Bank of Canada — Banque du Canada

June 1973

Technical Report 1

A Monte Carlo Study of the Estimation of an Overidentified Model with Temporally Dependent Residuals

Jean-Pierre Aubry
Research Department

ACKNOWLEDGEMENTS

In preparing the English version of this paper I am grateful to have had the comments of Donald Stephenson. I especially wish to thank Georgia Whitfield and Margaret Bailey for the work they did in clarifying the text.

AVANT-PROPOS

Un des problèmes auxquels nous faisons face lors de l'estimation d'un modèle économétrique est la présence de résidus corrélés instantanément et temporellement. Dans l'étude de Monte-Carlo qui suit, nous tentons de visualiser l'effet de ce problème sur les estimateurs obtenus par six différentes méthodes d'estimation.

Le modèle utilisé est un modèle fictif à deux équations sans variables endogènes retardées. Dans ce système nous faisons varier différents paramètres: la taille de l'échantillon, le niveau de corrélation instantanée, le niveau d'autocorrélation et la variance des résidus.

Cette étude est tirée de ma thèse de maîtrise (texte français) présentée à l'Université de Montréal.

ABSTRACT

When building an econometric model, the equation residuals may be both simultaneously and temporally correlated. In the following Monte Carlo Study I try to visualize the effect of that problem on the estimators generated by six different methods of estimation.

I postulate a two-equation model without a lagged endogenous variable. In this system I change various parameters: the sample period size, the level of simultaneous correlation, the level of autocorrelation, and the residual variances.

This study is the second part of my MA thesis, which is written in French, presented at the University of Montreal.

A Monte Carlo Study of the Estimation of an Overidentified Model with Temporally Dependent Residuals

1. Objectives

In this paper I compare the results of six different methods of estimation applied to a system of two overidentified equations that pose simultaneous correlation and autocorrelation problems. (To isolate the effects of these two problems I have not allowed lagged endogenous variables to act as explanatory variables.) When quarterly data are used, the probability of encountering both problems is high. The different properties of each estimation method are well known. (This study is intended to illustrate these properties and their robustness.)

2. Theory

The six methods of estimation to be used are:

1 OLS - Ordinary Least Squares (Biased estimates.)

2 GLS - Generalized Least Squares (Biased estimates. Cochrane-Orcutt technique.)

3 2SLS - Two-Stage Least Squares (Consistent estimates.)

4 3SLS - Three-Stage Least Squares (Consistent estimates, more efficient than the 2SLS estimates.)

5 A2SLS - Adapted Two-Stage Least Squares (Consistent estimates, more efficient than the 2SLS estimates. I obtain the 2SLS estimates and from the residuals

calculate an autocorrelation coefficient. Then I transform the variables of the last regression $(Z_t - \hat{P}Z_{t-1})$ and apply the OLS method.

6 A3SLS - Adapted Three-Stage Least Squares (Consistent estimates, more efficient than the 3SLS and the 2SLS estimates. Instead of using the OLS method in the last step of the A2SLS method I apply to all the equations of the system the ZGLS (Zellner Generalized Least Squares) estimate.)

NOTE For more information on these methods see the bibliography or my thesis. These six methods may be divided into three groups.

1 The methods used to estimate each equation of the system independently, OLS and GLS.

2 The methods known as Limited Information Methods, 2SLS and A2SLS.

3 The Full Information Methods, 3SLS and A3SLS.

Note that GLS, A2SLS and A3SLS take account of autocorrelation problems.

3. Generation of the System

An artificial model of two overidentified equations is used:

$$y_1(t) = \alpha_{11}x_1(t) + \alpha_{12}x_3(t) + \alpha_{13}y_2(t) + u_1(t)$$

$$y_2(t) = \alpha_{21}x_2(t) + \alpha_{22}x_4(t) + \alpha_{23}y_1(t) + u_2(t)$$

where there are four exogenous variables

$x_1(t)$: a constant variable

$X_2(t)$: a trend variable

$x_3(t)$: a cyclical variable (sine function)

$X_4(t)$: a dummy variable (Q4)

and $u_1(t)$ and $u_2(t)$ are the residuals

The true values of the coefficients are:

$\alpha_{11} = 10$ $\alpha_{12} = 1$ $\alpha_{13} = .5$

$\alpha_{21} = .7$ $\alpha_{22} = 2$ $\alpha_{23} = .2$

In order to obtain values for the endogenous variables, residuals are generated and added to each equation; the system is then solved for the two unknown values. The residuals are created as follows:

$$u_{1M}(I) = .8\ u_{1M}(I-1) + .6\ v(I)$$

$$u_{2M}(I) = .6\ u_{2M}(I-1) + .52\ v(I) + .178\ \varepsilon(I)$$

$$u_{1F}(I) = u_{1M}(I)$$

$$u_{2F}(I) = .6\ u_{2F}(I-1) + .78\ v(I) + .178\ \varepsilon(I)$$

where $\varepsilon(I)$ and $v(I)$ are independent identically and normally distributed variables $N(0,\lambda)$, and the subscripts M and F distinguish between a moderate and a higher level of simultaneous correlation. Three sets of observations are created, one each for three levels of $\sqrt{\lambda}$, the residual standard error (RSE). The three levels used are 1.0, .2, and 2.0 since, by definition, $RSE(\varepsilon)=RSE(v)=\sqrt{\lambda}$. We thus obtain:

$RSE(\varepsilon)$	1.000	.200	2.000
$RSE(u_1M)$ =	1.000	.200	2.000
$RSE(u_2M)$	.687	.137	1.374
$RSE(u_1F)$	1.000	.200	2.000
$RSE(u_2F)$	1.000	.200	2.000

The ratio of the RSE of each residual to the mean of the corresponding endogenous variable is approximately 5% in the first case ($RSE(\varepsilon) = 1.00$), 1% in the second case, and 10% in the third case. One can see that in each equation a Markov chain is generated with a coefficient of autocorrelation of .8 in the first equation and of .6 in the second. As well, there is no lagged value for the endogenous variable in either equation. This simplification is adopted in order to isolate the effect of autocorrelation. If there had been lagged endogenous values, the 2SLS and 3SLS methods would have given biased estimates and it would be impossible to compare their efficiency with that of the A2SLS and A3SLS methods.

Convergence is one of the properties influencing a choice of method. To demonstrate how fast the estimates from the above methods converge towards the true value, I took samples of 20, 50, and 100 observations. I also experimented with different levels of the residual variance. Finally, because the estimate of the coefficient of autocorrelation has an influence on the result, I wish to see what happens when the true values (.8 and .6) are used and when the GLS, A2SLS, and A3SLS methods are applied.

