ANNALES D'ÉCONOMIE ET DE STATISTIQUE. – N° 14 — 1989

Further Monte Carlo Evidence on Seemingly Unrelated Regressions with Unequal Number of Observations

Badi H. BALTAGI, Susan GARVIN, Stephen KERMAN *

ABSTRACT. – SCHMIDT'S [1977] results on seemingly unrelated regressions with unequal number of observations are replicated. These results are shown to be robust to the type of additional observations available *i. e.,* whether they are time-series or cross-sectional in nature. An important finding is that the extra observations may lead to better estimates of the variance-covariance matrix Σ or its inverse Σ^{-1}, but this does not necessarily lead to better estimates of the regression coefficients.

Résultats de simulations portant sur des régressions empilées avec des nombres d'observations différents

RÉSUMÉ. – On reproduit les résultats de SCHMIDT [1977] sur les régressions empilées dans le cas où il y a des nombres d'observations différents. On montre que ces résultats restent robustes quel que soit le type d'observation qu'on ajoute, c'est-à-dire qu'elles soient des données temporelles ou en coupe. Les observations supplémentaires peuvent conduire à de meilleures estimations de la matrice Σ de variances- covariances ou de son inverse Σ^{-1}, mais elles n'améliorent pas nécessairement l'estimation des coefficients de régression.

* B. H. BALTAGI : Texas A & M University, Department of Economics, College Station Texas 77843-4228, USA; S. GARVIN et S. KERMAN : University of Houston, Department of Economics, Houston, Texas 77004, USA. The authors would like to thank Anurag Malhotra for his active participation on an earlier draft of this paper, and two anonymous referees whose comments and suggestions improved the paper.

1 Introduction

Schmidt [1977] considers the problem of estimating a set of two seemingly unrelated regressions with unequal number of observations, and provides five different estimators of the disturbances' variance-covariance matrix Σ. All five estimators of Σ are consistent and the corresponding coefficient estimates are asymptotically efficient. In order to distinguish among the small sample performances of these estimators, Monte Carlo experiments are performed. One surprising result reported by Schmidt [1977] is that a feasible GLS estimator of the regression coefficients that ignores the extra observations in estimating Σ compares favorably to a feasible GLS estimator of the regression coefficients that uses all the extra observations in estimating Σ.

The objectives of this paper are two-fold: The first is to replicate, as closely as possible, the Monte Carlo experiments performed by Schmidt [1977], providing additional support or counter-evidence to his results. The second is to try to explain Schmidt's surprising findings by pursuing the following strategies: (1) We focus on the performance of the variance-covariance estimates themselves. Our conjecture is that better estimates of these variances need not imply better estimates of the regression coefficients; [1] (2) We consider a reparameterization by Hwang [1987] of this estimation problem in terms of the elements of Σ^{-1} rather than Σ. This reparameterization allows us to check whether better estimates of Σ^{-1} yield better estimates of the regression coefficients; (3) Finally, we focus on the type of additional observations, *i.e.*, whether it is time-series or cross-sectional in nature, to see whether this affects the results.

In the next section, we review Schmidt's [1977] model and the estimation methods to be compared. Section 3 describes the design of the experiment, Section 4 compares the various estimators of the variance-covariance matrix Σ and Section 5 compares the various estimators of the regression coefficients. Section 6 discusses the reparameterization in terms of Σ^{-1}, Section 7 discusses the results of varying the type of additional information and the last section provides the conclusion.

1. A similar result has been obtained in panel data studies by Maddala and Mount [1973], Taylor [1980] and Baltagi [1981].

2 The Model

SCHMIDT [1977] considers a set of two ZELLNER [1962] seemingly unrelated regressions

$$y_1 = X_1 \beta_1 + \varepsilon_1 \tag{1}$$

$$y_2 = X_2 \beta_2 + \varepsilon_2 \tag{2}$$

with unequal number of observations. In particular, there are T observations on the first equation and (T + E) observations on the second equation. The T observations are overlapping. Assuming that $(\varepsilon_{t1}, \varepsilon_{t2})$ are i. i. d. $N(0, \varepsilon)$, we have

