\Delta}.
\end{aligned}$$

Also,

$$\begin{aligned}
&\operatorname{tr}\left[\left(a_l a_m'\Sigma\otimes e_i e_j'\right)\left(\Sigma^{-1}B M_{\mathbf{1}_T}\xi_\tau\xi_\tau' M_{\mathbf{1}_T}B'\Sigma^{-1}\otimes M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^{\Delta}\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}\right)\right]\\
=&a_m' B M_{\mathbf{1}_T}\xi_\tau\xi_\tau' M_{\mathbf{1}_T}B'\Sigma^{-1}a_l e_j' M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^{\Delta}\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}e_i
\end{aligned}$$

such that

$$\begin{aligned}
&\frac{\sum_{l,m=1}^{T-1}\sum_{i,j=1}^{n}\operatorname{tr}\left[\left(a_m a_l'\Sigma\otimes e_i e_j'\right)\left(\Sigma^{-1}B M_{\mathbf{1}_T}\xi_\tau\xi_\tau' M_{\mathbf{1}_T}B'\Sigma^{-1}\otimes M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^{\Delta}\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}\right)\right]\left(h'\left(a_l a_m'\Sigma\otimes e_i e_j'\right)h\right)}{\sigma^{\tau\tau}\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^{\Delta}}\\
=&\frac{\sum_{l,m=1}^{T-1}\sum_{i,j=1}^{n}\xi_\tau' M_{\mathbf{1}_T}B'\Sigma^{-1}a_l a_m' B M_{\mathbf{1}_T}\xi_\tau\xi_\tau' M_{\mathbf{1}_T}B'\Sigma^{-1}a_m a_l' B M_{\mathbf{1}_T}\xi_\tau\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}e_i e_j' M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^{\Delta}e_j' M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^{\Delta}\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}e_i}{\sigma^{\tau\tau}\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^{\Delta}}\\
=&\textstyle\sum_{l,m=1}^{T-1}\xi_\tau' M_{\mathbf{1}_T}B'\Sigma^{-1}a_l a_l' B M_{\mathbf{1}_T}\xi_\tau\xi_\tau' M_{\mathbf{1}_T}B'\Sigma^{-1}a_m a_m' B M_{\mathbf{1}_T}\xi_\tau\left(\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^{\Delta}\right)/\sigma^{\tau\tau}\\
=&\left(\xi_\tau' M_{\mathbf{1}_T}B'\Sigma^{-1}B M_{\mathbf{1}_T}\xi_\tau\right)\left(\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^{\Delta}\right)=\sigma^{\tau\tau}\left(\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^{\Delta}\right)
\end{aligned}$$

and therefore

$$\textstyle\sum_{l,m=1}^{T-1}\sum_{i,j=1}^{n}\operatorname{tr}\left(\left(a_l a_m'\Sigma\otimes e_i e_j'\right)H\right)\left(h'\left(a_m a_l'\Sigma\otimes e_i e_j'\right)h\right)=0.$$

It thus follows that

$$\begin{aligned}
\operatorname{tr}V_\Omega\left(hh'\otimes hh'\right)&=2\left(\sigma^{\tau\tau}\right)^2\left(\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^{\Delta}\right)^2,\\
\operatorname{tr}V_\Omega\left(H\otimes hh'\right)&=\left(T-2\right)\sigma^{\tau\tau}\left(\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^{\Delta}\right)
\end{aligned}$$

Collecting these results and substitution in (2.9) and (2.10) leads to

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

and

$$\frac{\operatorname{tr} V_\Omega\left(H \otimes hh'\right)}{h'\Omega h} = T - 2$$

which establishes the results for T_1.

For the test T_2 note that in the same way as for $\operatorname{tr}\left(V_\Omega\left(hh' \otimes hh'\right)\right)$,

$$\begin{aligned}\operatorname{tr}\left(V_\Omega\left(xx' \otimes xx'\right)\right) &= \textstyle\sum_{i,j=1}^{n}\left(x'\left(\Sigma \otimes e_i e_j'\right)x\right)^2 \\ &\quad + \textstyle\sum_{l,m=1}^{(T-1)}\sum_{i,j=1}^{n}\left(x'\left(a_l a_m'\Sigma \otimes e_i e_j'\right)x\right)\left(x'\left(a_m a_l'\Sigma \otimes e_i e_j'\right)x\right)\end{aligned}$$

and that

$$x = \sqrt{n}\frac{BM_{\mathbf{1}_T}\xi_\tau \otimes M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta}{\xi_\tau' M_{\mathbf{1}_T}B'BM_{\mathbf{1}_T}\xi_\tau \tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta}.$$

It follows that

(A.5) $$x'\Omega x = \frac{\xi_\tau' M_{\mathbf{1}_T}B'\Sigma BM_{\mathbf{1}_T}\xi_\tau \tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta}{\left(\xi_\tau' M_{\mathbf{1}_T}B'BM_{\mathbf{1}_T}\xi_\tau\right)^2\left(\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta\right)^2}.$$

As before for the GLS case, the denominator of x cancels in $\operatorname{tr}\left(V_\Omega\left(xx' \otimes xx'\right)\right)/\left(x'\Omega x\right)^2$ such that without loss of generality we set $x = BM_{\mathbf{1}_T}\xi_\tau \otimes M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta$. Then consider

$$\begin{aligned}\textstyle\sum_{i,j=1}^{n}\left(x'\left(\Sigma \otimes e_i e_j'\right)x\right)^2 &= \textstyle\sum_{i,j=1}^{n}\left(\xi_\tau' M_{\mathbf{1}_T}B'\Sigma BM_{\mathbf{1}_T}\xi_\tau \tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}e_i e_j' M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta\right)^2 \\ &= \left(\xi_\tau' M_{\mathbf{1}_T}B'\Sigma BM_{\mathbf{1}_T}\xi_\tau\right)^2 \textstyle\sum_{i,j=1}^{n}\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}e_i e_i' M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta \tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}e_j e_j' M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta \\ &= \left(\xi_\tau' M_{\mathbf{1}_T}B'\Sigma BM_{\mathbf{1}_T}\xi_\tau\right)^2\left(\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta\right)^2\end{aligned}$$

and

$$\begin{aligned}&\textstyle\sum_{l,m=1}^{(T-1)}\sum_{i,j=1}^{n}\left(x'\left(a_l a_m'\Sigma \otimes e_i e_j'\right)x\right)\left(x'\left(a_m a_l'\Sigma \otimes e_i e_j'\right)x\right) \\ = &\textstyle\sum_{l,m=1}^{(T-1)}\sum_{i,j=1}^{n}\xi_\tau' M_{\mathbf{1}_T}B'a_l a_m'\Sigma BM_{\mathbf{1}_T}\xi_\tau \tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}e_i e_j' M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta \xi_\tau' M_{\mathbf{1}_T}B'a_m a_l'\Sigma BM_{\mathbf{1}_T}\xi_\tau \tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}e_i e_j' M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta \\ = &\textstyle\sum_{l,m=1}^{(T-1)}\sum_{i,j=1}^{n}\xi_\tau' M_{\mathbf{1}_T}B'a_l a_l'\Sigma BM_{\mathbf{1}_T}\xi_\tau \tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}e_i e_i' M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta \xi_\tau' M_{\mathbf{1}_T}B'a_m a_m'\Sigma BM_{\mathbf{1}_T}\xi_\tau \tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}e_j e_j' M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta \\ = &\left(\xi_\tau' M_{\mathbf{1}_T}B'\Sigma BM_{\mathbf{1}_T}\xi_\tau\right)^2\left(\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta\right)^2\end{aligned}$$

such that

$$\operatorname{tr}\left(V_\Omega\left(xx' \otimes xx'\right)\right) = 2\left(\xi_\tau' M_{\mathbf{1}_T}B'\Sigma BM_{\mathbf{1}_T}\xi_\tau\right)^2\left(\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta\right)^2.$$