NOTE: GLS(ρ) is the real value of the autocorrelation coefficient, as is A2SLS(ρ) and A3SLS(ρ).

4. Results and Interpretation

For each of the three levels of residual variance and for each of the two levels of simultaneous correlation, 100 samples were generated for each of the three sample sizes, after which the six methods of estimation were applied to each sample. All these results are compiled and summarized in Tables 1 to 5 and in Charts 1.1 - 1.6.

A) The Average of the Proportional Bias

The bias was estimated by taking the average of the 100 estimates of a coefficient and by subtracting this value from the true value. The statistic in Table 1, which is the average over the six coefficients of the bias divided by the true coefficient value, is constructed as shown below.

$$\left[\sum_{i=1}^{2} \sum_{j=1}^{3} \left| \left(\sum_{k=1}^{100} (\hat{\alpha}_{ij})_k / 100 - \alpha_{ij} \right) / \alpha_{ij} \right| \right] / 6$$

MSE OF EACH COEFFICIENT WHEN RSE IS 1

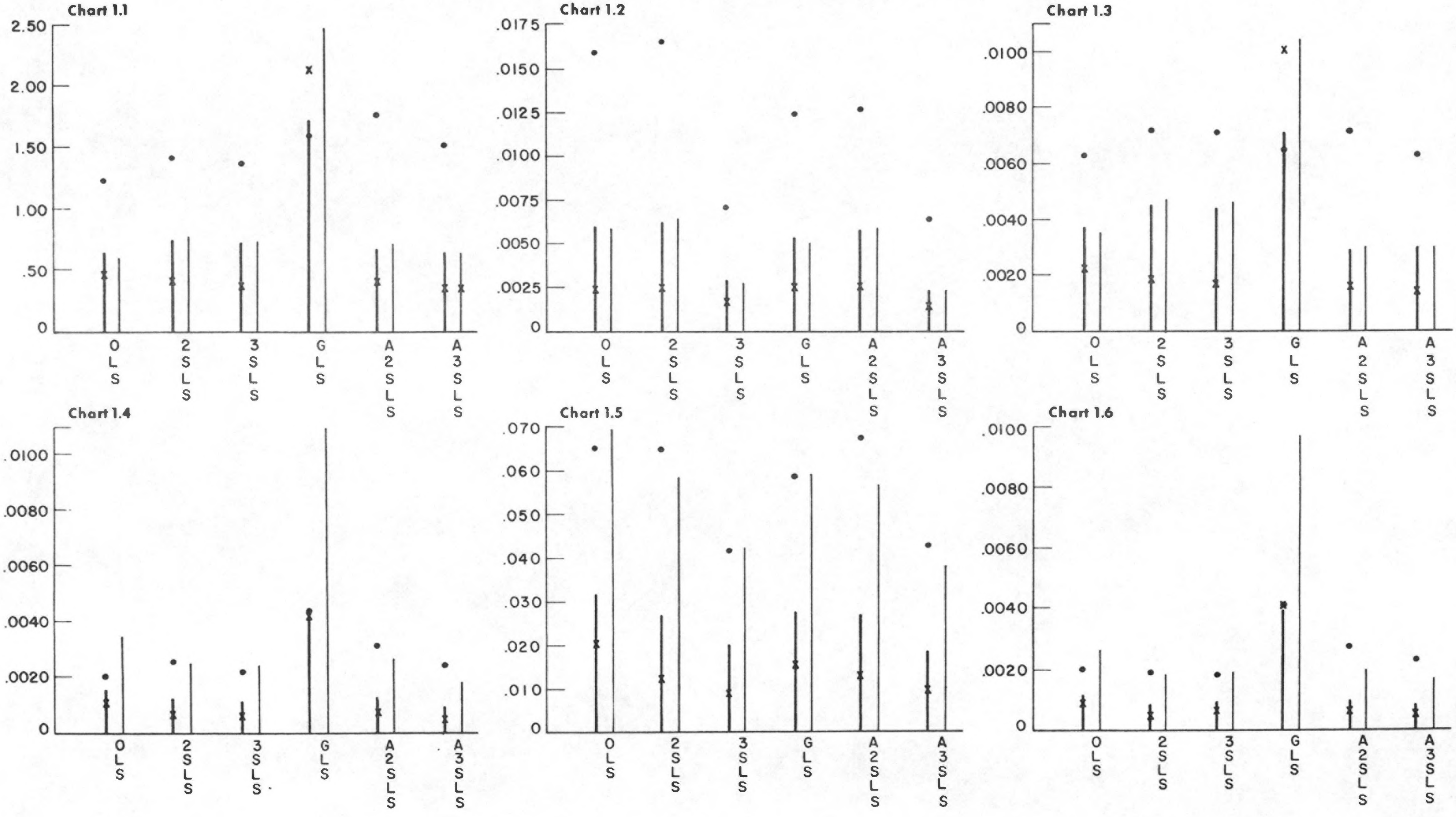

Chart 1

Chart 1.1:	MSE of the estimates of α_{11}
1.2:	MSE of the estimates of α_{12}
1.3:	MSE of the estimates of α_{13}
1.4:	MSE of the estimates of α_{21}
1.5:	MSE of the estimates of α_{22}
1.6:	MSE of the estimates of α_{23}

Thin line:	sample size 50 with high simultaneous correlation
Large line:	sample size 50 with moderate simultaneous correlation
Point (.):	sample size 20 with moderate simultaneous correlation
(X):	sample size 100 with moderate simultaneous correlation

The OLS Method of Estimation

Note in Table 1 that use of the OLS method produces inconsistent estimates and that the bias does not always decrease when the size of the sample is increased. When the residual standard error (RSE) is 1 or 2 the bias goes up. When the RSE is .2 the bias seems tc be decreasing to zero as the size of the sample increases without limit. In all three cases the bias should be proportional to the residual variance. On the other hand, when the simultaneous correlation of the residuals is higher, the probability of having a greater bias is higher. However, the impact of the level of simultaneous correlation decreases with a lower residual standard error. A close examination of the coefficients reveals that the coefficient on the explanatory endcgenous variable has the largest bias of all coefficients in each equation. For example, in a sample size of 50 observations with a residual variance of 1 the percentage bias of each coefficient is:

$\alpha 11$	$\alpha 12$	$\alpha 13$	$\alpha 21$	$\alpha 22$	$\alpha 23$
-.16%	-.02%	-.30%	-3.5%	-4.1%	11%