$$\Omega = \text{cov}(\varepsilon) = \begin{bmatrix} \sigma_{11} I_T & \sigma_{12} I_T & 0 \\ \sigma_{21} I_T & \sigma_{22} I_T & 0 \\ 0 & 0 & \sigma_{22} I_E \end{bmatrix} \tag{3}$$

where $\Sigma = [\sigma_{ij}]$ and $\varepsilon' = (\varepsilon_1', \varepsilon_2')$. Partition the observations on the second equation as follows:

$$X_2 = \begin{pmatrix} X_2^* \\ X_2^0 \end{pmatrix}, \qquad y_2 = \begin{pmatrix} y_2^* \\ y_2^0 \end{pmatrix} \tag{4}$$

where X_2^* and y_2^* contain the T overlapping observations and X_2^0, y_2^0 denote the remaining E observations. The GLS estimator of $\beta' = (\beta_1', \beta_2')$ can be written as follows:

$$\hat{\beta}_{GLS} = \begin{bmatrix} \sigma^{11} X_1' X_1 & \sigma^{12} X_1' X_2^* \\ \sigma^{21} X_2^{*\prime} X_1 & \sigma^{22} X_2^{*\prime} X_2^* + (1/\sigma_{22}) X_2^{0\prime} X_2^0 \end{bmatrix}^{-1} \begin{bmatrix} \sigma^{11} X_1' y_1 + \sigma^{12} X_1' y_2^* \\ \sigma^{21} X_2^{*\prime} y_1 + \sigma^{22} X_2^{*\prime} y_2^* + (1/\sigma_{22}) X_2^{0\prime} y_2^0 \end{bmatrix} \tag{5}$$

Let e_1 and e_2 be the vectors of OLS residuals for the first and second equation, with dimensions T and (T + E), respectively. Partition e_2' as $(e_2^{*\prime}, e_2^{0\prime})$ where e_2^* is $T \times 1$ and e_2^0 is $E \times 1$, and define the following quantities:

$$\begin{cases} s_{11} = e_1' e_1/T, \quad s_{12} = e_1' e_2^*/T, \quad s_{22}^* = e_2^{*\prime} e_2^*/T \\ s_{22}^0 = e_2^{0\prime} e_2^0/E, \quad s_{22} = e_2' e_2/(T+E). \end{cases} \tag{6}$$

Based on (6), the following estimators of Σ are considered by SCHMIDT [1977]:

(i) The "usual" estimator, which uses no extra observations in estimating Σ, *i. e.*, it relies on the overlapping residuals e_1 and e_2^*:

$$\hat{\sigma}_{11,\ \text{usual}} = s_{11}, \qquad \hat{\sigma}_{22,\ \text{usual}} = s_{22}^*, \qquad \hat{\sigma}_{12,\ \text{usual}} = s_{12}. \tag{7}$$

(ii) The WILKS' [1932] estimator which uses extra observations only in estimating σ_{22}:

$$\hat{\sigma}_{11,w}=s_{11}, \quad \hat{\sigma}_{22,w}=s_{22}, \quad \hat{\sigma}_{12,w}=s_{12}. \tag{8}$$

(iii) The SRIVASTAVA and ZAATAR's [1973] modification of Wilks which uses the extra observations in estimating σ_{22} and σ_{12}:

$$\hat{\sigma}_{11,sz}=s_{11}, \quad \hat{\sigma}_{22,sz}=s_{22}, \quad \hat{\sigma}_{12,sz}=s_{12}\,(s_{22}/s_{22}^{*})^{1/2}. \tag{9}$$

(iv) The HOCKING and SMITH'S [1968] estimator which uses the extra observations in estimating all elements of Σ:

$$\begin{cases} \hat{\sigma}_{11,HS}=s_{11}-(E/T+E)\,(s_{12}/s_{22}^{*})^{2}\,(s_{22}^{*}-s_{22}^{0}) \\ \hat{\sigma}_{22,HS}=s_{22}, \quad \hat{\sigma}_{12,HS}=s_{12}\,(s_{22}/s_{22}^{*}) \end{cases} \tag{10}$$

It is important to point out that the extra observations are used in obtaining the least squares residuals, but that some estimators like the "usual" estimator truncate the additional residuals e_2^0 in estimating Σ. Also, as clear from (5), any feasible GLS estimator of β ultimately uses the extra observations, even though this estimator may seem not to do so in the estimation of Σ.