Using (A.5) now leads to

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

which establishes the result. ■

Proof of Theorem 3.2. Note that $\tilde{\Sigma}^{1/2}M_{\mathbf{1}_T}B'C\Sigma_C^{-1}C'BM_{\mathbf{1}_T}\tilde{\Sigma}^{1/2}$ is a projection matrix. For $r=1$ it thus follows that $\sigma^{\tau\tau}$ 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 3.4 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. ■

Proof of Theorem 3.3. We derive explicit versions of the formulas (2.9) and (2.10). We first simplify the expression for the GLS estimator

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

where

$$\begin{aligned}\Omega_\Delta^{-1}Z_\Delta\left(Z_\Delta'\Omega_\Delta^{-1}Z_\Delta\right)^{-1}Z_\Delta'\Omega_\Delta^{-1} &= \textstyle\sum_{t_1,t_2=1}^{T-1}\left(\Sigma_\Delta^{-1}a_{t_1}a_{t_1}'\otimes\mathbf{1}_n\right)\frac{\Sigma_\Delta}{n}\left(a_{t_2}a_{t_2}'\Sigma_\Delta^{-1}\otimes\mathbf{1}_n'\right)\\ &= \Sigma_\Delta^{-1}\otimes\frac{\mathbf{1}_n\mathbf{1}_n'}{n}.\end{aligned}$$

Also let $\sigma_\Delta^{t,s} = a_{t-1}'\Sigma_\Delta^{-1}a_{s-1}$, $\sigma_{\Delta;t,s} = a_{t-1}'\Sigma_\Delta a_{s-1}$. Then,

$$h_\Delta = \frac{\left(\Sigma_\Delta^{-1}a_{\tau-1}\otimes M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta\right)}{a_{\tau-1}'\Sigma_\Delta^{-1}a_{\tau-1}\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta}$$

and

$$H_\Delta = \Sigma_\Delta^{-1}\otimes M_{\mathbf{1}_n} - \frac{\Sigma_\Delta^{-1}a_{\tau-1}a_{\tau-1}'\Sigma_\Delta^{-1}\otimes M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}}{a_{\tau-1}'\Sigma_\Delta^{-1}a_{\tau-1}\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta}.$$

Note that h_Δ and H_Δ differ from h and H only in that $a_{\tau-1}$ replaces $BM_{\mathbf{1}_T}\xi_\tau$ and Σ_Δ replaces Σ. Since $\operatorname{tr}V_\Omega\left(hh'\otimes hh'\right)/\left(h'\Omega h\right)^2$ and $\operatorname{tr}V_\Omega\left(H\otimes hh'\right)/h'\Omega h$ do not depend on $BM_{\mathbf{1}_T}\xi_\tau$ or Σ_Δ it follows that

$$\frac{\operatorname{tr}V_{\Omega_\Delta}\left(h_\Delta h_\Delta'\otimes h_\Delta h_\Delta'\right)}{\left(h_\Delta'\Omega h_\Delta\right)^2} = 2$$

and

$$\frac{\operatorname{tr}V_{\Omega_\Delta}\left(H_\Delta\otimes h_\Delta h_\Delta'\right)}{h_\Delta'\Omega h_\Delta} = T-2$$

as before and the result is established for T_1. For the test T_2 we note that

$$x_\Delta = \sqrt{n}\frac{a_{\tau-1}\otimes M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta}{\tilde{\Upsilon}_\tau^{\Delta\prime}M_{\mathbf{1}_n}\tilde{\Upsilon}_\tau^\Delta}$$

which differs from x only in that $a_{\tau-1}$ replaces $BM_{\mathbf{1}_T}\xi_\tau$. Since $\operatorname{tr}\left(V_\Omega\left(xx'\otimes xx'\right)\right)/\left(x'\Omega x\right)^2$ does not depend on $BM_{\mathbf{1}_T}\xi_\tau$ it follows that

$$\frac{\operatorname{tr}V_{\Omega_\Delta}\left(x_\Delta x_\Delta'\otimes x_\Delta x_\Delta'\right)}{\left(x_\Delta'\Omega x_\Delta\right)^2} = 2$$

as before. This establishes the result for T_2. ■

Acknowledgements:

We thank Matthew Harding, Alan Grant, Anne Hall and Jie Yang for excellent research assistance. Whitney Newey engaged in helpful conversations. We thank the Editor, Associate Editor and two referees who provided helpful comments on an earlier draft of the paper. Kuersteiner gratefully acknowledges financial support from NSF Grants SES 0095132 and 0523186.

References

Arellano, M., 1987, Computing Robust Standard Errors for the Within-Groups Estimators. Oxford Bulletin of Economics and Statistics XLIX, 431–434.

Bertrand, M., E. Duflo, and S. Mullainathan, 2004, How much should we trust Differences-In-Differences Estimates. Quarterly Journal of Economics 119, 249–275.

Chow, Y. S., and H. Teicher, 1997, Probability Theory: independence, interchantability, martingales - 3rd edition. Springer-Verlag New York, Inc.

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

Hansen, C. B., 2007, Generalized Least Squares Inference In Panel and Multilevel Models with serial correlation and fixed effects. Journal of Econometrics 140, 670–694.

Hausman, J., 1978, Specification tests in econometrics. Econometrica 46, 1251–71.

Hausman, J., 1995, State Regulation of Cellular Prices. Wireless Communications Forum 3.

Hausman, J., 2002, Mobile Telephone, in: M. Cave (Ed.), Handbook of Telecommunications Economics, North Holland, Amsterdam, pp 564–604.

Kezdi, G., 2004, Robust Standard Error Estimation in Fixed-Effects Panel Models. Hungarian Statistical Review 9, 95–116.

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, 381–394.

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

Moulton, B. R., 1986, Random Group Effects and the Precision of Regression Estimates. Journal of Econometrics 32, 385–397.

Mundlak, Y., 1978, On the Pooling of Time Series and Cross Section Data. Econometrica 46, 69–85.

Rothenberg, T. J., 1984a, Approximate Normality of Generalized Least Squares Estimates. Econometrica 52, 811–825.