The GLS Method of Estimation

The GLS results are surprising. Initially I thought that they would have a lower bias than the OLS estimates because the GLS method takes account of the autocorrelation problem. This was not the case and indeed opposite results often emerged as can be seen in Table 1. One can not attribute this outcome

Table 1

THE AVERAGE PROPORTIONAL BIAS

Residual Standard Error	Method of Estimation	Sample Size 20		Sample Size 50		Sample Size 100	
		Moderate	High	Moderate	High	Moderate	High
1.0	OLS	3.73	4.63	3.23	5.23	4.60	7.02
	2SLS	2.63	2.54	1.41	1.35	.42	.78
	3SLS	2.80	2.71	1.69	1.61	.41	.83
	GLS	6.16	10.09	10.59	16.19	12.77	19.01
	GLS(ρ)	13.32	19.90	14.51	21.24	14.21	20.83
	A2SLS	0.84	0.98	.83	.97	1.43	1.76
	A2SLS(ρ)	1.36	1.41	1.30	1.45	1.62	2.00
	A3SLS	1.00	1.28	.91	1.15	1.31	1.67
	A3SLS(ρ)	1.05	1.17	1.17	1.31	1.38	1.67
2.0	OLS	8.57	14.23	12.88	20.49	16.91	25.24
	2SLS	2.72	3.79	2.27	3.12	1.74	2.55
	3SLS	3.18	4.64	2.64	3.57	1.77	2.73
	GLS	23.04	34.81	34.98	49.14	39.90	54.39
	GLS(ρ)	40.70	54.32	42.45	55.72	41.85	55.00
	A2SLS	1.33	1.52	1.61	1.94	2.76	3.50
	A2SLS(ρ)	1.85	2.20	2.07	2.43	3.00	3.79
	A3SLS	1.76	2.25	2.00	2.83	2.47	3.30
	A3SLS(ρ)	1.73	2.92	1.85	2.38	2.41	3.09
0.2	OLS	2.98	3.01	1.61	1.65	.69	.70
	2SLS	2.93	2.92	1.53	1.50	.59	.52
	3SLS	3.17	3.23	1.67	1.68	.64	.58
	GLS	1.36	1.43	.34	.55	.51	.77
	GLS(ρ)	5.18	8.47	.63	.97	.69	.95
	A2SLS	0.69	0.68	.31	.35	.42	.46
	A2SLS(ρ)	1.09	1.10	.73	.76	.57	.65
	A3SLS	0.79	0.83	.36	.39	.38	.43
	A3SLS(ρ)	.87	.85	.64	.65	.49	.55

to faulty estimation of the coefficient of autocorrelation because the results of the GLS(ρ) estimates are even worse. Apparently the GLS estimates are not robust when there is simultaneous correlation in the residuals and an endogenous variable among the independent variables. Later, in the general interpretation, I shall explain this situation. When one compares the increase of the bias from a moderate level of simultaneous correlation to a higher level, one sees that the increase of the GLS bias is twice or three times that of the OLS bias.

The Four Other Methods of Estimation

The results of using the four other methods seem to follow the theory. When the 2SLS and the 3SLS methods are applied the bias decreases rapidly with the increased sample size. Use of these two methods produces convergent estimates of the coefficient in the absence of a lagged dependent variable. When the more efficient A2SLS and A3SLS methods are employed it is difficult to see whether or not the bias is asymptotically zero, because even with my small sample sizes the bias is already small. A higher level of residual standard error and a higher level of simultaneous correlation result in a higher bias. This is also the outcome when I use the OLS and GLS methods. I also find that a priori knowledge of the autocorrelation coefficient (as in the case of the GLS method) has an adverse effect on the A2SLS and A3SLS estimates.

B) Mean Square Error (MSE)

The mean square error is a statistic that measures the scattering of the estimate around the true value.

$$MSE(\alpha_{ij}) = \sum_{k=1}^{100} ((\hat{\alpha}_{ij})_k - \alpha_{ij})^2/100$$

One can prove that the MSE equals the sample variance plus the square of bias where the sample variance is defined as:

$$\sum_{k=1}^{100} ((\hat{\alpha}_{ij})_k - \bar{\hat{\alpha}}_{ij})^2/100$$

The MSE is generally lower when the simultaneous correlation is lcwer, according to Charts 1.1 - 1.6. Most of this result can be explained by the increasing bias accompanying a higher level of simultaneous correlation as depicted in Table 1. However, the sample size is much more important. In other words, it is very useful to have a large sample in order to minimize the MSE because the sample variance is greatly affected by the sample size. This is so in the case of a residual standard error of 1 for a moderate level of simultaneous correlation and for the coefficient α_{11}. For different sample sizes these variances are:

Sample Size	20	50	100
Variance α_{11} (OLS)	1.158	.632	.387

The result is the same for the other coefficients and methods. This is important because in most econometric work with economic relationships only a single sample is used, not, as here, 100 samples, so that a random factor can divert the result from the true value. The higher the variance the greater the probability that the result will be two or three standard errors away from the mean. For example, when I build a just-identified model ($x_3=x_4=0$), using the GLS and 2SLS methods, I obtain in two or three of the 100 samples of 20 observations certain coefficients that are completely different from the true values.

Also, if one omits the GLS method, the efficiency of the other methods is in increasing order: 2SLS, A2SLS, OLS, 3SLS, and A3SLS. The first three methods have approximately the same order of mean square error (MSE). A big gain is realized by using the full-information methods. It is surprising that the OLS estimates with their bias have a MSE of the same order as that of the 2SLS and A2SLS estimates. This occurs because the OLS estimates are concentrated around their mean and have little sample variance, whereas the opposite happens in the case of the 2SLS and the A2SLS estimates - demonstrating the robustness of the OLS estimates. For the sample size 50, for a moderate level of simultaneous correlation, and for the residual variance of 1.0, the order of the variance of the coefficients is: 2SLS, A2SLS, GLS, OLS, 3SLS, A3SLS.

It is evident that the small variance of the GLS estimators does not counterbalance the large bias (because of the simultaneous correlation) to produce a lower MSE. Hence one notes with interest the robustness of the OLS estimates. Their bias is not too high when the residual variance is moderate. The level of the sampling variance is low enough to counterbalance the bias effect in order to have a good MSE. On the other hand, if the 2SLS and the A2SLS estimates are interesting for their smaller bias, their large variance is important enough to push up the MSE. The 3SLS method and (especially) the A3SLS method have advantages: a small bias and a small variance, implying a small MSE. However, the application of these two methods is practically impossible when a large model is being used.

