The general linear seemingly unrelated regression problem (Part II

Plasmans, J.E.J.; Van Straelen, R.A.P.

Publication date:
1970

EIT
19

J. Plasmans and R. van Straelen

The general linear seemingly unrelated regression problem

II Feasible statistical estimation and an application

Research memorandum

TILBURG INSTITUTE OF ECONOMICS
DEPARTMENT OF ECONOMETRICS

The General Linear Seemingly Unrelated Regression Problem

Part II Feasible Statistical Estimation and an Application

by

J.E.J. Plasmans and R.A. Van Straelen

December 1970

Contents

The General Linear Seemingly Unrelated Regression Problem

Part I Models and Inference

§ 1 Specification of the seemingly unrelated regression models.

Appendix

A Some properties of pseudo-inverses and estimable functions

B A statistical test on the significance of latent roots of a covariance matrix

Part II Feasible Aitken Estimation and an Application

§ 2 Feasible Aitken estimation of the specified SUR-models with unknown error variance-covariance matrix

2.1 The "classical" SUR-model: contemporaneously correlated disturbances and positive definite covariance matrix

2.2 Intertemporal correlation of disturbances and non-constancy of variances and covariances

2.21 First order autocorrelation

2.22 Heteroscedasticity

2.3 The general linear SUR-model with an error variance-covariance matrix and non stochastic X_i-matrices of arbitrary rank

2.4 Feasible Aitken estimation of autoregressive models

§ 3 An application: a stochastic model for the generation of input-output production coefficients.

Appendix

A Analysis of a classical SUR-two equation model
1. The role of dependency between different sets of explanatory variables w.r.t. efficiency
2. The efficiency of Aitken-relative to OLS-estimation concerning the unexplained variation
3. Exact finite sample properties of the feasible Aitken estimator in the classical SUR-model

B Properties of some statistical estimators regarding unbiasedness, consistency etc...
1. Finite sample bias of the ML-estimator of the serial correlation coefficient
2. The bias and inconsistency of the OLS-autoregression estimators in autoregressive models
2.1 First order autoregression and autocorrelation without exogenous variables
2.2 First order autoregression and autocorrelation with exogenous variables
2.3 An additional specification error: estimation of a 1st order i.s.o. a 2nd order model
3. The consistency and asymptotic (in) efficiency of the feasible Aitken estimator

C A likelihood ratio test on vector equality with error variance covariance matrix of arbitrary rank

II Feasible Aitken estimation of the specified SUR-models with unknown error variance-covariance matrix

In the first part of this paper, entitled "Models and Inference", various linear SUR-models were specified together with the statistical properties of the relating Aitken estimators in the case of known variance-covariance matrices of the disturbances.
If the error variance-covariance matrix is not a priori known, a "convenient" statistical estimator of this matrix has to be defined in order to obtain a "good" estimator of the unknown parameter vector β. Various statistical properties of such "two-round" estimators will be discussed in this section.

Definition 2.1

An Aitken estimator of β_i in model (1.1)(generalized least squares) or of β in model (1.2a)(seemingly unrelated regression), based on an initial consistent estimator of Ω_{ii} or of Ω, is called a feasible Aitken estimator of β_i, resp. β.

A feasible Aitken estimator of β in model(1.2a) may be derived by substituting the unknown Ω-matrix (1.3) by a first stage consistent, positive definite matrix $\hat{\Omega} = \{ \hat{\sigma}_{ij} \}$ or

(2.1) $\beta^{*} = (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}y$, with estimated variance-covariance matrix

(2.2) $V^{*}(\beta^{*}) = E^{*}[(\beta^{*}-\beta)(\beta^{*}-\beta)'] = (X'\hat{\Omega}^{-1}X)^{-1}$

(see also (1.6) and (1.7)).

Now, we shall show that the (feasible) Aitken estimators (2.1-2) are consistent , and, under certain conditions even unbiased, while, for each SUR-estimator specified, some other statistical properties will be briefly discussed in the subsequent paragraphs (*)

Theorem 2.1

If $E(\varepsilon) = 0$, $E(\varepsilon\varepsilon') = \Omega$, $\underset{T\to\infty}{\text{plim}} \left(\frac{X'\Omega^{-1}X}{T}\right)^{-1}$ (limit if X is non-stochastic) exists and is equal to the finite matrix V and if the columns of X_i are asymptotically independent of ε_i (with Ω non-stochastic), then $\tilde{\beta}$, β^*, $\tilde{V}(\tilde{\beta})$ and $V^*(\beta^*)$ are consistent estimators of β and $V(\tilde{\beta})$ (or $V(\beta^*)$).

Proof

1. $\tilde{\beta}$ is a consistent estimator of β , because from (1.6), the sampling error is:

(2.3) $\tilde{\beta}-\beta = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\varepsilon$ with probability limit (inconsistency):

(*) Since most statistical properties are well-known from literature: see P.Dhrymes [7], pp. 153-167, A.Zellner [26] and [28], A.Zellner and D.Huang [27], only the most appealing properties will be thoroughly studied. See also appendix A for an exact finite sample analysi of two-equation models and §1 for the BLU-property of the Aitken estimator in models with known error variance-covariance matrices.

$$(2.4)\quad \text{inc}\ (\tilde{\beta}) = \underset{T\to\infty}{\text{plim}}\ \tilde{\beta}-\beta = \underset{T\to\infty}{\text{plim}}\left(\frac{X'\Omega^{-1}X}{T}\right)^{-1} \underset{T\to\infty}{\text{plim}}\ \frac{1}{T}(X'\Omega^{-1}\varepsilon) \quad (*)$$

$$= V\ \underset{T\to\infty}{\text{plim}}\ \frac{1}{T}\ (X'\Omega^{-1}\varepsilon) = 0$$

which holds by suitable choice of $\Omega^{-1}=H'H$ (see (1.4) and (1.5)).