Rothenberg, T. J., 1984b, Hypothesis Testing in Linear Models when the Error Covariance Matrix is Nonscalar. Econometrica 52, 827–842.

Rothenberg, T. J., 1988, Approximate Power Functions for Some Robust Tests of Regression Coefficients. Econometrica 56, 997–1019.

Table 1a: Results for the Levels Specification

gamma	rho	T	GLS-SC	GLS	GLS-AR	ROLS	OLS	GLS-SC	GLS	GLS-AR	ROLS	OLS	GLS-SC	GLS	GLS-AR	ROLS	OLS
			Full Sample					3 Periods					2 Periods				
0	0	5	0.0441	0.0624	0.0725	0.0524	0.0518	0.0452	0.0595	0.0709	0.0524	0.0310	0.0480	0.0524	0.0533	0.0524	0.0560
0.1			0.0578	0.0803	0.0903	0.0672	0.0663	0.0594	0.0764	0.0889	0.0672	0.0410	0.0607	0.0672	0.0697	0.0672	0.0699
0.6			0.5651	0.6271	0.6447	0.5616	0.5640	0.5717	0.6194	0.6438	0.5616	0.4723	0.5428	0.5616	0.5739	0.5616	0.5636
1			0.9430	0.9594	0.9639	0.9342	0.9384	0.9449	0.9576	0.9640	0.9342	0.9024	0.9283	0.9342	0.9419	0.9342	0.9363
0	0	10	0.0475	0.0907	0.0590	0.0527	0.0510	0.0466	0.0606	0.0586	0.0520	0.0174	0.0470	0.0527	0.0524	0.0527	0.0564
0.1			0.0714	0.1253	0.0884	0.0804	0.0797	0.0706	0.0893	0.0882	0.0773	0.0287	0.0738	0.0804	0.0812	0.0804	0.0851
0.6			0.7912	0.8696	0.8576	0.8336	0.8420	0.8211	0.8505	0.8580	0.8044	0.6630	0.8217	0.8336	0.8460	0.8336	0.8400
1			0.9955	0.9983	0.9987	0.9978	0.9984	0.9976	0.9983	0.9988	0.9960	0.9877	0.9975	0.9978	0.9985	0.9978	0.9981
0	0	15	0.0587	0.1367	0.0556	0.0527	0.0513	0.0472	0.0608	0.0555	0.0515	0.0137	0.0472	0.0527	0.0524	0.0527	0.0568
0.1			0.0933	0.1889	0.0978	0.0912	0.0924	0.0823	0.1018	0.0976	0.0846	0.0271	0.0834	0.0912	0.0942	0.0912	0.0977
0.6			0.8875	0.9489	0.9469	0.9360	0.9431	0.9252	0.9411	0.9471	0.8825	0.7460	0.9303	0.9360	0.9433	0.9360	0.9414
1			0.9991	0.9998	0.9999	0.9999	1.0000	0.9998	1.0000	0.9999	0.9991	0.9967	0.9999	0.9999	1.0000	0.9999	0.9999
0	0	20	0.0847	0.2070	0.0535	0.0496	0.0518	0.0441	0.0572	0.0536	0.0490	0.0116	0.0446	0.0496	0.0525	0.0496	0.0569
0.1			0.1328	0.2771	0.1158	0.1056	0.1094	0.0953	0.1180	0.1155	0.0877	0.0267	0.0971	0.1056	0.1100	0.1056	0.1163
0.6			0.9442	0.9809	0.9891	0.9845	0.9887	0.9806	0.9860	0.9891	0.9171	0.8012	0.9824	0.9845	0.9887	0.9845	0.9875
1			0.9999	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9983	1.0000	1.0000	1.0000	1.0000	1.0000

Table 1b: Results for the Levels Specification

gamma	rho	T	GLS-SC	GLS	GLS-AR	ROLS	OLS	GLS-SC	GLS	GLS-AR	ROLS	OLS	GLS-SC	GLS	GLS-AR	ROLS	OLS
			Full Sample					3 Periods					2 Periods				
0	0.4	5	0.0425	0.0604	0.1005	0.0525	0.1010	0.0435	0.0581	0.1051	0.0525	0.0649	0.0473	0.0525	0.0682	0.0525	0.0555
0.1			0.0519	0.0726	0.1145	0.0627	0.1169	0.0539	0.0692	0.1188	0.0627	0.0762	0.0568	0.0627	0.0789	0.0627	0.0664
0.6			0.4496	0.5121	0.5516	0.4548	0.5766	0.4322	0.4796	0.5391	0.4548	0.4910	0.4356	0.4548	0.5046	0.4548	0.4563
1			0.8648	0.8962	0.9013	0.8572	0.9171	0.8467	0.8729	0.8912	0.8572	0.8800	0.8458	0.8572	0.8872	0.8572	0.8605
0	0.4	10	0.0440	0.0871	0.0702	0.0533	0.1386	0.0465	0.0599	0.0713	0.0536	0.0484	0.0478	0.0533	0.0578	0.0533	0.0557
0.1			0.0587	0.1081	0.0873	0.0687	0.1642	0.0588	0.0748	0.0862	0.0680	0.0607	0.0624	0.0687	0.0722	0.0687	0.0714
0.6			0.5747	0.6939	0.6637	0.5839	0.7579	0.5612	0.6082	0.6203	0.5892	0.5730	0.5670	0.5839	0.6057	0.5839	0.5889
1			0.9506	0.9753	0.9708	0.9465	0.9836	0.9401	0.9536	0.9548	0.9489	0.9458	0.9411	0.9465	0.9548	0.9465	0.9483
0	0.4	15	0.0511	0.1266	0.0623	0.0551	0.1557	0.0475	0.0622	0.0632	0.0541	0.0372	0.0495	0.0551	0.0552	0.0551	0.0578
0.1			0.0710	0.1558	0.0856	0.0738	0.1902	0.0664	0.0844	0.0848	0.0741	0.0525	0.0671	0.0738	0.0753	0.0738	0.0763
0.6			0.6575	0.8024	0.7499	0.6832	0.8483	0.6604	0.7042	0.7080	0.6746	0.6189	0.6667	0.6832	0.6986	0.6832	0.6891
1			0.9760	0.9921	0.9910	0.9816	0.9962	0.9780	0.9841	0.9844	0.9791	0.9714	0.9796	0.9816	0.9846	0.9816	0.9829
0	0.4	20	0.0634	0.1736	0.0595	0.0533	0.1708	0.0453	0.0599	0.0617	0.0531	0.0274	0.0481	0.0533	0.0531	0.0533	0.0571
0.1			0.0882	0.2126	0.0910	0.0789	0.2153	0.0703	0.0888	0.0894	0.0760	0.0437	0.0723	0.0789	0.0800	0.0789	0.0827
0.6			0.7516	0.8857	0.8454	0.7931	0.9234	0.7758	0.8104	0.8149	0.7487	0.6642	0.7797	0.7931	0.8071	0.7931	0.8000
1			0.9901	0.9977	0.9981	0.9952	0.9995	0.9945	0.9963	0.9962	0.9910	0.9839	0.9945	0.9952	0.9963	0.9952	0.9958