C) Confidence Interval Tests

I calculated the number of times out of the 100 repetitions that the estimated value of each coefficient is within a 95% confidence interval of the true value. Note in Table 2 the sum of the tests for the three coefficients of each equation. Note also the sum over both equations, and the aggregation of the two systems (the moderate and high level of simultaneous correlation). When these three types of sums are examined, the following is seen:

maximum possible per equation - 300 times,
optimum observed from samples - 285 times,
maximum possible per system - 600 times,
optimum observed from samples - 570 times,
maximum possible over aggregated systems - 1200 times,
optimum observed from samples - 1140 times.

Table 2

CONFIDENCE INTERVAL TESTS

Residual Standard Error	Method of Estimation	Sample Size 20			Sample Size 50			Sample Size 100		
		Moderate	High	1200	Moderate	High	1200	Moderate	High	1200
1.0	OLS	189+226=415	187+225=412	827	189+214=403	196+207=403	806	179+170=369	173+179=352	721
	2SLS	177+246=423	179+244=423	846	188+228=416	188+226=414	830	194+244=438	196+241=437	875
	3SLS	186+240=426	192+240=432	858	185+216=401	185+216=401	802	196+230=426	193+225=418	844
	GLS	207+231=438	202+221=423	861	188+183=371	150+163=313	684	126+122=248	102+105=207	
	GLS(ρ)	245+231=483	209+223=432	915	148+163=311	114+129=243	554	114+ 99=213	98+ 88=213	399
	A2SLS	232+262=494	234+266=500	994	279+285=564	276+285=561	1125	282+281=563	279+281=560	1123
	A2SLS(ρ)	299+289=588	299+291=590	1178	298+287=585	298+286=584	1169	299+286=585	300+286=586	1171
	A3SLS	240+270=510	239+274=513	1023	272+285=557	272+284=556	1113	283+285=568	277+285=562	1130
	A3SLS(ρ)	300+289=589	300+291=591	1180	298+292=590	298+291=559	1179	299+290=589	300+290=590	
2.0	OLS	193+212=405	192+207=399	804	192+164=356	182+149=331	687	159+105=264	137+ 93=232	494
	2SLS	183+246=429	184+246=430	859	189+231=420	187+232=419	839	196+245=441	198+212=440	881
	3SLS	188+241=429	191+241=432	861	177+219=396	184+222=406	802	193+231=424	188+229=417	841
	GLS	191+172=363	170+152=322	685	115+ 94=209	97+ 84=181	390	93+ 68=161	89+ 58=147	308
	GLS(ρ)	152+148=300	114+117=231	531	93+ 80=173	91+ 78=169	342	91+ 70=161	89+ 61=150	311
	A2SLS	236+263=499	234+263=497	996	279+285=564	278+284=562	1126	280+282=562	276+279=555	1117
	A2SLS(ρ)	299+284=583	300+282=582	1165	298+288=586	296+287=583	1169	300+286=586	300+ 28=328	914
	A3SLS	238+265=503	238+268=506	1009	272+285=557	273+284=557	1114	279+280=559	275+282=557	1116
	A3SLS(ρ)	300+286=586	300+283=583	1169	297+289=586	296+287=583	1169	300+285=585	300+286=586	1171
0.2	OLS	132+192=324	134+210=344	668	154+213=367	157+224=381	748	197+232=429	197+236=433	862
	2SLS	123+196=319	123+214=337	656	151+223=374	151+229=380	754	192+240=432	191+243=434	866
	3SLS	111+173=284	113+194=307	591	142+206=348	144+214=358	706	186+231=417	184+234=418	835
	GLS	190+244=434	192+251=443	877	257+267=524	257+267=524	1048	251+271=522	243+267=510	1032
	GLS(ρ)	277+276=553	274+276=550	1103	275+275=550	263+274=537	1087	267+258=525	261+262=523	1048
	A2SLS	226+269=495	225+271=496	991	282+293=575	282+291=513	1148	290+287=577	289+286=575	1152
	A2SLS(ρ)	295+297=595	269+295=594	1189	297+294=591	297+293=590	1181	300+291=591	300+290=590	1181
	A3SLS	224+269=493	221+271=492	985	283+291=574	280+291=571	1145	288+289=577	288+288=576	1153
	A3SLS(ρ)	298+296=594	299+295=594	1188	297+296=593	297+296=593	1186	299+293=592	299+294=593	1185

Further results appear in Table 2.

1 The level of simultaneous correlation seems to have little effect on the test.

2 The best results were secured with a lower level of autocorrelation (second equation of the system).

3 The effect of the level of residual variance is hard to isolate. The best results were obtained with a moderate level of residual variance and with a sample of moderate size.

The choice of method has the greatest effect on the results. For the interpretation that follows I use the results of tests for sample size 50 with a moderate level of simultaneous correlation and a residual variance of 1. The set of results is typical.

Use of the GLS method (except in the case of a residual variance of .2) gives the poorest estimation of the confidence interval. In a typical case, I had only 62% and 52% (known) success instead of 95%. The size of the bias explains a great part of this failure. As well, the distribution of the number of times that the real value of a coefficient is within the confidence interval differs greatly with each coefficient. Of the total of 148 times for the three coefficients of the first equation, the coefficient of the cycle variable X_3 was within the interval ninety-four times, and in the second equation the dummy had eighty-eight successes of the 163 successes for all three coefficients. The true value of the endogenous variables is very often above the upper confidence limit; the real coefficients of the trend and dummy variables are often below the lower confidence limit. This dispersion is much more explicit with the samples of size 100:

	α11	α12	α13	α21	α22	α23
within	20	94	12	22	84	16
below	0	0	88	0	0	84
above	80	6	0	78	16	0
% of the bias	-13.4	-2.5	18.8	-8.6	-3.8	29.5

The OLS Method

As stated above the OLS method produces some biased estimates, but this bias is not important enough to explain the poor results shown in Table 2. The standard error of each coefficient is a biased estimate; compared to the GLS, 2SLS, and 3SLS estimates, the OLS estimates follow a normal distribution pattern. Although in each case the residuals are not normally distributed, the pattern followed by each OLS estimate is similar to the GLS pattern but less assymetric.

Sample Size 50

	α11	α12	α13	α21	α22	α23
within	51	91	47	59	96	59
below	23	5	27	2	0	39
above	26	4	26	39	4	2

It will be apparent later that this underestimation of the confidence interval is generated by an underestimation of the residual variances.