Finally, we would like to point out that SCHMIDT [1977] uses the maximum likelihood estimator as basis for comparison. We do not consider MLE in this paper, rather we focus on true GLS for comparison purposes. True GLS means GLS knowing the true Σ. We believe that this latter comparison gives us a better feel of how a particular feasible GLS estimator compares with the same estimator which uses the true σ's.

3 The Monte Carlo Design

Following SCHMIDT [1977] and KMENTA and GILBERT [1968], the following two equation model was considered:

$$y_1=10+2X_{11}-5X_{12}+\varepsilon_1 \tag{11}$$

$$y_2=-10+6X_{21}+3X_{22}+\varepsilon_2 \tag{12}$$

Note that this Monte Carlo set up is invariant to the use of these specific regression parameters, in other words, there is no need to vary the intercepts and slopes of (11) and (12) since the Monte Carlo results are invariant to their values, see BREUSCH [1980, p. 337]. This study uses only the first set of X's used by KMENTA and GILBERT [1968, p. 1186] for their experiments, *i. e.*, the case where the cross-equation correlation between the X's is low. Alternative sets of explanatory variables are considered using additional observations that are time-series or cross-sectional in origin, to assess whether the type of additional information affects the properties of the estimates. These are discussed in more detail in Section 7.

Following Schmidt, we set the variances of ε_1 and ε_2 equal to one ($\sigma_{11}=\sigma_{22}=1$) and we consider three alternative values of the correlation between ε_1 and ε_2: namely $\sigma_{12}=\rho=.3, .6, .925$. Three different values of the extra observations are used: E=5, 10 and 20. All the extra observations are on the second equation. Also, three different sample sizes are considered: T=10, 20 and 50. For our study all possible combinations of T, E and ρ are entertained.

For each experiment, (X matrix, value of ρ, value of T and value of E), a sample was generated using a pseudo-random normal deviate generator and the four feasible GLS estimators described in the previous section along with true GLS and OLS are performed. Each experiment is replicated 500 times and the MSE's are obtained for the σ's and the regression coefficients. Also, a count measure is obtained which gives the number of times an estimator is closer to the true value of the parameter than another estimator, and whether this frequency count is significantly different from 50%. We now turn to the results of the experiments.

4 A Comparison of Various Estimators of Σ

Table 1 reports the MSE of various estimators of Σ. [2] The following observations can be drawn from our results:

(1) Only HS has a different estimator of σ_{11}, and in only 8 cases is the MSE of $\hat{\sigma}_{11,\ HS}$ worse than that of $\hat{\sigma}_{11,\ usual}$. A count of the number of times that $\hat{\sigma}_{11,\ HS}$ was closer to the true σ_{11} than $\hat{\sigma}_{11,\ usual}$ reveals one case in favor of "usual", 7 cases in favor of "HS" and 19 cases where one cannot reject indifference at the 95% confidence level. [3] (2) Only the "usual" estimator has a different estimator of σ_{22}, and in all cases the MSE of $\hat{\sigma}_{22,\ usual}$ is worse than that of $\hat{\sigma}_{22}$ for all other feasible GLS estimators. In fact there are 22 cases where $\hat{\sigma}_{22,\ HS}$ is closer to the true σ_{22} than $\hat{\sigma}_{22,\ usual}$ and 5 cases of indifference at the 95% confidence level. (3) For σ_{12}, both SZ and HS have different estimators from that of the

2. For the Wilks' estimator, $\hat{\Sigma}$ is not necessarily positive definite. SCHMIDT [1977] reported the same problem in his study. The worst cases occurred for T=10 and E=20: In fact, for $\rho=.3$, $\hat{\Sigma}$ was negative definite in only 1 replication out of 500. For $\rho=.6$, $\hat{\Sigma}$ was negative definite in 8 cases out of 500 and for $\rho=.925$, in 62 cases out of 500. For this reason, we focus on the other feasible GLS estimators in the remainder of this study.
3. Using 500 replications, counts greater than 272 or smaller than 228 are significantly different from 250 at the .05 level.

TABLE 1

M.S.E. of the Variance Estimates.