2. β^{*} is a consistent estimator of β, since from the corresponding sampling error (see (2.1) and (2.3)):

(2.6) $\beta^{*}-\beta = (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}\varepsilon$ and the consistency of $\hat{\Omega}$:

$$(2.7)\quad \underset{T\to\infty}{\text{plim}}\left(\frac{X'\hat{\Omega}^{-1}X}{T}\right)^{-1} = \underset{T\to\infty}{\text{plim}}\left(\frac{X'\Omega^{-1}X}{T}\right)^{-1} = V \ (**),$$

the sampling error (2.6) converges to zero in probability:

$$(2.8)\quad \text{inc}\ (\beta^{*}) = \underset{T\to\infty}{\text{plim}}\ \beta^{*}-\beta = V\ \underset{T\to\infty}{\text{plim}}\ \frac{X'\hat{\Omega}^{-1}\varepsilon}{T} = V\ \underset{T\to\infty}{\text{plim}}\ \frac{X'\Omega^{-1}\varepsilon}{T} = 0,$$

(*) Slutsky's theorem, H.Cramèr |6|, p.255. Note also that for X non stochastic the consistency follows from:

$$(2.5)\quad \underset{T\to\infty}{\text{plim}}\left(\frac{X'\Omega^{-1}X}{T}\right)^{-1} = \lim_{T\to\infty}\left(\frac{X'\Omega^{-1}X}{T}\right)^{-1} = V \text{ or}$$

$$\lim_{T\to\infty}\ (X'\Omega^{-1}X)^{-1} = V\ \lim_{T\to\infty}\frac{1}{T} = 0.$$

(**) See Slutsky's theorem again.

which is obtained by expanding $\hat{\Omega}^{-1}$ around some parameter value (see also (2.7) and further appendix B: the sole difficulties occur when X contains lagged dependent variables).

3. Since

$$(2.9)\quad \underset{T\to\infty}{\text{plim}}\ T(X'\hat{\Omega}^{-1}X)^{-1} = \underset{T\to\infty}{\text{plim}}\ T(X'\Omega^{-1}X)^{-1} = V \quad (*),$$

$\tilde{V}(\tilde{\beta})$ and $V^*(\beta^*)$ are consistent estimators of $V(\tilde{\beta})$ (or of $V(\beta^*)$). Δ

Theorem 2.2 (**)

The feasible Aitken estimator (2.1) is unbiased if the error vector $\varepsilon(t) = [\varepsilon_1(t), \varepsilon_2(t), \ldots, \varepsilon_M(t)]'$ follows an M-dimensional symmetric continuous probability distribution about zero for all t, provided that the mathematical expectation of β^* exists.

Proof

By the symmetry condition, the probability density function of $\varepsilon(t)$, say $f[\varepsilon(t)]$, satisfies $f[\varepsilon_1(t), \varepsilon_2(t), \ldots$ $\ldots, \varepsilon_M(t)] = f[-\varepsilon_1(t), -\varepsilon_2(t), \ldots, -\varepsilon_M(t)]$ and the sampling error (2.6), written as $\beta^*-\beta = C(\varepsilon)\varepsilon$, is an even function of ε because $\hat{\Omega}$ is an even function of ε (since

---

(*) See Slutsky's theorem again.

(**) See N. Kakwani [10], who discussed the classical SUR-model with contemporaneously correlated disturbances and a positive definite covariance matrix.

$\hat{\Omega}$ is invariant w.r.t.a change of sign of ε, i.e. if all elements of ε change sign) (*).

So, $C(\varepsilon) = C(-\varepsilon)$ or $\beta^{*}-\beta=\beta-\beta^{*}$ in probability, i.e. $\beta^{*}-\beta$ has the same probability density function as $\beta-\beta^{*}$, so that β^{*} is symmetrically distributed around the value β. Hence β^{*} is an unbiased estimator if its mathematical expectation exists. Δ

Remark 2.1

Notice that the feasible Aitken estimator is unbiased if the $\varepsilon(t)$-vectors are multivariately symmetrically distributed, although the first stage estimator of Ω is generally biased (but consistent).

(*) That $\hat{\Omega}$ is an even function of ε is readily verified for the classical SUR-model because then $\hat{\Omega}=\hat{\Sigma} \otimes I_T$ with

$$\hat{\Sigma}=\{\hat{\sigma}_{ij}\}=\{\frac{\hat{\varepsilon}_i'\hat{\varepsilon}_j}{T}\}=\{\frac{\varepsilon_i'Q_iQ_j\varepsilon_j}{T}\} \text{ and } Q_i=I_T-X_i(X_i'X_i)^{-1}X_i'$$

$(i,j=1,2,...,M)$.