The 2SLS and 3SLS Methods

The results of using the 2SLS and 3SLS methods are no better than those of using the OLS method. Of course the bias cannot have a serious effect. This is the classical case of underestimation of the standard error of the residuals because these two methods do not take account of the autocorrelation problem.

The A2SLS and A3SLS Methods

By using these two methods I made a significant gain. In a typical set of results, the real values are ninety-four times out of 100 within a 95% confidence interval employing the A2SLS method and ninety-three times employing the A3SLS method. As will be evident later (see Table 4), it appears that the estimation of the residual variances explains part of this success. On the other hand knowledge of autocorrelation coefficients provides an overestimation of the standard errors.

I calculate the follcwing statistic:

$$\left[\sum_{i=1}^{2}\sum_{j=1}^{3}\left(\frac{\sum_{k=1}^{100} S.E.(\hat{\alpha}_{ij})_k/100}{\sqrt{\sum_{k=1}^{100}[(\hat{\alpha}_{ij})_k-\bar{\hat{\alpha}}_{ij}]^2/100}}\right)\right]/6$$

It is an average over the coefficients of the ratio of the mean standard errors of the coefficients to their sample standard errors. The values for this statistic are given in Table 3.

All these results confirm what I have stated above. The mean of the standard errors are much nearer to the sample standard errors of the coefficients when the A2SLS and A3SLS methods are used rather than the other methods.

D) Residual Variances

The average of the estimates of the residual variances is tabulated in Table 4.

Clearly the sample size is important - the larger the sample size the better the estimate of the residual variance. Important also is the level of the true residual variances. The bias is proportional to the residual variance. For example, when the residual standard error is .2, the bias is around 15% for a sample size of 20 and is just 2.5% for a sample size of 100. On the other hand, when the residual standard error is 1, the bias is 13% for a sample size of 100.

Table 3

AVERAGE OVER COEFFICIENTS OF ($\frac{\text{Average of single-sample estimates of SE of a coefficient}}{\text{SE of a coefficient estimated from all samples}}$)

Residual Standard Error	Method of Estimation	Sample Size 20		Sample Size 50		Sample Size 100	
		Moderate	High	Moderate	High	Moderate	High
1.0	OLS	62.25	63.20	65.99	66.70	72.00	73.02
	2SLS	60.81	61.08	64.15	63.86	70.10	70.11
	3SLS	56.18	56.33	55.56	54.11	56.89	55.47
	GLS	69.81	69.03	78.90	77.67	82.88	81.51
	GLS(ρ)	99.80	103.71	102.61	108.29	106.24	111.40
	A2SLS	90.13	91.69	108.98	107.33	116.82	115.03
	A2SLS(ρ)	132.07	133.73	122.40	121.44	128.29	127.49
	A3SLS	90.19	92.35	110.79	109.07	116.96	114.46
	A3SLS(ρ)	129.09	132.86	124.26	122.59	129.83	129.17
2.0	OLS	69.93	70.45	72.60	76.82	78.57	82.90
	2SLS	61.55	62.47	64.14	63.90	70.07	70.07
	3SLS	56.64	57.30	55.22	53.77	56.36	54.85
	GLS	66.83	67.39	77.70	79.45	82.75	84.79
	GLS(ρ)	110.69	117.28	117.49	122.45	117.72	119.16
	A2SLS	91.58	94.15	107.45	105.57	116.91	115.36
	A2SLS(ρ)	132.40	135.52	120.09	118.75	127.86	127.22
	A3SLS	941.96	95.12	108.13	104.98	116.31	113.01
	A3SLS(ρ)	134.83	130.39	120.33	115.37	128.47	126.77
0.2	OLS	66.58	64.69	69.02	67.46	73.47	72.75
	2SLS	66.45	64.57	69.00	67.40	73.47	72.74
	3SLS	60.80	60.56	59.84	58.76	59.99	59.01
	GLS	69.69	70.60	82.96	82.58	86.80	86.32
	GLS(ρ)	94.76	95.46	93.23	93.50	84.64	86.92
	A2SLS	105.18	101.85	124.99	120.45	125.53	122.22
	A2SLS(ρ)	188.52	178.24	148.33	143.97	139.55	137.29
	A3SLS	120.20	101.16	126.26	123.94	125.28	122.28
	A3SLS(ρ)	169.69	163.87	148.88	147.91	139.76	140.21

Table 4

ESTIMATION OF THE RESIDUAL VARIANCE

Residual Standard Error	Method of Estimation	Sample Size 20				Sample Size 50				Sample Size 100			
		Moderate		High		Moderate		High		Moderate		High	
		(1.000)	(.472)	(1.000)	(1.000)	(1.000)	(.472)	(1.000)	(1.000)	(1.000)	(.472)	(1.000)	(1.000)
1.0	OLS	.494	.347	.504	.734	.694	.380	.697	.795	.828	.417	.814	.873
	2SLS	.503	.358	.522	.766	.708	.397	.725	.841	.846	.439	.848	.936
	3SLS	.529	.370	.552	.793	.723	.400	.742	.850	.857	.441	.860	.941
	GLS	.291	.242	.283	.498	.302	.254	.281	.519	.306	.264	.275	.538
	GLS(ρ)	.297	.252	.267	.511	.295	.257	.263	.519	.319	.264	.284	.537
	A2SLS	.615	.413	.649	.864	.750	.417	.757	.874	.864	.450	.860	.954
	A2SLS(ρ)	1.163	.450	1.287	1.926	.848	.422	.869	.884	.891	.450	.891	.955
	A3SLS	.599	.417	.634	.875	.764	.427	.768	.896	.876	.451	.872	.957
	A3SLS(ρ)	1.174	.430	1.258	.890	.871	.424	.893	.893	.907	.448	.907	.952
		(4.000)	(1.888)	(4.000)	(4.000)	(4.000)	(1.889)	(4.000)	(4.000)	(4.000)	(1.889)	(4.000)	(4.000)
2.0	OLS	1.990	1.311	2.010	2.711	2.708	1.384	2.646	2.807	3.085	1.483	2.895	2.999
	2SLS	2.153	1.503	2.372	3.257	2.944	1.629	3.114	3.488	3.337	1.813	3.382	3.907
	3SLS	2.273	1.555	2.525	3.385	3.011	1.648	3.195	3.536	3.405	1.824	3.437	3.935
	GLS	1.023	.791	.891	1.514	.950	.801	.743	1.501	.914	.799	.674	1.477
	GLS(ρ)	.896	.757	.680	1.407	.870	.778	.653	.447	.901	.797	.672	1.491
	A2SLS	2.480	1.703	2.659	3.605	3.018	1.698	3.093	3.587	3.411	1.847	3.394	3.916
	A2SLS(ρ)	4.464	1.834	5.051	3.832	3.357	1.720	3.469	3.637	3.512	1.852	3.511	1.829
	A3SLS	2.468	1.701	2.682	3.625	3.090	1.739	3.170	3.684	3.470	1.842	3.456	3.939
	A3SLS(ρ)	4.270	1.701	5.470	3.606	3.442	1.721	3.595	3.655	3.583	1.829	3.587	3.908
		(.040)	(.0189)	(.040)	(.040)	(.040)	(.019)	(.040)	(.040)	(.040)	(.019)	(.040)	(.040)
0.2	OLS	.024	.017	.024	.032	.034	.017	.034	.035	.039	.018	.039	.038
	2SLS	.024	.017	.024	.033	.034	.018	.034	.035	.039	.018	.039	.038
	3SLS	.025	.018	.026	.034	.035	.018	.035	.035	.039	.019	.040	.038
	GLS	.013	.010	.013	.022	.013	.011	.013	.023	.014	.012	.014	.025
	GLS(ρ)	.013	.011	.013	.024	.013	.011	.013	.024	.026	.013	.026	.026
	A2SLS	.040	.024	.041	.041	.043	.020	.042	.038	.042	.019	.042	.039
	A2SLS(ρ)	.112	.029	.117	.048	.055	.020	.056	.039	.045	.019	.045	.039
	A3SLS	.038	.024	.038	.042	.043	.020	.042	.039	.042	.019	.042	.039
	A3SLS(ρ)	.112	.027	.118	.045	.057	.020	.058	.038	.045	.019	.045	.039