	T	E	ρ	$\hat{\sigma}_{11}$ Usual (SZ)	$\hat{\sigma}_{11}$ HS	$\hat{\sigma}_{12}$ Usual	$\hat{\sigma}_{12}$ SZ	$\hat{\sigma}_{12}$ HS	$\hat{\sigma}_{22}$ Usual	$\hat{\sigma}_{22}$ SZ(HS)
1	10	5	0.3	0.2190	0.2206	0.0753	0.0744	0.0785	0.1939	0.1523
2	10	10	0.3	0.2230	0.2304	0.0725	0.0696	0.0753	0.1950	0.1113
3	10	20	0.3	0.2164	0.2169	0.0752	0.0738	0.0812	0.1569	0.0734
4	20	5	0.3	0.1095	0.1093	0.0505	0.0504	0.0515	0.0966	0.0821
5	20	10	0.3	0.0972	0.0970	0.0456	0.0440	0.0443	0.0915	0.0734
6	20	20	0.3	0.1080	0.1069	0.0523	0.0483	0.0477	0.0976	0.0524
7	50	5	0.3	0.0439	0.0437	0.0200	0.0198	0.0198	0.0421	0.0385
8	50	10	0.3	0.0431	0.0432	0.0220	0.0215	0.0212	0.0392	0.0337
9	50	20	0.3	0.0441	0.0443	0.0212	0.0207	0.0207	0.0365	0.0260
10	10	5	0.6	0.2199	0.2248	0.1297	0.1248	0.1262	0.1845	0.1483
11	10	10	0.6	0.2173	0.2189	0.1214	0.1122	0.1128	0.1908	0.1084
12	10	20	0.6	0.2108	0.2117	0.1198	0.1104	0.1113	0.1562	0.0747
13	20	5	0.6	0.1105	0.1094	0.0744	0.0718	0.0713	0.0989	0.0843
14	20	10	0.6	0.0973	0.0939	0.0662	0.0605	0.0587	0.0948	0.0747
15	20	20	0.6	0.1119	0.1066	0.0751	0.0639	0.0601	0.1020	0.0544
16	50	5	0.6	0.0428	0.0421	0.0268	0.0261	0.0259	0.0411	0.0378
17	50	10	0.6	0.0442	0.0435	0.0295	0.0278	0.0269	0.0402	0.0339
18	50	20	0.6	0.0436	0.0431	0.0283	0.0257	0.0247	0.0375	0.0268
19	10	5	0.925	0.2200	0.2120	0.2311	0.2200	0.2159	0.1702	0.1409
20	10	10	0.925	0.2120	0.1898	0.2146	0.1941	0.1860	0.1757	0.1030
21	10	20	0.925	0.2087	0.1837	0.2063	0.1825	0.1714	0.1561	0.0739
22	20	5	0.925	0.1111	0.1021	0.1168	0.1093	0.1054	0.1023	0.0873
23	20	10	0.925	0.0999	0.0833	0.1043	0.0912	0.0850	0.0966	0.0739
24	20	20	0.925	0.1174	0.0884	0.1150	0.0915	0.0812	0.1074	0.0571
25	50	5	0.925	0.0416	0.0389	0.0401	0.0382	0.0375	0.0405	0.0373
26	50	10	0.925	0.0452	0.0404	0.0432	0.0392	0.0371	0.0426	0.0352
27	50	20	0.925	0.0425	0.0354	0.0413	0.0348	0.0317	0.0395	0.0285

"usual" estimator. The MSE of $\hat{\sigma}_{12,\,SZ}$ is always smaller than $\hat{\sigma}_{12,\,usual}$. The same is true for $\hat{\sigma}_{12,\,HS}$ except in 4 cases, all of which have $\rho = .3$. The count reveals 14 cases where $\hat{\sigma}_{12,\,SZ}$ is better than $\hat{\sigma}_{12,\,usual}$ and 13 cases of indifference at the 95% confidence level. Also, there are 8 cases where $\hat{\sigma}_{12,\,HS}$ is better than $\hat{\sigma}_{12,\,usual}$ and 19 cases of indifference.