2.1 Contemporaneously correlated disturbances and positive definite covariance matrix.

Under the assumptions that

- $E(\varepsilon)=0$ and $E(\varepsilon\varepsilon') = \Omega = \Sigma \otimes I_T$ (neither autocorrelation nor heteroscedasticity)
- the $\{X_i\}$ matrices $(i=1,2,...,M)$ are non stochastic (so, surely no lagged dep. variables),

some statistical properties, such as asymptotic probability distribution, efficiency, etc... of the estimators $\tilde{\beta}$ and β^*, with the latter being:

$$(2.10)\quad \beta^* = \left[X'(\hat{\Sigma}^{-1} \otimes I_T)\, X\right]^{-1} X' (\hat{\Sigma}^{-1} \otimes I_T) y \quad \text{or}$$

$$(2.11)\quad \begin{bmatrix} \beta_1^* \\ \beta_2^* \\ \vdots \\ \beta_M^* \end{bmatrix} = \begin{bmatrix} \hat{\sigma}^{11} X_1'X_1 & \hat{\sigma}^{12} X_1'X_2 & \cdots & \hat{\sigma}^{1M} X_1'X_M \\ \hat{\sigma}^{21} X_2'X_1 & \hat{\sigma}^{22} X_2'X_2 & \cdots & \hat{\sigma}^{2M} X_2'X_M \\ \vdots & \vdots & & \vdots \\ \hat{\sigma}^{M1} X_M'X_1 & \hat{\sigma}^{M2} X_M'X_2 & \cdots & \hat{\sigma}^{MM} X_M'X_M \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{M} \hat{\sigma}^{1i} X_1' y_i \\ \sum_{i=1}^{M} \hat{\sigma}^{2i} X_2' y_i \\ \vdots \\ \sum_{i=1}^{M} \hat{\sigma}^{Mi} X_M' y_i \end{bmatrix},$$

with $\hat{\Sigma} = \{\hat{\sigma}_{ij}\}$ a first round positive definite estimate of Σ, based on the OLS residual vectors $\hat{\varepsilon}_i, \hat{\varepsilon}_j$ $(i,j=1,2,...,M)$, will be studied.

Theorem 2.3

Consider model (1.2a) with

1. $E(\varepsilon\varepsilon') = \Omega = \Sigma \otimes I_T$

2. the matrix $\lim_{T\to\infty} \left(\frac{X'\Sigma^{-1}\otimes I_T X}{T}\right)^{-1} = V$ exists, is finite and non singular

3. the matrices $X_i (i=1,2,\ldots,M)$ are non stochastic

4. the error vectors $\varepsilon(t)=[\varepsilon_1(t), \varepsilon_2(t),\ldots,\varepsilon_M(t)]'$ are assumed to be mutually independent distributed with mean $E[\varepsilon(t)]=0$ and (constant) variance covariance matrix $E[\varepsilon(t)\ \varepsilon'(t)] = \Sigma = \{\sigma_{ij}\}$ $(\forall t)$ (*)

5. the matrices X_i are <u>uniformly bounded</u> and the error vectors satisfy for any $\eta>0$:

$$(2.12)\quad \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \int_{|\phi_t|>\eta\sqrt{T}} |\phi_t|^2 \, dF_t(\phi) = 0 \quad (**) ,$$

(*) Although all $\varepsilon(t)$ are assumed to be mutually independent with the same mean and covariance matrix, they need <u>not</u> be mutually independent and <u>identically</u> distributed!

(**) A sequence of stochastic variables z_t with corresponding distribution functions $F_t(.)$ converge <u>uniformly</u> in z_t if for some $r_o>o$:

$$(2.13)\quad \forall\eta>o\ \exists c\in Re\ \forall t: \int_{|\xi_t|\geq c} |\xi_t|^{r_o} dF_t(\xi)<\eta \qquad (\xi_t \text{ are convenient values of } z_t\ ) ,$$

i.e. if the tails of distribution about the r_o'th absolute moment are assumed to vanish. Condition (2.12) is commonly known as the <u>Lindeberg</u> condition and points to convergence saying that the terms $\frac{\phi_t}{\sqrt{T}}$ become uniformly small if T increases.

where $F_t(.)$ is the distribution function of the M-dimensional error vectors $\varepsilon(t)$,
then,
$\sqrt{T}(\beta^{\star}-\beta)$ and $\sqrt{T}(\tilde{\beta}-\beta)$ have the same asymptotic probability distribution, which is normal with mean zero and variance covariance matrix V.

Proof (⋆)

1. Asymptotically: $\sqrt{T}(\tilde{\beta}-\beta) \sim N(0,V)$

From the sampling error (2.3), we have to find the limit distribution of

$$(2.14)\quad \sqrt{T}(\tilde{\beta}-\beta)=\left(\frac{X'\Sigma^{-1}\otimes I_T X}{T}\right)^{-1}\frac{X'\Sigma^{-1}\otimes I_T\,\varepsilon}{\sqrt{T}},\ \text{and}$$

since the limit of $\left(\frac{X'\Sigma^{-1}\otimes I_T X}{T}\right)^{-1}$ exists, is finite and non-singular, we only have to bother with the asymptotic probability distribution of the vector

$\frac{X'\Sigma^{-1}\otimes I_T\,\varepsilon}{\sqrt{T}}$, which contains $\sum_{i=1}^{M} k_i = K$ elements.

(⋆) The proof of this theorem is based upon chapter 3 and pp. 161-167 of Prof.P.Dhrymes's book [7].