* Real values are shown in parentheses.

When the two best methods are used there is a small error of estimation in the case of the smallest residual variance for any sample size. It is obviously difficult to say whether the estimators are convergent or not. Theoretically the OLS, 2SLS, 3SLS, and GLS methods do not generally produce convergent estimates. However, the bias goes down as the sample size increases and as the true residual variance decreases. On the other hand, the A2SLS and A3SLS quickly converge to the true value. In Table 4 the efficiency and order of these methods are evident.

A3SLS, A2SLS, 3SLS, 2SLS, OLS, and GLS Methods

The GLS estimates used alone produce the lowest residual standard error but at what a price! The GLS residual standard error is 100% or 150% less than the residual standard error of the OLS or 2SLS estimates, each of which is an underestimation.

If one refers to Johnston for a very simple model

$$y_t = \beta x_t + u_t \qquad \text{where } u_t = \rho\, u_{t-1} + v_t$$

and applies the OLS, one can prove that

$$E\left(\frac{e'e}{n-1}\right) \cong \sigma_u^2 \left(1 - \frac{1+\rho^2}{1-\rho^2}\right)/(n-1)$$

where e is the vector of OLS residuals. If $\rho = .8$ and $n = 20$,

$$E\left(\frac{e'e}{n-1}\right) = .81\, \sigma_u^2$$

Using a simple model without a dependant variable one obtains an underestimation of 20%. Using my more complex model one should not be surprised by the OLS, 2SLS, and 3SLS results.

Unlike the simultaneous correlation the level of autocorrelation is important. A larger autocorrelation coefficient should produce a larger bias. In the sample proceeded by Johnston we have:

Sample Size	20	50	100
% of the bias (ρ=.6)	6	2	1
% of the bias (ρ=.8)	19	7	4

The same thing happened in my studies. The first equation, where the autocorrelation coefficient is higher than in the second equation, has a bias that is more pronounced. With a moderate level of simultaneous correlation and with a residual standard error of 1 the result is:

Sample Size	20	50	100
% of the bias (ρ=.6)	26	19	11
% of the bias (ρ=.8)	50	30	13

This problem is only important when there is a small sample error.

E) Autocorrelation Coefficients

I tabulate, in Table 5, the average of the estimates of the autocorrelation coefficients and the sample variance of these estimates. The estimates are generated from the OLS

Table 5

ESTIMATES OF THE AUTOCORRELATION COEFFICIENT AND SAMPLE VARIANCES OF THESE ESTIMATES

Residual Standard Error	Method of Estimation	Sample Size 20				Sample Size 50				Sample Size 100			
		Moderate		High		Moderate		High		Moderate		High	
		Eqn.1	Eqn.2	Eqn.1	Eqn.2	Eqn.1	Eqn.2	Eqn.1	Eqn.2	Eqn.1	Eqn.2	Eqn.1	Eqn.2
1.0	*(1)	.410	.380	.416	.385	.667	.508	.673	.513	.744	.557	.751	.562
		(.058)	(.043)	(.057)	(.046)	(.014)	(.017)	(.013)	(.016)	(.007)	(.008)	(.007)	(.609)
	**(2)	.377	.360	.369	.361	.614	.525	.600	.519	.687	.584	.671	.584
		(.056)	(.047)	(.057)	(.049)	(.015)	(.017)	(.016)	(.017)	(.0081)	(.0089)	(.0084)	(.0091)
2.0	*(1)	.435	.413	.462	.430	.689	.533	.713	.547	.765	.582	.788	.595
		(.058)	(.042)	(.058)	(.044)	(.013)	(.013)	(.015)	(.014)	(.006)	(.007)	(.006)	(.007)
	**(2)	.379	.363	.371	.364	.614	.526	.599	.520	.687	.584	.671	.574
		(.057)	(.048)	(.058)	(.050)	(.016)	(.016)	(.018)	(.017)	(.0082)	(.0089)	(.0086)	(.0092)
0.2	*(1)	.415	.363	.414	.363	.674	.499	.673	.501	.743	.548	.742	.548
		(.051)	(.040)	(.051)	(.044)	(.010)	(.016)	(.010)	(.016)	(.0055)	(.0081)	(.0055)	(.0086)
	**(2)	.395	.362	.381	.358	.629	.531	.614	.523	.694	.586	.678	.576
		(.049)	(.044)	(.052)	(.047)	(.012)	(.016)	(.013)	(.016)	(.0067)	(.0081)	(.0072)	(.0086)

*(1) Calculated from the OLS residuals

**(2) Calculated from the 2SLS residuals

residuals (for GLS) and the 2SLS residuals (for A2SLS and A3SLS), using the following definition:

$$\hat{\rho} = \frac{\sum_{t=2}^{n} \hat{e}_t \hat{e}_{t-1}}{\sum_{t=2}^{n} \hat{e}_{t-1}^2}$$

In studying Table 5, the underestimation of the autocorrelation coefficient strikes one immediately. The most important parameter is the sample size; as it increases, both the underestimation of the coefficient and the magnitude of the variance decrease. Theoretically the estimate of ρ, using the OLS residuals, is inconsistent, whereas using the 2SLS residuals (when there is no lagged endogenous variable) the estimate is consistent. With the increasing sample size shown in Table 5, both estimates seem to converge to the true values, and the first value is less underestimated. But it is difficult to know what will happen when the sample size is increased even further. Theoretically the first estimate has a certain asymptotic bias and the second estimate approaches the true value.