Table 1c: Results for the Levels Specification

gamma	rho	T	GLS-SC	GLS	GLS-AR	ROLS	OLS	GLS-SC	GLS	GLS-AR	ROLS	OLS	GLS-SC	GLS	GLS-AR	ROLS	OLS
			Full Sample					3 Periods					2 Periods				
0	0.8	5	0.0347	0.0512	0.1053	0.0449	0.1504	0.0383	0.0510	0.1192	0.0449	0.0981	0.0402	0.0449	0.1069	0.0449	0.0559
0.1			0.0455	0.0648	0.1183	0.0526	0.1634	0.0467	0.0607	0.1309	0.0526	0.1082	0.0469	0.0526	0.1181	0.0526	0.0635
0.6			0.4264	0.4918	0.5205	0.3349	0.5344	0.3748	0.4220	0.4876	0.3349	0.4476	0.3177	0.3349	0.4686	0.3349	0.3432
1			0.8522	0.8863	0.8758	0.7133	0.8615	0.7927	0.8253	0.8389	0.7133	0.8059	0.6988	0.7133	0.8233	0.7133	0.7212
0	0.8	10	0.0395	0.0796	0.0735	0.0522	0.2747	0.0421	0.0556	0.0913	0.0524	0.1089	0.0464	0.0522	0.1041	0.0522	0.0567
0.1			0.0495	0.0956	0.0877	0.0563	0.2875	0.0484	0.0631	0.0991	0.0571	0.1161	0.0510	0.0563	0.1113	0.0563	0.0617
0.6			0.4432	0.5694	0.5462	0.2503	0.5778	0.2899	0.3319	0.3780	0.2634	0.3830	0.2364	0.2503	0.3612	0.2503	0.2541
1			0.8701	0.9280	0.9143	0.5628	0.8487	0.6602	0.7030	0.7262	0.5872	0.7130	0.5448	0.5628	0.6834	0.5628	0.5646
0	0.8	15	0.0500	0.1235	0.0641	0.0534	0.3269	0.0457	0.0593	0.0772	0.0529	0.1020	0.0482	0.0534	0.0852	0.0534	0.0560
0.1			0.0618	0.1423	0.0794	0.0592	0.3351	0.0520	0.0673	0.0841	0.0597	0.1099	0.0537	0.0592	0.0904	0.0592	0.0618
0.6			0.4566	0.6300	0.5551	0.2477	0.6160	0.2710	0.3123	0.3322	0.2601	0.3666	0.2331	0.2477	0.3186	0.2477	0.2493
1			0.8714	0.9439	0.9297	0.5534	0.8645	0.6191	0.6633	0.6728	0.5790	0.6950	0.5355	0.5534	0.6357	0.5534	0.5542
0	0.8	20	0.0611	0.1711	0.0639	0.0544	0.3786	0.0454	0.0599	0.0772	0.0550	0.0910	0.0485	0.0544	0.0773	0.0544	0.0565
0.1			0.0749	0.1911	0.0790	0.0607	0.3900	0.0517	0.0673	0.0834	0.0610	0.0980	0.0550	0.0607	0.0836	0.0607	0.0629
0.6			0.4788	0.6836	0.5753	0.2544	0.6702	0.2726	0.3143	0.3297	0.2718	0.3546	0.2394	0.2544	0.3061	0.2544	0.2548
1			0.8760	0.9540	0.9383	0.5672	0.8937	0.6282	0.6724	0.6714	0.6035	0.6907	0.5492	0.5672	0.6311	0.5672	0.5663

Table 1d: Results for the Levels Specification

gamma	rho	T	GLS-SC	GLS	GLS-AR	ROLS	OLS	GLS-SC	GLS	GLS-AR	ROLS	OLS	GLS-SC	GLS	GLS-AR	ROLS	OLS
			Full Sample					3 Periods					2 Periods				
0	0.9	5	0.0331	0.0486	0.0935	0.0384	0.1553	0.0336	0.0463	0.1079	0.0384	0.0989	0.0337	0.0384	0.1014	0.0384	0.0561
0.1			0.0427	0.0606	0.1069	0.0453	0.1674	0.0424	0.0553	0.1183	0.0453	0.1101	0.0400	0.0453	0.1116	0.0453	0.0643
0.6			0.4202	0.4833	0.5190	0.3223	0.5478	0.3660	0.4146	0.4857	0.3223	0.4563	0.3048	0.3223	0.4646	0.3223	0.3491
1			0.8462	0.8826	0.8847	0.7059	0.8729	0.7881	0.8229	0.8454	0.7059	0.8170	0.6902	0.7059	0.8261	0.7059	0.7321
0	0.9	10	0.0408	0.0819	0.0696	0.0471	0.3105	0.0368	0.0493	0.0931	0.0469	0.1202	0.0415	0.0471	0.1214	0.0471	0.0582
0.1			0.0515	0.0965	0.0829	0.0499	0.3204	0.0416	0.0559	0.1008	0.0495	0.1255	0.0444	0.0499	0.1284	0.0499	0.0610
0.6			0.4215	0.5454	0.5421	0.1984	0.5556	0.2474	0.2879	0.3523	0.2112	0.3443	0.1851	0.1984	0.3346	0.1984	0.2098
1			0.8438	0.9109	0.9147	0.4542	0.8014	0.5915	0.6381	0.6841	0.4822	0.6369	0.4357	0.4542	0.6153	0.4542	0.4637
0	0.9	15	0.0577	0.1382	0.0617	0.0483	0.3832	0.0402	0.0535	0.0810	0.0487	0.1215	0.0437	0.0483	0.1079	0.0483	0.0549
0.1			0.0697	0.1547	0.0766	0.0516	0.3890	0.0442	0.0591	0.0857	0.0517	0.1286	0.0461	0.0516	0.1130	0.0516	0.0586
0.6			0.4343	0.6007	0.5451	0.1669	0.5726	0.2062	0.2423	0.2774	0.1770	0.3030	0.1551	0.1669	0.2689	0.1669	0.1724
1			0.8338	0.9171	0.9257	0.3701	0.7787	0.4891	0.5387	0.5681	0.4004	0.5581	0.3531	0.3701	0.5063	0.3701	0.3741
0	0.9	20	0.0811	0.2009	0.0598	0.0531	0.4547	0.0416	0.0564	0.0782	0.0526	0.1166	0.0471	0.0531	0.1046	0.0531	0.0576
0.1			0.0924	0.2214	0.0749	0.0539	0.4604	0.0461	0.0607	0.0825	0.0549	0.1215	0.0482	0.0539	0.1097	0.0539	0.0599
0.6			0.4513	0.6432	0.5526	0.1535	0.6137	0.1885	0.2221	0.2481	0.1668	0.2768	0.1429	0.1535	0.2389	0.1535	0.1571
1			0.8258	0.9221	0.9273	0.3351	0.7902	0.4484	0.4959	0.5150	0.3723	0.5136	0.3180	0.3351	0.4502	0.3351	0.3368