If we denote the t'th columm of X_i' by $p_i(t)$, we observe that:

$$(2.15)\quad \frac{X'\Sigma^{-1}\otimes I_T\varepsilon}{\sqrt{T}} = \frac{1}{\sqrt{T}} \begin{bmatrix} \sigma^{11}X_1' & \sigma^{12}X_1' \cdots\cdots & \sigma^{1M}X_1' \\ \sigma^{21}X_2' & \sigma^{22}X_2' \cdots\cdots & \sigma^{2M}X_2' \\ \vdots & \vdots & \vdots \\ \sigma^{M1}X_M' & \sigma^{M2}X_M' \cdots\cdots & \sigma^{MM}X_M' \end{bmatrix} \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_M \end{bmatrix}$$

$$= \frac{1}{\sqrt{T}} \begin{bmatrix} \sum_{i=1}^{M} \sum_{t=1}^{T} \sigma^{1i} p_1(t)\varepsilon_i(t) \\ \sum_{i=1}^{M} \sum_{t=1}^{T} \sigma^{2i} p_2(t)\varepsilon_i(t) \\ \vdots \\ \sum_{i=1}^{M} \sum_{t=1}^{T} \sigma^{Mi} p_M(t)\varepsilon_i(t) \end{bmatrix} = \frac{1}{\sqrt{T}} \begin{bmatrix} \sum_{t=1}^{T} p_1(t) \sum_{i=1}^{M} \sigma^{1i}\varepsilon_i(t) \\ \sum_{t=1}^{T} p_2(t) \sum_{i=1}^{M} \sigma^{2i}\varepsilon_i(t) \\ \vdots \\ \sum_{t=1}^{T} p_M(t) \sum_{i=1}^{M} \sigma^{Mi}\varepsilon_i(t) \end{bmatrix}$$

$$= \frac{1}{\sqrt{T}} \begin{bmatrix} \sum_{t=1}^{T} W_1(t) \\ \sum_{t=1}^{T} W_2(t) \\ \vdots \\ \sum_{t=1}^{T} W_M(t) \end{bmatrix} \varepsilon(t) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} W(t)\varepsilon(t) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} z(t)$$

where the sum signs have been interchanged (which is allowed since the σ_{ij}'s are finite and the matrices X_i are uniformly bounded)
and $W_i(t)$ are(k_ixM)-matrices $p_i(t)\sigma^{i\cdot}$ (i=1,2,...,M; t=1,2,...,T)

$\sigma^{i.}$ are the i'th rows of Σ^{-1}

$W(t)$ are $(K\times M)$-matrices $[W_1(t), W_2(t), \ldots, W_M(t)]'$ and

$z(t)$ are the K-dimensional vectors $W(t)\ \varepsilon(t)$ which are mutually independent (since the $\varepsilon(t)$ are mutually independent and the X_i non stochastic) with

(2.16) $E[z(t)] = E[W(t)\varepsilon(t)] = W(t)E[\varepsilon(t)] = 0$ and

(2.17) $E[z(t)z'(t)] = E[W(t)\varepsilon(t)\varepsilon'(t)W'(t)] = W(t)\Sigma W'(t)$ (*)

So, from the mutual independence of the $z(t)$ vectors and from (2.15)and(2.17):

(2.18) $\frac{1}{T}E[X'\Sigma^{-1}\otimes I_T\varepsilon\varepsilon'\Sigma^{-1}\otimes I_T X] = \frac{1}{T}X'\Sigma^{-1}\otimes I_T X = \frac{1}{T}\sum_{t=1}^{T} W(t)\Sigma W'(t)$

and we find that the sequence of K-dimensional vectors

(2.19) $z_T = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} z(t) = \frac{1}{\sqrt{T}}(X'\Sigma^{-1}\otimes I_T\varepsilon)$

converges to a random variable, say z, which is K- variate normally distributed with mean zero and variance covariance matrix

$\lim_{T\to\infty}\frac{1}{T}(X'\Sigma^{-1}\otimes I_T X)$, or asymptotically

---

(*) From which it is clear that, although it were assumed that the $\varepsilon(t)$ would be identically multivariately distributed (pay attention to footnote on p.9), the variance covariance matrices of the $z(t)$ vary over time, so that the $z(t)$ are <u>not identically</u> distributed.

$$(2.20)\quad \frac{X'\Sigma^{-1}\otimes I_T\varepsilon}{\sqrt{T}} \sim N(O,\ \lim_{T\to\infty} \frac{1}{T}(X'\Sigma^{-1}\otimes I_T X)),$$

because by the analogon of the Lindeberg condition (2.12) for the uniform convergence of the K-vectors z(t), after transforming to the univariate case:

$$(2.21)\quad r_T=\gamma' z_T = \sum_{t=1}^{T} \frac{\gamma' z(t)}{\sqrt{T}} = \sum_{t=1}^{T} q_t,$$

with γ an appropriate K-vector of real constants, we find, denoting the distribution function of q_t bij $G_t(.)$ and of z(t) by $\Psi_t(.)$, for $\forall\eta>o$:

$$(2.22)\quad \lim_{T\to\infty} \sum_{t=}^{T} \int_{|q_t|>\eta} |q_t|^2 dG_t(q) = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \int_{|\gamma' z(t)|>\eta\sqrt{T}} |\gamma' z(t)|^2 d\Psi_t(z)$$

$$\leq \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \int_{|\gamma||z(t)|\geq|\gamma' z(t)|>\eta\sqrt{T}} |\gamma|^2 |z(t)|^2 d\Psi_t(z) \quad \text{(Cauchy Schwarz)}$$

$$= \lim_{T\to\infty} \frac{1}{T}|\gamma|^2 \sum_{t=}^{T} \int_{|z(t)|>\eta_1\sqrt{T}} |z(t)|^2 \phi_t(z) dz(t) = 0 \quad \text{with } \eta_1 = \frac{\eta}{|\gamma|}$$

and the Lindeberg-Feller central limit theorem (*) for a sequence of mutually independent scalar random variables r_t with zero mean and variances $\frac{1}{T}\gamma'(X'\Sigma^{-1}\otimes I_T X)\gamma$ (T=1,2,...) may be applied , such that (2.20) is verified.

(*) See K.Chung [5], p. 187.