The results of using the truncated form of the GLS, A2SLS, and A3SLS methods (with the true values of the autocorrelation coefficient) indicate that the underestimation of ρ gives better estimates and that a better method of evaluating the autocorrelation coefficient generates poorer estimates of the coefficients of the structural form. Residual variance does not have an important effect, nor does the level of simultaneous correlation. On the other hand, the autocorrelation coefficient of the first equation is more underestimated than that of the second equation. I shall attempt to explain this in the following section.

5. General Interpretation

One of the most surprising results of the exercise is the very poor performance of the GLS method. The good results of lower residual variances can be explained by the fact that the relation is practically exact and the disturbances are not important. The variances of my estimates are so small that these estimates are near to the true values. But when the residual variance is much more important, the random disturbances can have a great effect.

The major difference from the case in which the GLS estimates are at their best is the dependence between the residuals and the explanatory variables. A look at Table 1 will show how influential is the simultaneous correlation on the bias (sometimes an increase of 50% of the average of the percentage bias).

Another important factor is the autoregressive pattern followed by the independent variables. If the structure of an independent variable is virtually similar to that of the dependent variable, such an independent variable will be used tc explain too large a part of the variance of the dependent variable. This will be to the disadvantage of the other independent variables or of the residuals. Only when a larger sample size is employed does use of the GLS method distinguish the difference between the two patterns. This is exactly what happens in my model. The dependent variable and the endogenous variable, which appears as an explanatory variable in the same equation, follow practically the same autoregressive pattern. The coefficient of the endcgenous variable has the largest positive bias. When a larger sample size is used, the GLS method begins to distinguish the autoregressive structure of the residuals and of the explanatory variables. In my model the endcgenous dependent variables corner all the autocorrelation. Thus correction of the autocorrelation by the GLS method results in "overcorrecticn" of this problem. It is therefore better to underestimate the real value of the autocorrelation coefficient.

The autoregressive structure of the exogenous variable can have an effect. When I constructed the model I chose the sample size to be the period of the sinusoidal variable instead of a period of eight cbservations. This sinusoidal variable along with the endogenous explanatory variable increase the autoregressive structure of the explanatory variables and purge practically all the autoregression from the residuals. The OLS estimate of the autocorrelation coefficient for a sample size of 20 is not significantly different from zero.

If the results of the just-identified model ($x_3=x_4=0$) are also compared, the estimates of the coefficients of the endogenous variables are overestimated even more because two of the variables have a strong autoregressive structure (specifically x_2); all the autoregressive structure of the residuals is explained by the endogenous explanatory variables.

According to Rao and Griliches the autoregressive structure of an exogenous variable can affect the results. Given the following model

$$y_t = \beta x_t + u_t$$

$$x_t = \gamma x_{t-1} + v_t$$

$$u_t = \rho u_{t-1} + w_t$$

where

$$E(v_t) = E(w_t) = E(v_t w_t) = E(w_t w_{t-1}) = E(v_t v_{t-1}) = 0$$

$$E(v_t^2) = \sigma_v^2,\ E(w_t^2) = \sigma_w^2, \qquad |\gamma| \leq 1,\ |\rho| < 1$$

$$E(uu') = R$$

Rao and Griliches prove that:

1 The variance of the GLS estimates equals

$$v(b_{GLS}) \cong \frac{\sigma_u^2}{\sum_1^T x_t^2} \frac{1-\rho^2}{1+\rho^2-2\rho\gamma} \qquad \text{for T reasonably large}$$

2 The variance of the OLS estimates equals

$$v(b_{OLS}) \cong \sigma_u^2 \frac{1+\rho\gamma}{1-\rho\gamma} \qquad \text{for T large}$$

$v(b_{OLS})$ is proportional to γ

3 The mean of the Cochrane-Orcutt estimator of the autocorrelation coefficient

$$E(\hat{\rho}) \cong \rho - \frac{\rho-\gamma}{T-1-(\frac{1+\rho\gamma}{1-\rho\gamma})},$$

implies that the bias of ρ is a function of ρ, T, and λ .

Another way of visualizing this problem and its interpretation is to compare it with the study of a similar case: the regression on the endogenous variable lagged one period where the residuals are autocorrelated.

$$y_t = \beta y_{t-1} + v_t$$

$$v_t = \rho v_{t-1} + \varepsilon_t \qquad \text{where } |\beta|<1,\ |\rho|<1$$

Here the lagged endogenous variable takes the place of another endogenous variable that follows practically the pattern of y . In this case one can prove that:

$$(1) \qquad \text{plim } (\hat{\beta}_{OLS} - \beta) = \frac{\rho(1-\beta^2)}{1+\beta\rho}$$

$$\text{plim } (\hat{\beta}_{OLS} - \beta) = \begin{cases} .43 \text{ for } \rho=.8 \text{ and } \beta=.5 \\ .44 \text{ for } \rho=.6 \text{ and } \beta=.3 \end{cases}$$

$$(2) \qquad \text{plim } (\hat{\rho}-\rho) = \frac{-\rho(1-\beta^2)}{1-\beta\rho} = - \text{ plim } (\hat{\beta}_{OLS} - \beta)$$

This is the same situation as that in my more complex model - the coefficient of the endogenous variable is overestimated and the autocorrelation coefficient is underestimated.

From these Monte Carlo studies the conclusions that appear to be important are: first, the robustness of the OLS estimates; second, the bad performance of the GLS estimates; third, the advantage of using the 2ASLS method instead of the 2SLS or 3SLS method; fourth, the little difference between the A2SLS and the A3SLS method.

BIBLIOGRAPHY

Amemiya, Takeshi: "Specification Analysis in the Estimation of Parameters of a Simultaneous Equation Model with Autoregressive Residuals." Econometrica, Vol.34, April 1966, pp.283-306.

Anderson, R.L.: "The Problem of Autocorrelation in Regression Analysis." Journal of the American Statistical Association, Vol.49, March 1949, pp.91-105.