Table 2a: Results for the First Difference Specification

gamma	rho	T	GLS-SC	GLS	GLS-AR	ROLS	OLS	GLS-SC	GLS	GLS-AR	ROLS	OLS	GLS-SC	GLS	GLS-AR	ROLS	OLS
			Full Sample					3 Periods					2 Periods				
0	0	5	0.0441	0.0624	0.0685	0.0515	0.0386	0.0423	0.0558	0.0777	0.0515	0.0833	0.0432	0.0531	0.0683	0.0518	0.0313
0.1			0.0578	0.0803	0.0855	0.0590	0.0446	0.0532	0.0691	0.0930	0.0590	0.0948	0.0507	0.0609	0.0771	0.0594	0.0365
0.6			0.5651	0.6271	0.6364	0.3290	0.2844	0.4673	0.5150	0.5513	0.3290	0.4059	0.3506	0.3818	0.4114	0.3259	0.2462
1			0.9430	0.9594	0.9616	0.6987	0.6601	0.8822	0.9056	0.9172	0.6987	0.7716	0.7537	0.7798	0.7993	0.7017	0.6169
0	0	10	0.0475	0.0907	0.0574	0.0490	0.0445	0.0420	0.0556	0.0808	0.0490	0.2035	0.0428	0.0524	0.0696	0.0506	0.1104
0.1			0.0714	0.1253	0.0861	0.0579	0.0515	0.0522	0.0687	0.0944	0.0579	0.2160	0.0505	0.0614	0.0791	0.0579	0.1214
0.6			0.7912	0.8696	0.8552	0.3241	0.3012	0.4646	0.5113	0.5458	0.3241	0.5878	0.3481	0.3799	0.4066	0.3217	0.4453
1			0.9955	0.9983	0.9986	0.6963	0.6794	0.8787	0.9012	0.9138	0.6963	0.8837	0.7479	0.7732	0.7928	0.6932	0.7998
0	0	15	0.0587	0.1367	0.0549	0.0500	0.0473	0.0392	0.0519	0.0786	0.0500	0.2383	0.0405	0.0502	0.0694	0.0481	0.1257
0.1			0.0933	0.1889	0.0965	0.0571	0.0536	0.0499	0.0647	0.0933	0.0571	0.2522	0.0481	0.0592	0.0786	0.0553	0.1383
0.6			0.8875	0.9489	0.9464	0.3251	0.3067	0.4632	0.5124	0.5488	0.3251	0.6158	0.3476	0.3811	0.4080	0.3198	0.4790
1			0.9991	0.9998	0.9999	0.6925	0.6815	0.8799	0.9022	0.9145	0.6925	0.8979	0.7518	0.7780	0.7965	0.6930	0.8232
0	0	20	0.0847	0.2070	0.0531	0.0480	0.0472	0.0380	0.0514	0.0807	0.0480	0.2562	0.0406	0.0498	0.0708	0.0487	0.1165
0.1			0.1328	0.2771	0.1151	0.0563	0.0548	0.0480	0.0628	0.0954	0.0563	0.2683	0.0487	0.0588	0.0797	0.0566	0.1269
0.6			0.9442	0.9809	0.9891	0.3276	0.3115	0.4653	0.5146	0.5490	0.3276	0.6394	0.3506	0.3830	0.4095	0.3246	0.4630
1			0.9999	1.0000	1.0000	0.7001	0.6899	0.8789	0.9018	0.9129	0.7001	0.9069	0.7521	0.7773	0.7945	0.6982	0.8123

Table 2b: Results for the First Difference Specification

gamma	rho	T	GLS-SC	GLS	GLS-AR	ROLS	OLS	GLS-SC	GLS	GLS-AR	ROLS	OLS	GLS-SC	GLS	GLS-AR	ROLS	OLS
			Full Sample					3 Periods					2 Periods				
0	0.4	5	0.0425	0.0604	0.0920	0.0522	0.0373	0.0426	0.0565	0.0937	0.0522	0.0718	0.0432	0.0516	0.0911	0.0505	0.0244
0.1			0.0519	0.0726	0.1055	0.0616	0.0453	0.0524	0.0679	0.1092	0.0616	0.0833	0.0495	0.0604	0.1013	0.0569	0.0284
0.6			0.4496	0.5121	0.5452	0.4226	0.3685	0.4389	0.4878	0.5326	0.4226	0.4764	0.3104	0.3405	0.3979	0.3222	0.2124
1			0.8648	0.8962	0.9008	0.8244	0.7934	0.8557	0.8811	0.8904	0.8244	0.8616	0.6891	0.7184	0.7544	0.6899	0.5708
0	0.4	10	0.0440	0.0871	0.0677	0.0497	0.0420	0.0432	0.0562	0.0881	0.0497	0.1783	0.0432	0.0527	0.0843	0.0505	0.0943
0.1			0.0587	0.1081	0.0844	0.0602	0.0523	0.0532	0.0693	0.1028	0.0602	0.1991	0.0500	0.0599	0.0912	0.0563	0.1033
0.6			0.5747	0.6939	0.6582	0.4252	0.3967	0.4886	0.5371	0.5691	0.4252	0.6598	0.3042	0.3340	0.3766	0.2956	0.3896
1			0.9506	0.9753	0.9702	0.8272	0.8142	0.8960	0.9170	0.9209	0.8272	0.9431	0.6793	0.7087	0.7335	0.6467	0.7380
0	0.4	15	0.0511	0.1266	0.0609	0.0502	0.0456	0.0417	0.0554	0.0878	0.0502	0.2204	0.0406	0.0502	0.0834	0.0478	0.1171
0.1			0.0710	0.1558	0.0840	0.0604	0.0555	0.0522	0.0680	0.1033	0.0604	0.2394	0.0475	0.0576	0.0912	0.0544	0.1262
0.6			0.6575	0.8024	0.7475	0.4239	0.4025	0.4971	0.5475	0.5806	0.4239	0.6987	0.2929	0.3219	0.3673	0.2822	0.4188
1			0.9760	0.9921	0.9908	0.8272	0.8210	0.9052	0.9243	0.9306	0.8272	0.9546	0.6692	0.6990	0.7235	0.6292	0.7610
0	0.4	20	0.0634	0.1736	0.0587	0.0492	0.0466	0.0386	0.0527	0.0906	0.0492	0.2462	0.0405	0.0501	0.0863	0.0487	0.1040
0.1			0.0882	0.2126	0.0898	0.0599	0.0567	0.0503	0.0656	0.1074	0.0599	0.2670	0.0457	0.0569	0.0929	0.0547	0.1150
0.6			0.7516	0.8857	0.8437	0.4271	0.4103	0.5057	0.5545	0.5874	0.4271	0.7244	0.2904	0.3208	0.3677	0.2807	0.3948
1			0.9901	0.9977	0.9980	0.8284	0.8245	0.9083	0.9260	0.9312	0.8284	0.9596	0.6611	0.6909	0.7157	0.6243	0.7375