To summarize, this evidence is in favor of the proposition that estimators of the variances which use the extra observations have better MSE and count performance than those procedures that do not use the extra observations fully. In fact, only HS uses extra observations for estimating σ_{11}, and except for one case (using the count measure) and 8 cases (using the MSE measure) its performance is better than that of other estimators that do not use the extra observations. Only the "usual" estimator does not use the extra observations in estimating σ_{22} and in all cases it has the worst MSE performance. For σ_{12}, both SZ and HS use the extra observations, and in all cases these methods perform better or no different than the usual estimator which does not use the extra observations. The next question to pose is the following: Does this better performance for the variance estimates translate into better performance for the coefficient estimates? This is answered in the next section.

5 A Comparison of Various Estimators of β

Due to space limitations, and following SCHMIDT [1977] we focus on β_{11} (the coefficient of X_{11} in the first equation). Table 2 gives the number of times a specific estimator of β_{11} is closer to β_{11} than the true GLS estimator. The following observations can be drawn from our results:

(i) As expected, OLS is dominated by true GLS for all cases where $\rho > .3$ and only when the sample size is large (T=20 or 50) and ρ is small (.3) does it fair well with respect to true GLS.

TABLE 2

Number of Times in 500 Trials that an Estimator of β_{11} ***Beats the GLS Estimator.***

	T	E	ρ	Usual	SZ	HS	OLS
1.	10	5	0.3	221	220	222	215
2.	10	10	0.3	214	206	210	210
3.	10	20	0.3	229	229	236	222
4.	20	5	0.3	208	207	209	207
5.	20	10	0.3	258	256	260	236
6.	20	20	0.3	228	225	226	221
7.	50	5	0.3	220	219	220	220
8.	50	10	0.3	255	250	256	236
9.	50	20	0.3	237	234	237	238
10.	10	5	0.6	226	234	228	182
11.	10	10	0.6	209	209	211	181
12.	10	20	0.6	216	208	216	194
13.	20	5	0.6	221	223	223	176
14.	20	10	0.6	252	244	256	211
15.	20	20	0.6	202	206	201	198
16.	50	5	0.6	241	239	236	191
17.	50	10	0.6	246	254	247	200
18.	50	20	0.6	225	242	231	204
19.	10	5	0.925	190	196	191	111
20.	10	10	0.925	195	194	202	112
21.	10	20	0.925	185	189	186	126
22.	20	5	0.925	214	219	213	110
23.	20	10	0.925	207	210	211	118
24.	20	20	0.925	204	185	201	122
25.	50	5	0.925	260	239	257	115
26.	50	10	0.925	240	248	238	125
27.	50	20	0.925	225	228	225	131

Note: Counts greater than 272 or smaller than 228 are significantly different from 250 at the .05 level.

(ii) True GLS dominates *all* other feasible GLS estimators in 14 cases out of 27 at the 95% confidence level. It is no different from *all* other feasible GLS estimators in 9 cases. These are cases where $T \geqslant 20$. The remaining 4 cases have mixed results with some feasible GLS estimators no different from true GLS and other feasible GLS estimators dominated by true GLS. These results are important for reference purpose because they point out the specific cases where estimating Σ rather than using true Σ did (or did not) matter for the estimation of β_{11}. For example, it does not make sense to compare two feasible GLS estimators in cases where both are no different from true GLS.

TABLE 3

M.S.E. of $\hat{\beta}_{11}$ ***Relative to the M.S.E. of GLS***

	T	E	ρ	Usual	SZ	HS	OLS
1.	10	5	0.3	1.1297	1.1352	1.1282	1.0787
2.	10	10	0.3	1.1435	1.1415	1.1413	1.1175
3.	10	20	0.3	1.0899	1.0917	1.0898	1.1059
4.	20	5	0.3	1.0764	1.0723	1.0765	1.1419
5.	20	10	0.3	1.0327	1.0321	1.0324	1.0964
6.	20	20	0.3	1.0474	1.0442	1.0475	1.1202
7.	50	5	0.3	1.0275	1.0292	1.0276	1.1118
8.	50	10	0.3	1.0100	1.0096	1.0097	1.0969
9.	50	20	0.3	1.0166	1.0156	1.0165	1.0830
10.	10	5	0.6	1.2034	1.2096	1.2004	1.4594
11.	10	10	0.6	1.2329	1.2326	1.2303	1.5752
12.	10	20	0.6	1.1925	1.1850	1.1908	1.5500
13.	20	5	0.6	1.0897	1.0848	1.0393	1.6237
14.	20	10	0.6	1.0592	1.0623	1.0587	1.5150
15.	20	20	0.6	1.0883	1.0918	1.0878	1.5779
16.	50	5	0.6	1.0299	1.0347	1.0306	1.5377
17.	50	10	0.6	1.0314	1.0311	1.0308	1.5016
18.	50	20	0.6	1.0295	1.0274	1.0292	1.4760
19.	10	5	0.925	2.1539	2.1494	2.1473	5.5905
20.	10	10	0.925	2.2804	2.2627	2.2683	6.1846
21.	10	20	0.925	2.2185	2.1816	2.2519	6.1106
22.	20	5	0.925	1.3688	1.3748	1.3640	6.5022
23.	20	10	0.925	1.3655	1.4026	1.3644	6.0284
24.	20	20	0.925	1.3995	1.4331	1.3912	6.2458
25.	50	5	0.925	1.0713	1.0882	1.0728	6.0229
26.	50	10	0.925	1.1296	1.1324	1.1275	5.9558
27.	50	20	0.925	1.1050	1.1143	1.1024	5.7339