Finally, from (2.14) and (2.20), we obtain asymptotically:

(2.23) $\sqrt{T}(\tilde{\beta}-\beta) \sim N(0, \lim_{T\to\infty} \frac{1}{T} V(X'\Sigma \otimes I_T X)V) = N(0,V).$

2. Asymptotically: both $\sqrt{T}(\beta^*-\beta)$ and $\sqrt{T}(\tilde{\beta}-\beta) \sim N(0,V)$

If the sequence of vectors $\sqrt{T}(\beta^*-\beta)$ converges in probability (weak convergence) to a random variable, say b, then the corresponding distribution functions $F_{\beta^*_T}(.)$ (T=1,2,...) converge to the distribution function $F_b(.)$ of b (*), i.e. the vector sequence $\sqrt{T}(\beta^*-\beta)$ also converges to b in distribution. So, the asymptotic distribution of $\sqrt{T}(\beta^*-\beta)$ can be derived if its probability limit is evaluated and since:

(2.24) $$\operatorname*{plim}_{T\to\infty} \sqrt{T}(\beta^*-\beta) = \operatorname*{plim}_{T\to\infty}\left(\frac{X'\hat{\Sigma}^{-1}\otimes I_T X}{T}\right)^{-1} \operatorname*{plim}_{T\to\infty}\frac{X'\hat{\Sigma}^{-1}\otimes I_T \varepsilon}{\sqrt{T}} \quad \text{(Slutsky)}$$

$$= \lim_{T\to\infty}\left(\frac{X'\Sigma^{-1}\otimes I_T X}{T}\right)^{-1} \operatorname*{plim}_{T\to\infty}\frac{X'\Sigma^{-1}\otimes I_T \varepsilon}{\sqrt{T}} \quad (**)$$

$$= \operatorname*{plim}_{T\to\infty} \sqrt{T}\,(\tilde{\beta}-\beta)$$

or from part 1 of this theorem, both $\sqrt{T}(\beta^*-\beta)$ and $\sqrt{T}(\tilde{\beta}-\beta)$ have an identical asymptotic probability density which is N(0,V) Δ

(*) See M.Loève, [15], p. 168

(**) Since $\hat{\Sigma} = \{\hat{\sigma}_{ij}\} = \{\frac{\hat{\varepsilon}_i'\hat{\varepsilon}_j}{T}\}$ is consistent (see theorem 2.1)

Corollary 2.1

If the error vectors $\varepsilon(t)=[\varepsilon_1(t),\varepsilon_2(t),\varepsilon_3(t),\ldots,\varepsilon_M(t)]'$ $(t=1,2,\ldots,T)$ are mutually independent M-dimensional identically distributed and have non-vanishing finite variances (variance sums), then as $T\to\infty$:

$$(2.25)\quad \sqrt{T}(\beta^{*}-\beta) \text{ and } \sqrt{T}(\tilde{\beta}-\beta)\sim N(O,V)$$

$$\text{with } V = \lim_{T\to\infty}\left(\frac{X'\Sigma^{-1}\otimes I_T X}{T}\right)^{-1} \qquad (*)$$

Proof

If it is verified that the Lindeberg condition (2.12) is satisfied under the accompanying assumptions, the results of theorem 2.3 may be used to prove the conjecture (2.25) for large T.

From (2.22), we derive:

$$(2.26)\quad \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\int_{|z(t)|>\eta_1\sqrt{T}}|z(t)|^2 d\Psi_t(z)=\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\int_{|W(t)\varepsilon(t)|>\eta_1\sqrt{T}}|W(t)\varepsilon(t)|\, d\Psi_t(W\varepsilon)$$

(*) Notice that if the joint probability distribution of $\varepsilon(t)$ is M-dimensional normal with mean zero and variance-covariance matrix Σ (for all t), then the uniform convergence condition (2.12) is not needed at all because then for each sample size $T: \tilde{\beta}\sim N(\beta,(X'\Sigma^{-1}\otimes I_T X)^{-1})$ while β^{*} and $\tilde{\beta}$ have the asymptotic distribution (2.25) (see also properties of maximum likelihood estimators, e.g. in H.Cramèr [6], P.Dhrymes [7]).

$$\leq \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \int_{||W(t)||\ ||\varepsilon(t)|| \geq ||W(t)\varepsilon(t)|| > \eta_1\sqrt{T}} ||W(t)||^2\ ||\varepsilon(t)||^2\ d\Psi_t(W\varepsilon) = (\star)$$

$$= \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} tr(W'(t)W(t)) \int_{|\varepsilon(t)| > \eta_2\sqrt{T}} |\varepsilon(t)|^2\, dF_t(\varepsilon) \text{ with } \eta_2 = \frac{\eta_1}{\left[tr(W'(t)W(t))\right]^{\frac{1}{2}}}$$

$$\leq \lim_{T\to\infty} \frac{1}{T}\ MC^2 \sum_{t=1}^{T} \int_{|\varepsilon(t)| > \eta_3\sqrt{T}} tr(\varepsilon(t)\varepsilon'(t)) f_t(\varepsilon)\, d\varepsilon(t) \text{ with } \eta_3 = \frac{\eta_1}{C}$$

and C is the maximum length of any columm vector of W(t) (⋆⋆) we find from the assumed existence of finite variances or of the variance-covariance matrices of $\varepsilon(t)$:

(2.29) $\lim_{t\to\infty} \int_{\varepsilon(t)} tr\left[\varepsilon(t)\varepsilon'(t)\right]\ dF_t(\varepsilon) = tr\ \Sigma$

(⋆) Since:

(2.27) $|W(t)\varepsilon(t)|^2 = \varepsilon'(t)W'(t)W(t)\varepsilon(t) \leq tr(W'(t)W(t))tr(\varepsilon(t)\varepsilon'(t))$ or in general, the Euclidean(vector)norm of $\varepsilon(t)$ is <u>consistent</u> with the trace (matrix) norm of W(t):

(2.28) $||W(t)\varepsilon(t)|| = |W(t)\varepsilon(t)| \leq ||W(t)||\ ||\varepsilon(t)|| = \left[tr(W'(t)W(t))\right]^{\frac{1}{2}}$ $|\varepsilon(t)|$, where $||\ \ ||$ is the norm indication.