Table 2c: Results for the First Difference Specification

gamma	rho	T	GLS-SC	GLS	GLS-AR	ROLS	OLS	GLS-SC	GLS	GLS-AR	ROLS	OLS	GLS-SC	GLS	GLS-AR	ROLS	OLS
			Full Sample					3 Periods					2 Periods				
0	0.8	5	0.0347	0.0512	0.0933	0.0452	0.0467	0.0361	0.0496	0.0926	0.0452	0.0725	0.0361	0.0453	0.1139	0.0447	0.0407
0.1			0.0455	0.0648	0.1076	0.0571	0.0580	0.0471	0.0614	0.1072	0.0571	0.0869	0.0421	0.0512	0.1208	0.0505	0.0457
0.6			0.4264	0.4918	0.5193	0.4664	0.4625	0.4338	0.4839	0.5190	0.4664	0.5364	0.2531	0.2804	0.3786	0.2759	0.2443
1			0.8522	0.8863	0.8824	0.8693	0.8757	0.8532	0.8812	0.8819	0.8693	0.9108	0.6001	0.6325	0.6983	0.6200	0.5824
0	0.8	10	0.0395	0.0796	0.0694	0.0486	0.0434	0.0408	0.0534	0.0708	0.0486	0.1492	0.0414	0.0506	0.0878	0.0495	0.0754
0.1			0.0495	0.0956	0.0836	0.0598	0.0541	0.0504	0.0659	0.0849	0.0598	0.1712	0.0453	0.0549	0.0919	0.0538	0.0807
0.6			0.4432	0.5694	0.5452	0.4965	0.4857	0.4708	0.5198	0.5417	0.4965	0.7004	0.2205	0.2458	0.2918	0.2392	0.2894
1			0.8701	0.9280	0.9155	0.8975	0.8988	0.8860	0.9084	0.9138	0.8975	0.9672	0.5228	0.5574	0.5917	0.5471	0.6069
0	0.8	15	0.0500	0.1235	0.0620	0.0484	0.0450	0.0417	0.0549	0.0654	0.0484	0.1896	0.0422	0.0515	0.0794	0.0502	0.0924
0.1			0.0618	0.1423	0.0782	0.0615	0.0579	0.0536	0.0696	0.0810	0.0615	0.2123	0.0467	0.0564	0.0851	0.0535	0.0970
0.6			0.4566	0.6300	0.5533	0.5022	0.4946	0.4825	0.5325	0.5479	0.5022	0.7493	0.1996	0.2241	0.2564	0.2134	0.2938
1			0.8714	0.9439	0.9301	0.9026	0.9050	0.9008	0.9205	0.9249	0.9026	0.9771	0.4728	0.5046	0.5301	0.4849	0.5869
0	0.8	20	0.0611	0.1711	0.0624	0.0488	0.0465	0.0406	0.0546	0.0676	0.0488	0.2224	0.0409	0.0504	0.0856	0.0498	0.0850
0.1			0.0749	0.1911	0.0778	0.0614	0.0591	0.0524	0.0687	0.0829	0.0614	0.2496	0.0449	0.0542	0.0891	0.0525	0.0890
0.6			0.4788	0.6836	0.5733	0.5053	0.5013	0.4964	0.5451	0.5600	0.5053	0.7770	0.1709	0.1935	0.2341	0.1810	0.2408
1			0.8760	0.9540	0.9384	0.9016	0.9063	0.9044	0.9229	0.9266	0.9016	0.9808	0.4042	0.4371	0.4691	0.4105	0.4874

Table 2d: Results for the First Difference Specification

gamma	rho	T	GLS-SC	GLS	GLS-AR	ROLS	OLS	GLS-SC	GLS	GLS-AR	ROLS	OLS	GLS-SC	GLS	GLS-AR	ROLS	OLS
			Full Sample					3 Periods					2 Periods				
0	0.9	5	0.0331	0.0486	0.0830	0.0393	0.0499	0.0332	0.0448	0.0823	0.0393	0.0744	0.0309	0.0384	0.1065	0.0376	0.0481
0.1			0.0427	0.0606	0.0968	0.0503	0.0622	0.0426	0.0562	0.0957	0.0503	0.0896	0.0358	0.0446	0.1144	0.0432	0.0537
0.6			0.4202	0.4833	0.5178	0.4599	0.4988	0.4271	0.4770	0.5175	0.4599	0.5652	0.2411	0.2698	0.3722	0.2655	0.2712
1			0.8462	0.8826	0.8908	0.8677	0.9024	0.8509	0.8799	0.8908	0.8677	0.9285	0.5893	0.6223	0.6979	0.6097	0.6175
0	0.9	10	0.0408	0.0819	0.0663	0.0412	0.0465	0.0350	0.0469	0.0666	0.0412	0.1443	0.0369	0.0451	0.0894	0.0439	0.0801
0.1			0.0515	0.0965	0.0794	0.0533	0.0581	0.0450	0.0593	0.0796	0.0533	0.1672	0.0397	0.0482	0.0945	0.0479	0.0858
0.6			0.4215	0.5454	0.5428	0.4825	0.5105	0.4485	0.4976	0.5421	0.4825	0.7046	0.1909	0.2152	0.2773	0.2112	0.2797
1			0.8438	0.9109	0.9178	0.8900	0.9127	0.8750	0.9001	0.9171	0.8900	0.9691	0.4702	0.5045	0.5593	0.4954	0.5763
0	0.9	15	0.0577	0.1382	0.0599	0.0431	0.0459	0.0366	0.0495	0.0600	0.0431	0.1806	0.0394	0.0484	0.0783	0.0473	0.0901
0.1			0.0697	0.1547	0.0752	0.0561	0.0602	0.0489	0.0636	0.0761	0.0561	0.2041	0.0426	0.0520	0.0815	0.0501	0.0941
0.6			0.4343	0.6007	0.5454	0.4950	0.5155	0.4634	0.5111	0.5443	0.4950	0.7528	0.1674	0.1893	0.2260	0.1857	0.2637
1			0.8338	0.9171	0.9269	0.9027	0.9186	0.8876	0.9105	0.9262	0.9027	0.9794	0.4014	0.4346	0.4682	0.4272	0.5263
0	0.9	20	0.0811	0.2009	0.0590	0.0442	0.0479	0.0379	0.0505	0.0596	0.0442	0.2127	0.0389	0.0489	0.0790	0.0483	0.0796
0.1			0.0924	0.2214	0.0739	0.0576	0.0609	0.0491	0.0640	0.0746	0.0576	0.2377	0.0422	0.0514	0.0814	0.0512	0.0837
0.6			0.4513	0.6432	0.5516	0.5034	0.5218	0.4735	0.5226	0.5511	0.5034	0.7824	0.1366	0.1552	0.1911	0.1526	0.2004
1			0.8258	0.9221	0.9277	0.9046	0.9192	0.8920	0.9136	0.9265	0.9046	0.9825	0.3152	0.3461	0.3762	0.3402	0.4047