Table 3 gives the mean squared error of various estimators of β_{11} relative to that of true GLS. This clearly demonstrates the magnitude of the gain in using true GLS rather than feasible GLS. This gain is maximum when ρ is large (.925) and T is small (10). Table 3 also shows how OLS deteriorates in relative MSE performance as ρ increases, and how feasible GLS improves in relative MSE performance as T increases. Focusing on the feasible GLS estimators, there is not much difference in relative MSE performance across these estimators. In fact, a simple counting of the number of times each feasible GLS estimator beats the "usual" estimator confirms this result. At the 95% confidence level, we can not reject the hypothesis that either the Hocking-Smith (HS) or the Srivastava and Zaatar (SZ) are any different from the "usual" estimator. This replicates the surprising result reported by SCHMIDT [1977, p. 376] : "*...It is certainly remarkable that procedures that essentially ignore the extra observations in estimating* Σ [e. g. the usual estimator] *do not generally do badly relative to procedures that use the extra observations fully* [e. g. HS estimator]."

In order to shed some light on this result, we focus on a specific case, that for which $T=10$, $E=5$ and $\rho=.925$, a case where true GLS dominates all feasible GLS estimators (*i. e.* a case where knowing the true variances makes a difference) HS and SZ estimators of the variances strictly dominate those of the usual estimator in terms of MSE, and this translates into the smaller MSE for the corresponding estimators of the coefficients. By the count measure, however, we cannot reject that the performance of the coefficient estimators are the same as that of the "usual" estimator.

Let us consider another case: For $T=20$, $E=10$ and $\rho=.925$, HS and SZ estimators of the variance strictly dominate those of the usual estimator in terms of MSE, but this translates into a larger MSE for the regression coefficients of SZ and a smaller MSE for HS. Using the count measure, the variance estimates dominate for HS and SZ over the usual but we cannot reject the fact that the regression coefficient performances are no different. This is an example of cases where strict dominance in MSE of the variances translated in the case of SZ into worse performance of the MSE of β_{11} and in the case of HS into a better performance for β_{11}. The moral of the story is that better estimates of the variances need not imply better estimates of the regression coefficients.

Let us now focus on what happens to a specific estimator of β_{11} as we fix T and ρ and increase E (the number of extra observations). The MSE for $\hat{\beta}_{11}$ does not necessarily fall as E increases from 5 to 10 or 20. This is another duplication of the result obtained by SCHMIDT [1977, p. 373]. If we focus on the variance estimates, say the MSE of σ_{22}, the SZ and the HS estimator are identical, and both improve with extra observations. However, the MSE of σ_{11} and the MSE of σ_{12} of all feasible GLS estimators do not necessarily fall as E increases with T and ρ fixed.