(⋆⋆) Since all elements of W(t) are assumed to be bounded and non stochastic, there should exist a positive number C which may be put equalto the maximum length of the vectors contained in all W(t).

and from the mutual independent and identical distributio of the $\varepsilon(t)$-vectors, by convenient (large) choice of η_1:

$$(2.30)\quad \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} |z(t)|^2 d\Psi_t(z) \le \lim MC^2 \int_{\varepsilon(t)>\eta_3\sqrt{T}} tr\left[\varepsilon(t)\varepsilon'(t)\right] f_t(\varepsilon)\, d\varepsilon(t)=0,$$

which is in accordance with the Lindeberg condition (2.12) so that the results of theorem (2.3) may be applied since all assumptions for it are satisfied. Δ

Theorem 2.4

The Aitken estimator (1.6) with $\Omega=\Sigma\otimes I_T$ is efficient w.r.t. OLS unless $\sigma_{ij}=0$ for all $i\neq j$ or all $X_i\,(i=1,2,\ldots,M)$ are equal.

Proof

This theorem can easily be proved from theorem 1.1 and corollary 1.1. Indeed, substituting $\Omega_{ij}=\sigma_{ij}I_T=0$ or $X_1=X_2=\ldots\ldots=X_M$ and application of the above theorem and corollary show that $\tilde{\beta}$ is efficient w.r.t. $\tilde{\beta}$ unless both exceptions are satisfied, in which case corollary 1.1 demonstrates that $\tilde{\beta}=\hat{\beta}$.

This may also be indicated for the underlying contemporaneously correlated SUR-model as follows (*).

Denote the ratio of the generalized variance of the Aitken estimator (1.6) w.r.t. that of the OLS-estimator (1.17) as:

$$(2.31)\quad \alpha=\frac{|(X'\Omega^{-1}X)^{-1}|}{|(X'\Phi^{-1}X)^{-1}|}$$, where from theorem 1.1 $0\le\alpha\le1$.

(*) See also A,Zellner. D.Huang [27], pp. 306-307

Using a general determinantal inequality (*) (see def.(2.11):

$$(2.32)\quad |X'\Omega^{-1}X| = |X'\Sigma^{-1}\otimes I_T X| \leq |X_1'X_1|\sigma^{11}.|X_2'X_2|\sigma^{22}.....|X_M'X_M|\ \sigma^{MM},$$

where equality only holds if $\forall_{\substack{i,j\\ i\neq j}}\ \sigma_{ij}=0$ or $X_i'X_j=0$,

i.e. if contemporaneous disturbance terms in different equations are uncorrelated or if the explanatory variables of different equations are mutually orthogonal.
Combining (2.31)and(2.32):

$$(2.33)\quad \frac{1}{\alpha} = |X'\Sigma^{-1}\otimes I_T X|\ \ |(X'\Phi^{-1}X)^{-1}|$$

$$= |X'\Sigma^{-1}\otimes I_T X|\sigma_{11}|(X_1'X_1)^{-1}|\sigma_{22}|(X_2'X_2)^{-1}|.....\sigma_{MM}|(X_M'X_M)^{-1}|$$

$$\leq |\sigma_{11}\sigma^{11}I_{k_1}|\ \ |\sigma_{22}\sigma^{22}I_{k_2}\ \ |.....|\sigma_{MM}\sigma^{MM}I_{k_M}|\quad \text{or}$$

$$(2.34)\quad (\sigma_{11}\sigma^{11})^{-k_1}(\sigma_{22}\sigma^{22})^{-k_2}......(\sigma_{MM}\sigma^{MM})^{-k_M}\ \ \leq\alpha\leq 1$$

from which it is clear that the l.h.s. becomes unity if $\sigma_{ij}=0$ for $\forall_{i\neq j}$ (then Aitken estimator is OLS estimator) Δ

(*) See R.Bellman [4], p.127: for any (n×n) matrix A, the general inequality: (2.32) $|A| \leq \prod_{i=1}^{n} a_{ii}$ holds.

There is only equality if $a_{ij}=0$ for $\forall_{\substack{i,j\\ i\neq j}}$.

Remark 2.2

1. Since the expression on the l.h.s. of (2.34) represents the maximal gain that can be realized, we find that the maximal gain in applying Aitken's estimator w.r.t. OLS-estimator occurs when the disturbances of different equations are strongly correlated and when the explanatory variables in different equations are really orthogonal. (*)

2. When considering feasible Aitken estimator (2.10), the results on efficiency hold only asymptotically or for large T:

(2.35) $|(X'\hat{\Sigma}^{-1}\otimes I_T X)|^{-1} \leq |(X'\hat{\phi}^{-1}X)^{-1}|$ and because of the consistency of $\hat{\Sigma}$:

(2.36) $$\operatorname*{plim}_{T\to\infty} \left|\left(\frac{X'\hat{\Sigma}^{-1}\otimes I_T X}{T}\right)^{-1}\right| \leq \operatorname*{plim}_{T\to\infty} \left|\left(\frac{X'\hat{\phi}^{-1}X}{T}\right)^{-1}\right| \quad (**)$$

Proposition 2.1

An unbiased first round estimator for the variance covariance matrix Σ (and hence for Ω) is given by

(2.37) $$\hat{\Sigma}=\{\hat{\sigma}_{ij}\} =\{\frac{\hat{\varepsilon}_i'\hat{\varepsilon}_j}{T-k_i-k_j+k_i r_H^2}\} ,$$

Where $r_H^2 = \sum_{i=1}^{k_i} r_i^2$, $k_i \leq k_j$ and r_H^2 is Hooper's trace corre-

---

(*) See appendix A for a complete 2-equation analysis and for the effect of intercorrelation between X_1 and X_2.