Cragg, J.G.: "On the Relative Small-Sample Properties of Several Structural Equation Estimators." Econometrica, Vol.35, January 1967, pp.89-110.

Cragg, J.G.: "On the Sensitivity of Simultaneous - Equations Estimators to the Stochastic Assumptions of the Models." Journal of the American Statistical Association, Vol.61, March 1966, pp.136-151.

Dbravarajan: Estimation of Simultaneous Economic Relationships with Temporally Dependent Disturbance Terms. Mimeo. (Paper presented to the Econometric Society meetings, August 23-27, 1970, at Boulder, Colorado.)

Dhrymes, Phoabus J.: "On the Treatment of Certain Recurrent Non-linearities in Regression Analysis." Southern Economic Journal, Vol.33, March 1966, pp.187-196.

Dhrymes, Phoabus J., John Berner and Robert Cumins: Limited Information Estimation of Simultaneous Equation Models with Lagged Endcgenous Variables and Autoregressive Errors. Mimeo. Montreal, University of Montreal, October 1970, 30 l. (Discussion Paper No.183.)

Fair, Ray C.: "The Estimation of Simultaneous Equation Models with Lagged Endogenous Variables and First Order Serially Correlated Errors." Econometrica, Vol.38, May 1970, pp.507-516.

Fisher, Franklin M.: "The Choice of Instrumental Variables in the Estimation of Economy-Wide Econometric Models." International Economic Review, Vol.6, September 1965, pp.245-274.

Grenander, Ulf: "On the Estimation of Regression Coefficients in the Case of an Autocorrelated Disturbance." Annals of Mathematical Statistics, Vol.25, June 1954, pp.252-272.

Grenander, Ulf and Murray Rosenblatt: Statistical Analysis of Stationary Time Series. New York, Wiley, 1957. 300 p.

Johnston, J.: Econometric Methods. New York, McGraw-Hill, 1972. 437 p.

Kakwaki, N.C.: "The Unbiasedness of Zellner's Seemingly Unrelated Regression Equations Estimators." Journal of the American Statistical Association, Vol.62, March 1967, pp.141-142.

Klein, L.R.: "The Efficiency of Estimation in Econometric Models." In: Essays in Economics and Econometrics, edited by Ralph W. Pfouts. Chapel Hill, N.C., The University of North Carolina Press, 1960, 240 p.

Kmenta, Jan: On Estimation of Simultaneous Equation Systems with Autocorrelated Disturbances. East Lansing, Michigan State University, May 1970. (Workshop Paper No.6910.)

Kmenta, Jan and Roy F. Gilbert: "Estimation of Seemingly Unrelated Regressions and Autoregressive Disturbances." Journal of the American Statistical Association, Vol.65, March 1970, pp.186-197.

Kmenta, Jan and Roy F. Gilbert: "Small Sample Properties of Seemingly Unrelated Regressions." Journal of the American Statistical Association, Vol.63, December 1968, pp.1180-1200.

Koopmans, Tjalling C. and William C. Hood: "The Estimation of Simultaneous Linear Economic Relationships." In: Studies in Economic Method, edited by Wm. C. Hood and Tjalling C. Koopmans. New York, Wiley, 1953 pp.112-199. (Cowles Commission for Research in Economics Monograph No.14.)

Koopmans, T.C., H. Rubin and R.B. Leipnik: "Measuring the Equation Systems of Dynamic Economics." In: Statistical Inference in Dynamic Economic Models, edited by Tjalling C. Koopmans. New York, Wiley, 1950, pp.53-237. (Cowles Commission for Research in Economics Monograph No.10.)

Madansky, Albert: "On the Efficiency of Three-Stage Least-Squares Estimation." Econometrica, Vol.32, January-April 1964, pp.51-56.

Nagar, A.L.: "A Monte Carlo Study of Alternative Simultaneous Equation Estimators." Econometrica, Vol.28, July 1960, pp.573-590.

Rao, Potluri and Zvi Griliches: "Small-Sample Properties of Several Two-Stage Regression Methods in the Context of Auto-Correlated Errors." Journal of the American Statistical Association, Vol.64, March 1969, pp.253-269.

Sargan, J.D.: "The Estimation of Economic Relationships Using Instrumental Variables." Econometrica, Vol.26, July 1958, pp.393-415.

Sargan, J.D.: "The Estimation of Relationships with

Autocorrelated Residuals by Means of Instrumental Variables." Journal of the Royal Statistical Society Series B, Vol.21, No.1, 1959, pp.91-105.

Sargan, J.D.: "The Maximum Likelihood Estimation of Economic Relationships with Autoregressive Residuals." Econometrica, Vol.29, July 1961, pp.414-426.

Telser Lester G.: "Interative Estimation of a Set of Linear Regression Equations." Journal of the American Statistical Association, Vol.59, September 1964, pp.845-862.

Theil, Henri: Economic Forecasts and Policy. Amsterdam, North-Holland, 1961. 567 p.

Theil, Henri: Principles of Econometrics. New York, Wiley, 1971, pp.429-539.

Wagner, H.: "A Monte Carlo Study of Estimates of Simultaneous Linear Structural Equations." Econometrica, Vol.26, January 1958, pp.117-133.

Wickens, M.R.: "The Consistency and Efficiency of Generalized Least Squares in Simultaneous Equation Systems with Autocorrelated Errors." Econometrica, Vol.37, October 1969, pp.651-659.

Wise, J.: "Regression Analysis of Relationships Between Autocorrelated Time Series." Journal of the Royal Statistical Society Series B, Vol.18, No.2, 1956, pp.240-256.

Zellner, Arnold: "Econometric Estimation with Temporally Dependent Disturbance Terms." International Economic Review, Vol.2, May 1961, pp.164-178.

Zellner, Arnold: "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias." Journal of the American Statistical Association, Vol.57, June 1962, pp.348-368.

Zellner, Arnold: "Estimators for Seemingly Unrelated Regression Equations: Some Exact Finite Results." Journal of the American Statistical Association, Vol.58. December 1963, pp.977-992.

Zellner, Arnold and David S. Huang: "Further Properties of Efficient Estimators for Seemingly Unrelated Regression Equations." International Economic Review, Vol.3, September 1962, pp.300-313.

Zellner, Arnold and Henri Theil: "Three-Stage Least Squares: Simultaneous Estimation of Simultaneous Equations." Econometrica, Vol.30, January 1962, pp.54-78.

Imprimé au Canada

Printed in Canada