6 A Reparameterization in Terms of Σ^{-1}

Recently, HWANG [1987] reparameterized this estimation problem in terms of the elements of Σ^{-1} rather than Σ. In particular, Σ^{-1} depends upon $\delta_1=\sigma_{11}-\sigma_{12}^2/\sigma_{22}$, $\theta_1=\sigma_{12}/\sigma_{22}$ and σ_{22}, and one would estimate $(\delta_1, \theta_1, \sigma_{22})$ rather than $(\sigma_{11}, \sigma_{12}, \sigma_{22})$. Or equivalently, one could reparameterize in terms of $\delta_2=\sigma_{22}-\sigma_{21}^2/\sigma_{11}$, $\theta_2=\sigma_{21}/\sigma_{11}$ and σ_{11}. With these reparameterizations one can easily show that HS, SZ and the "usual" estimators differ only in estimating one of these parameters. In particular HS differs from the "usual" estimator only in its estimate of θ_2. Similarly, HS differs from SZ only in its estimate of θ_1. HWANG [1987] argues that this reparameterization helps explain why the HS estimator, for example, does not necessarily use more observations in the estimation of Σ^{-1} than the other estimators. For example, HWANG [1987] shows that the HS estimator of β can be derived using $\hat{\delta}_1 = s_{11} - s_{12}^2/s_{22}^*$, $\hat{\theta}_1 = s_{12}/s_{22}^*$, and $\hat{\sigma}_{22} = s_{22}$. This shows that the HS estimators of δ_1 and θ_1 rely solely on the T common observations whereas the HS estimator of $\hat{\sigma}_{22}$ uses the entire (T+E) observations. This is different from the Σ parameterization given in (10), where HS was shown to use the extra observations in estimating *all* the elements of Σ.

We will use this reparaterization to focus the comparison of the various regression coefficient estimators according to the performance of the corresponding estimate of θ. In particular, when comparing HS vs. "usual", we focus on their corresponding estimates of θ_2. Even though there are 8 cases where $\hat{\theta}_{2,\,HS}$ dominates $\hat{\theta}_{2,\,usual}$ we cannot reject that the performance of the corresponding estimates of β_{11} are no different. In terms of MSE performance, we can find cases where better estimates of θ_2 translate into better *and* worse estimates of β_{11}. Once again we have the result that a better estimate of a certain crucial parameter of Σ^{-1} (that differentiates between two feasible GLS estimators) does not necessarily lead to a better estimate of the corresponding regression coefficients.

Using $\hat{\theta}_1$ one can also look at whether HS is better than SZ. Our results indicate that for larger ρ and larger T, the MSE performance of $\hat{\theta}_{1,\,HS}$ is better than that of $\hat{\theta}_{1,\,SZ}$. There are 11 cases where one cannot reject that $\hat{\theta}_{1,\,HS}$ dominates $\hat{\theta}_{1,\,SZ}$ using the count measure and in the remaining 16 cases one cannot reject indifference at the 95% confidence level. Only in one case, however, does this dominance in $\hat{\theta}_1$ translate into dominance of $\hat{\beta}_{11,\,HS}$ over $\hat{\beta}_{11,\,SZ}$ in terms of the count measure (this case is for T=20, E=20 and $\rho = .925$). Once again better estimates of a crucial parameter in Σ^{-1} need not lead to better estimates of the regression coefficients. Except for one case, we cannot reject the fact that the performance of the HS and SZ estimates of β_{11} are identical.

7 Varying the Type of Additional Information

So far, we have focused on how the extra observations affect the estimates of the variances and how they in turn affect the regression coefficient estimates. Next, we focus on the type of additional information that is available in the second equation, *i.e.* on the choice of X and the additional observations. Rather than replicating the X's used by KMENTA and GILBERT [1968] to generate the additional observations for the second equation, see SCHMIDT [1977], we turn to economic data for the source of our observations.

TABLE 4

Varying the Type of Additional Observations

I. Cross-Sctional Data [1975 and 1976 Cigarette Consumption Data across 46 U.S. States, *see* Baltagi and Levin (1986)].

			Number of Times in 100 Trials That an Estimator of β_{11} Beats the USUAL Estimator			
T	E	ρ	GLS	SZ	HS	OLS
10	5	.3	57	54	50	53
10	10	.3	55	55	62	44
10	5	.6	55	61	48	34
10	10	.6	60	51	62	36
10	5	.925	58	50	46	27
10	10	.925	57	50	49	25

II. Time-Series Data [General Electric and Westinghouse Data, 1935-1954, *see* Grunfeld (1958)].

			Number of Times in 100 Trials that an Estimator of β_{11} Beats the USUAL Estimator			
T	E	ρ	GLS	SZ	HS	OLS
10	5	.3	58	58	49	45
10	10	.3	55	57	45	45
10	5	.6	55	56	50	40
10	10	.6	59	57	48	37
10	5	.925	57	53	60	37
10	10	.925	64	54	53	31

Note: Counts greater than 60 or smaller than 40 are significantly different from 50 at the .05 level.