(**) See, however, appendix A for a 2-equation model.

lation and r_i^2 are the squared canonical correlation coefficients (*)

Proof

(2.38) $E(\hat{\varepsilon}_i'\hat{\varepsilon}_j)=E\left[\varepsilon_i'(I_T-X_i(X_i'X_i)^{-1}X_i')(I_T-X_j(X_j'X_j)^{-1}X_j')\varepsilon_j\right](i=1,2,\dots,M)$

$$=\sigma_{ij}tr\left[I_T-X_i(X_i'X_i)^{-1}X_i'-X_j(X_j'X_j)X_j'+X_i(X_i'X_i)^{-1}X_i'X_j(X_j'X_j)^{-1}X_j'\right]$$

$$=\sigma_{ij}\left[T-k_i-k_j+tr(X_i'X_i)^{-1}X_i'X_j(X_j'X_j)^{-1}X_j'X_i\right](k_i\leq k_j)\quad ,$$

where the last matrix between[ ] has k_i eigenvalues $\lambda_i=r_i^2$ or (2.38) becomes:

(2.39) $E(\hat{\varepsilon}_i'\hat{\varepsilon}_j)=\sigma_{ij}(T-k_i-k_j+\sum_{i=1}^{k_i} r_i^2)=\sigma_{ij}(T-k_i-k_j+k_i\ r_H^2)\quad ,$

Since for $i=j$, $r_H^2=1$, (2.37) provides an unbiased estimate of Σ, and hence of Ω (**) Δ

2.2 Intertemporal correlation of disturbances and non-constancy of variances and covariances.

2.21 First order autocorrelation

(i) Model I (A1)

A feasible Aitken estimator of the parameter vector β

(*) See A.Zellner and D.Huang [27] pp 308-309. This theorem might be interesting when the error vectors $\varepsilon(t)$ are not T-dimensionally symmetrically distributed, and an unbiased estimator is still desired.

(**) If the explanatory variables in the i'th and j'th equations are the same, then $X_i'X_j=X_i'X_i$ and $r_H^2=1$ so that the denominator in (2.37) becomes $T-k_i$.
If, an the contrary, the explanatory variables in the i'th and j'th equations are mutually orthogonal, then $X_i'X_j=0$ and $r_H^2=0$ so that the denominator becomes $T-k_i-k_j$.

in model (1.34), where the disturbances $\varepsilon_i(t)$ are assumed to be both contemporaneously and serially correlated (by first order autoregressive scheme (1.28) and assumptions (1.28-31)), can be derived by the following three step procedure:

1. Estimate the parameter vectors β_i of equations (1.33) by OLS to obtain the consistent estimates:

$$(2.40)\quad \hat{\varepsilon}_i = y_i - X_i\hat{\beta}_i \quad \text{with } \hat{\beta}_i = (X_i'X_i)^{-1}X_i'y_i \quad (i=1,2,\ldots,M)$$

with consistent estimate of the autoregressive parameters ρ_i (see app.B1 with $\varepsilon_i(o)$ either zero or stochastic:(B5) and (B.42)):

$$(2.41)\quad \hat{\rho}_i = \frac{\sum_{t=1}^{T} \hat{\varepsilon}_i(t)\hat{\varepsilon}_i(t-1)}{\sum_{t=2}^{T} \hat{\varepsilon}_i^2(t-1)}$$

2. Obtain a consistent estimate of the contemporaneous covariance matrix Σ by substituting the consistent estimator (2.41) into (1.32) to compute:

$$(2.42)\quad \hat{P}_i^{-1} = \begin{bmatrix} (1-\hat{\rho}_i^2)^{\frac{1}{2}} & 0 & 0 & \cdots & 0 \\ -\hat{\rho}_i & 1 & 0 & \cdots & 0 \\ 0 & -\hat{\rho}_i & 1 & \cdots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & \cdots -\hat{\rho}_i & 1 \end{bmatrix} = \hat{R}_i$$

and apply OLS an the transformed equations (see (1.33)):

(2.43) $\hat{R}_i y_i = \hat{R}_i X_i \beta_i + \hat{R}_i P_i \eta_i$ $\qquad$ (i =1,2,...,M)

to yield estimates of the elements σ_{ij} of Σ:

$$(2.44)\quad \hat{\sigma}_{ij} = \frac{(\hat{R}_i y_i - \hat{R}_i X_i \hat{\hat{\beta}}_i)'(\hat{R}_i y_i - \hat{R}_i X_i \hat{\hat{\beta}}_i)}{T} \qquad (\star)$$

with $\hat{\hat{\beta}}_i$ the OLS-estimator of β_i in eq.(2.43).