Table 3: Nonstationary Design with Empirical Covariance Matrix

	Full Sample				3 Periods				2 Periods			
gamma	**GLS-SC**	**GLS**	**GLS-B**	**GLS-AR**	**GLS-SC**	**GLS**	**GLS-B**	**GLS-AR**	**GLS-SC**	**GLS**	**GLS-B**	**GLS-AR**
	Level Specification											
0	0.0486	0.0966	0.0200	0.0919	0.0433	0.0565	0.0020	0.0978	0.0478	0.0537	0.0000	0.0827
0.1	0.0785	0.1434	0.0498	0.1136	0.0666	0.0847	0.0068	0.1213	0.0635	0.0702	0.0000	0.1045
0.6	0.8918	0.9423	0.8560	0.7660	0.7736	0.8097	0.6998	0.7315	0.6207	0.6379	0.6084	0.7170
1	0.9996	0.9999	0.9902	0.9903	0.9954	0.9970	0.9458	0.9832	0.9659	0.9694	0.8582	0.9822
	First Difference Specification											
0	0.0486	0.0966	0.0410	0.0874	0.0421	0.0552	0.0162	0.0935	0.0409	0.0516	0.0000	0.1811
0.1	0.0785	0.1434	0.0696	0.1078	0.0619	0.0794	0.0298	0.1113	0.0461	0.0566	0.0000	0.1908
0.6	0.8918	0.9423	0.8654	0.7616	0.7480	0.7864	0.6406	0.6301	0.2431	0.2702	0.0538	0.4398
1	0.9996	0.9999	0.9934	0.9902	0.9922	0.9947	0.9540	0.9527	0.5665	0.5978	0.5782	0.7311

Table 4: Results for the Levels Specification

gamma	rho	n	T	GLS-SC	GLS	GLS-AR	ROLS	OLS	GLS-SC	GLS	GLS-AR	ROLS	OLS	GLS-SC	GLS	GLS-AR	ROLS	OLS
				Full Sample					3 Periods					2 Periods				
0.0	0.8	15	10	0.1539	0.3431	0.0820	0.0556	0.2811	0.0388	0.0818	0.1057	0.0567	0.1205	0.0395	0.0556	0.1146	0.0556	0.0718
0.1				0.1569	0.3481	0.0854	0.0577	0.2846	0.0399	0.0856	0.1069	0.0586	0.1236	0.0407	0.0577	0.1167	0.0577	0.0735
0.6				0.2611	0.4792	0.2264	0.1134	0.3873	0.0985	0.1708	0.1907	0.1174	0.2040	0.0869	0.1134	0.1906	0.1134	0.1296
1.0				0.4254	0.6541	0.4649	0.2195	0.5355	0.2063	0.3182	0.3353	0.2286	0.3423	0.1775	0.2195	0.3225	0.2195	0.2323
0.0	0.8	25	10	0.0583	0.1517	0.0773	0.0517	0.2796	0.0383	0.0658	0.0969	0.0519	0.1123	0.0413	0.0517	0.1077	0.0517	0.0617
0.1				0.0638	0.1602	0.0838	0.0543	0.2855	0.0422	0.0700	0.1007	0.0543	0.1169	0.0436	0.0543	0.1107	0.0543	0.0640
0.6				0.2409	0.4224	0.3241	0.1522	0.4473	0.1494	0.2088	0.2464	0.1573	0.2555	0.1303	0.1522	0.2408	0.1522	0.1606
1.0				0.5307	0.7273	0.6645	0.3251	0.6597	0.3543	0.4415	0.4711	0.3422	0.4697	0.2935	0.3251	0.4445	0.3251	0.3309
0.0	0.8	50	10	0.0395	0.0796	0.0735	0.0522	0.2747	0.0421	0.0556	0.0913	0.0524	0.1089	0.0464	0.0522	0.1041	0.0522	0.0567
0.1				0.0495	0.0956	0.0877	0.0563	0.2875	0.0484	0.0631	0.0991	0.0571	0.1161	0.0510	0.0563	0.1113	0.0563	0.0617
0.6				0.4432	0.5694	0.5462	0.2503	0.5778	0.2899	0.3319	0.3780	0.2634	0.3830	0.2364	0.2503	0.3612	0.2503	0.2541
1.0				0.8701	0.9280	0.9143	0.5628	0.8487	0.6602	0.7030	0.7262	0.5872	0.7130	0.5448	0.5628	0.6834	0.5628	0.5646
0.0	0.8	100	10	0.0392	0.0590	0.0701	0.0509	0.2725	0.0456	0.0529	0.0888	0.0501	0.1052	0.0481	0.0509	0.1018	0.0509	0.0531
0.1				0.0631	0.0889	0.0992	0.0609	0.2933	0.0601	0.0685	0.1045	0.0616	0.1213	0.0578	0.0609	0.1162	0.0609	0.0631
0.6				0.7785	0.8279	0.8203	0.4446	0.7658	0.5414	0.5653	0.6073	0.4661	0.5986	0.4349	0.4446	0.5703	0.4446	0.4452
1.0				0.9953	0.9974	0.9969	0.8530	0.9715	0.9307	0.9385	0.9420	0.8735	0.9303	0.8475	0.8530	0.9158	0.8530	0.8545

Table 5: Results for the Levels Specification - Non-Gaussian Innovation Distributions

gamma	rho	n	T	GLS-SC	GLS	GLS-SC	GLS	GLS-SC	GLS	GLS-SC	GLS	GLS-SC	GLS	GLS-SC	GLS
				Full Sample				3 Periods				2 Periods			
				t-dist		**Chi-Sq**		**t-dist**		**Chi-Sq**		**t-dist**		**Chi-Sq**	
0.0	0.8	50	5	0.0365	0.0536	0.0459	0.0651	0.0376	0.0513	0.0473	0.0625	0.0423	0.0470	0.0494	0.0544
0.1				0.0416	0.0606	0.0480	0.0680	0.0425	0.0560	0.0490	0.0635	0.0458	0.0510	0.0498	0.0550
0.6				0.2677	0.3221	0.1082	0.1391	0.2385	0.2764	0.1011	0.1242	0.2164	0.2299	0.0956	0.1042
1.0				0.6181	0.6745	0.2216	0.2689	0.5552	0.6005	0.2000	0.2337	0.4998	0.5178	0.1871	0.2003
0.0	0.8	50	10	0.0404	0.0825	0.0466	0.0920	0.0423	0.0557	0.0453	0.0596	0.0465	0.0517	0.0488	0.0541
0.1				0.0459	0.0918	0.0485	0.0938	0.0457	0.0602	0.0467	0.0604	0.0501	0.0561	0.0494	0.0550
0.6				0.2744	0.3866	0.1068	0.1762	0.1809	0.2152	0.0819	0.1022	0.1662	0.1783	0.0823	0.0892
1.0				0.6255	0.7365	0.2163	0.3135	0.4232	0.4707	0.1516	0.1820	0.3804	0.3978	0.1470	0.1584
0.0	0.8	50	15	0.0493	0.1216	0.0552	0.1336	0.0440	0.0592	0.0489	0.0645	0.0472	0.0530	0.0490	0.0552
0.1				0.0546	0.1321	0.0572	0.1370	0.0465	0.0618	0.0501	0.0650	0.0498	0.0552	0.0502	0.0552
0.6				0.2832	0.4443	0.1156	0.2232	0.1652	0.1965	0.0817	0.1024	0.1560	0.1677	0.0801	0.0872
1.0				0.6293	0.7733	0.2237	0.3667	0.3813	0.4283	0.1420	0.1720	0.3539	0.3710	0.1367	0.1478
0.0	0.8	50	20	0.0626	0.1737	0.0660	0.1790	0.0461	0.0616	0.0497	0.0648	0.0482	0.0534	0.0503	0.0553
0.1				0.0690	0.1845	0.0682	0.1830	0.0490	0.0639	0.0510	0.0649	0.0511	0.0566	0.0512	0.0568
0.6				0.3082	0.5048	0.1303	0.2765	0.1672	0.1976	0.0820	0.1026	0.1599	0.1719	0.0797	0.0878
1.0				0.6445	0.8112	0.2414	0.4257	0.3874	0.4318	0.1414	0.1702	0.3603	0.3777	0.1362	0.1462