First we consider the original data set used by ZELLNER [1962] in his seminal paper on seemingly unrelated regressions. This is the GRUNFELD [1958] investment data on two firms, General Electric and Westinghouse. In this case, the extra observations for $T=10$ (and $E=5$, 10) are of a time series nature. Next, we consider extra observations that are cross-sectional in nature. These data are extracted from a panel data set of cigarette consumption for 46 states over the period 1963 to 1980, see BALTAGI and LEVIN [1986]. In particular, we consider cigarette consumption for the years 1975, 1976, in which case the extra observations denote new states. The results for both data sets are, in general, the same as reported for the KMENTA and GILBERT [1968] type X. Feasible GLS estimators that seem to ignore the extra observations in estimating Σ (but not necessarily in estimating Σ^{-1} or β) do not generally do badly relative to feasible GLS estimators that seem to use the extra observations fully. See Table 4. Note that in all but one case, the "usual" is no different from true GLS, and for that case where true GLS is superior, the "usual" estimator is no different from the SZ and HS estimators.

8 Conclusion

Our Monte Carlo results replicate those of SCHMIDT [1977] and confirm his results. We find that Schmidt's results are robust to the type of additional observations available, whether they are time-series or cross-section in nature. An important result in this paper is that the extra observations may lead to better estimates of the variance-covariance matrix of the disturbances Σ or its inverse Σ^{-1}, but that does not necessarily lead to better estimates of the β's.

• References

BALTAGI, B. H. (1981). — "Pooling: An Experimental Study of Alternative Testing and Estimation Procedures in a Two-Way Error Component Model", *Journal of Econometrics*, 17, pp. 21-49.

BALTAGI, B. H. and LEVIN, D. (1986). — "Estimating Dynamic Demand for Cigarettes Using Panel Data: The Effects of Bootlegging, Taxation and Advertising Reconsidered", *The Review of Economics and Statistics*, 68, pp. 148-155.

BREUSCH, T. S. (1980). — "Useful Invariance Results for Generalized Regression Models", *Journal of Econometrics*, 13, pp. 327-340.

GRUNFELD, Y. (1958). — "The Determinants of Corporate Investment", Unpublished doctoral dissertation (University of Chicago, Chicago, I1.).

HOCKING, R. R. and SMITH, W. B. (1968). — "Estimating of parameters in the Multivariante Normal Distribution with Missing Observations", *Journal of the American Statistical Association* 63, pp. 154-173.

HWANG, H. S. (1987). — "Estimation of Linear SUR Model with Unequal Numbers of Observations", mimeo, Texas A & M University.

KMENTA, J. and GILBERT, R. F., (1968). — Small Sample Properties of Alternative Estimators of Seemingly Unrelated Regressions, *Journal of the American Statistical Association*, 63, pp. 1180-1200.

MADDALA, G. S. and MOUNT, T. D. — "A Comparative Study of Alternative Estimators for Variance Components Models Used in Econometric Applications" *Journal of the American Statistical Association*, 68, pp. 324-328.

SCHMIDT, P. (1977). — "Estimation of seemingly unrelated regressions with unequal numbers of observations", *Journal of Econometrics*, 5, pp. 365-377.

SRIVASTAVA, J. M. and ZAATAR, M. K. (1973). — "Monte Carlo Comparison of Four Estimators of Dispersion Matrix of a Bivariate Normal Population, Using Incomplete Data", *Journal of the American Statistical Association*, 68, pp. 180-183.

TAYLOR, W. E. (1980). — "Small Sample Considerations in Estimation from Panel Data", *Journal of Econometrics*, 13, pp. 203-223.

WILKS, S. S. (1932). — "Moments and Distributions of Estimates of Population Parameters from Fragmentary Samples", *Annals of Mathematical Statistics*, 3, pp. 167-195.

ZELLNER, A. (1962). – "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias", *Journal of the American Statistical Association*, 57, pp. 348-368.