Table 6: 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	

Table 7: Correlation Structure

1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998
1.00										
0.36	1.00									
0.01	0.33	1.00								
-0.21	0.32	0.61	1.00							
0.05	-0.01	0.03	0.14	1.00						
-0.18	-0.27	-0.14	-0.17	-0.07	1.00					
-0.24	-0.22	-0.13	-0.14	-0.14	0.46	1.00				
-0.28	-0.25	-0.16	-0.13	-0.12	0.42	0.45	1.00			
-0.47	-0.24	-0.22	-0.18	-0.09	-0.08	-0.06	-0.02	1.00		
-0.53	-0.36	-0.29	-0.09	-0.06	-0.07	-0.06	0.01	0.45	1.00	
-0.62	-0.29	-0.03	-0.04	-0.35	-0.10	-0.08	-0.02	0.12	0.24	1.00

Table Legends

Table 1a:

The entries in the table are estimated rejection probabilities for a one sided test of the hypothesis that gamma=0 in Model (4.2) at the nominal significance level of 5%. The test statistics used are for the level version defined in Theorem 3.1. Rows corresponding to gamma=0 refer to rejection probabilities under the null hypothesis. Rows corresponding to gamma>0 refer to rejection probabilities under the corresponding alternative. The column labeled GLS-SC reports results for Wald statistics based on the feasible GLS estimator with the size correction given in Equation (3.2). The column GLS contains results based on the feasible GLS estimator without size correction. GLS-AR denotes the test statistic based on parametric covariance matrix estimators described in Section 4. ROLS is the Wald statistic based on the standard OLS estimator with robust standard errors and the covariance matrix is computed in the same way as for GLS. OLS is the usual OLS estimator with the non-robust variance estimator. Results for 3 Periods and 2 Periods are based on time averaging the sample prior to carrying out the tests over three or two time periods. The rejection probabilities are based on 50,000 replications. The cross-sectional sample size is kept fixed at n=50 for all cases.

Table 1b:

The entries in the table are estimated rejection probabilities for a one sided test of the hypothesis that gamma=0 in Model (4.2) at the nominal significance level of 5% when rho=.4. The test statistics used are for the level version defined in Theorem 3.1. See Table 1a for further details.

Table 1c:

The entries in the table are estimated rejection probabilities for a one sided test of the hypothesis that gamma=0 in Model (4.2) at the nominal significance level of 5% when rho=.8. The test statistics used are for the level version defined in Theorem 3.1. See Table 1a for further details.

Table 1d:

The entries in the table are estimated rejection probabilities for a one sided test of the hypothesis that gamma=0 in Model (4.2) at the nominal significance level of 5% when rho=.9. The test statistics used are for the level version defined in Theorem 3.1. See Table 1a for further details.

Table 2a:

The entries in the table are estimated rejection probabilities for a one sided test of the hypothesis that gamma=0 in Model (4.2) at the nominal significance level of 5% when rho=0. The test statistics used are for the first difference version defined in (3.6) and (3.7). See Table 1a for further details.

Table 2b:

The entries in the table are estimated rejection probabilities for a one sided test of the hypothesis that gamma=0 in Model (4.2) at the nominal significance level of 5% when rho=.4. The test statistics used are for the first difference version defined in (3.6) and (3.7). See Table 1a for further details.

Table 2c:

The entries in the table are estimated rejection probabilities for a one sided test of the hypothesis that gamma=0 in Model (4.2) at the nominal significance level of 5% when rho=.8. The test statistics used are for the first difference version defined in (3.6) and (3.7). See Table 1a for further details.

Table 2d:

The entries in the table are estimated rejection probabilities for a one sided test of the hypothesis that gamma=0 in Model (4.2) at the nominal significance level of 5% when rho=.9. The test statistics used are for the first difference version defined in (3.6) and (3.7). See Table 1a for further details.

Table 3:

The entries in the table are estimated rejection probabilities for a one sided test of the hypothesis that gamma=0 in Model (4.2) at the nominal significance level of 5%. The innovations of (4.2) are drawn from a multivariate Gaussian distribution with a correlation matrix corresponding to the one reported in Table 7. Rows corresponding to gamma=0 refer to rejection probabilities under the null hypothesis. Rows corresponding to gamma>0 refer to rejection probabilities under the corresponding alternative. The column labeled GLS is the feasible GLS estimator without size correction, GLS-SC reports results for Wald statistics based on the feasible GLS estimator with the size correction given in Equation (3.2). GLS-B is the t-statistic based on GLS with critical values computed by the bootstrap as described in Section 4. GLS-AR denotes the test statistic based on parametric covariance matrix estimators described in Section 4. Results for 3 Periods and 2 Periods are based on time averaging the sample over three or two time periods prior to carrying out the tests. The rejection probabilities are based on 50,000 replications for GLS, OLS-SC and GLS-AR. The rejection probabilities for GLS-B are based on 5,000 replications with 200 repetitions of the bootstrap at each replication to compute critical errors. The sample size is n=50 and T=11.

Table 4:

The entries in the table are estimated rejection probabilities for a one sided test of the hypothesis that gamma=0 in Model (4.2) at the nominal significance level of 5% when rho=.8. The test statistics used are for the level version defined in Theorem 3.1. The Monte Carlo design in this table is the same as in Table 1c except for the fact that *T=10* is kept fixed while *n* varies from 15 to 100.

Table 5:

The entries in the table are estimated rejection probabilities for a one sided test of the hypothesis that gamma=0 in Model (4.2) at the nominal significance level of 5% when rho=.8. The test statistics used are for the level version defined in Theorem 3.1. The Monte Carlo design in this table is the same as in Table 1c except for the fact that the innovations in (4.1) are drawn from a t distribution with 4 degrees of freedom and a demeaned chi-square distribution with 4 degrees of freedom respectively.

Table 7:

Serial correlations between regression residuals from OLS estimation of the cellular telephone data using Equation (1.1) with a fixed effect for each MSA, a time effect for each year and a single indicator variable to allow for the effect of price regulation.