Substituting (2.41) into (1.32) and (1.35) and (2.44) into (1.37), we obtain an estimate of the variance covariance matrix Ω as:

$$(2.45)\quad \hat{\Omega} = \hat{P}(\hat{\Sigma} \otimes I_T)\hat{P}'$$

where $\hat{P}$ is the (MT×MT)- block diagonal matrix of the $\hat{P}_i$'s and $\hat{\Sigma} = \{\hat{\sigma}_{ij}\}$.

3. Finally, the vector β in model (1.34) is estimated as:

$$(2.46)\beta^{\star} = (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}y = (X'\hat{P}'^{-1}\hat{\Sigma}^{-1} \otimes I_T \hat{P}^{-1}X)^{-1}X'\hat{P}'^{-1}\hat{\Sigma}^{-1} \otimes I_T \hat{P}^{-1}y.$$

Proposition 2.2

The estimated variance-covariance matrix $\hat{\Omega}$ in (2.45) is a consistent estimator for Ω and hence $\beta^{\star}$ in (2.46) is a consistent estimator for β.

Proof

Since the X_i's are assumed to be non stochastic, $\hat{\rho}_i$ and $\hat{P}_i$ are consistent estimates of ρ_i and P_i and by Slutsky's theorem:

($\star$) The denominator of (2.44) might also be $(T-k_i)^{\frac{1}{2}}(T-k_j)^{\frac{1}{2}}$ to consider, if desired, finite sample effects, but since, usually, only a consistent estimator of Σ is required, (2.44) will equally do.Notice also that, if

$$\hat{\rho}_i = \frac{\sum_{t=1}^{T} \hat{\varepsilon}_i(t)\hat{\varepsilon}_i(t-1)}{\sum_{t=2}^{T} \hat{\varepsilon}_i^2(t-1)} + \frac{k_i}{T},\ \hat{R}_i P_i \simeq I_T.$$

$$(2.47)\quad \underset{T\to\infty}{\text{plim}}\ \hat{R} = \underset{T\to\infty}{\text{plim}}\ \hat{P}^{-1} = \left[\underset{T\to\infty}{\text{plim}}\ \hat{P}\right]^{-1} = P^{-1} = R$$

and the estimated residuals $\hat{\hat{\eta}}_i$ of the transformed equations (2.43) may be written as:

$$(2.48)\quad \hat{\hat{\eta}}_i = (\hat{R}_i y_i - \hat{R}_i X_i \hat{\hat{\beta}}_i) = \left[I_T - \hat{R}_i X_i (X_i' \hat{R}_i' \hat{R}_i X_i)^{-1} X_i' \hat{R}_i'\right]\eta_i = (I_T - M_i)\eta_i$$

$$= Q_i \eta_i,$$

the probability limit of $\hat{\sigma}_{ij}$ can be written as:

$$(2.49)\quad \underset{T\to\infty}{\text{plim}}\ \hat{\sigma}_{ij} = \underset{T\to\infty}{\text{plim}}\left[\frac{\eta_i'(I_T - M_i)'(I_T - M_j)\eta_j}{T}\right]$$

$$= \underset{T\to\infty}{\text{plim}}\ \frac{\eta_i'\eta_j}{T} - \underset{T\to\infty}{\text{plim}}\ \frac{\eta_i' M_i \eta_j}{T} - \underset{T\to\infty}{\text{plim}}\ \frac{\eta_i' M_j \eta_j}{T} + \underset{T\to\infty}{\text{plim}}\ \frac{\eta_i' M_i M_j \eta_j}{T}$$

$$= \sigma_{ij} - \underset{T\to\infty}{\text{plim}}\ \frac{\eta_i'\hat{R}_i X_i}{T}\ \underset{T\to\infty}{\text{plim}}\left(\frac{X_i'\hat{R}_i'\hat{R}_i X_i}{T}\right)^{-1} \underset{T\to\infty}{\text{plim}}\ \frac{X_i'\hat{R}_i'\eta_j}{T}$$

$$- \underset{T\to\infty}{\text{plim}}\ \frac{\eta_i'\hat{R}_j X_j}{T}\ \underset{T\to\infty}{\text{plim}}\left(\frac{X_j'\hat{R}_j'\hat{R}_j X_j}{T}\right)^{-1} \underset{T\to\infty}{\text{plim}}\ \frac{X_j'\hat{R}_j'\eta_j}{T}$$

$$+ \underset{T\to\infty}{\text{plim}}\ \frac{\eta_i'\hat{R}_i X_i}{T}\ \underset{T\to\infty}{\text{plim}}\left(\frac{X_i'\hat{R}_i'\hat{R}_i X_i}{T}\right)^{-1} \underset{T\to\infty}{\text{plim}}\ \frac{X_i'\hat{R}_i'\hat{R}_j X_j}{T}$$

$$\underset{T\to\infty}{\text{plim}}\left(\frac{X_j'\hat{R}_j'\hat{R}_j X_j}{T}\right)^{-1} \underset{T\to\infty}{\text{plim}}\ \frac{X_j'\hat{R}_j'\eta_j}{T} \quad \text{(Slutsky)}$$

$$= \sigma_{ij}$$

and, again using Slutsky's theorem,

$$(2.50)\quad \operatorname*{plim}_{T\to\infty} \hat{\Omega} = \operatorname*{plim}_{T\to\infty}\left[\hat{P}(\hat{\Sigma}\otimes I_T)\ \hat{P}'\right] = \operatorname*{plim}_{T\to\infty} \hat{P}\ \operatorname*{plim}_{T\to\infty} (\hat{\Sigma}\otimes I_T)\ \operatorname*{plim}_{T\to\infty} \hat{P}'$$

$$= P\ \{\operatorname*{plim}_{T\to\infty} \hat{\sigma}_{ij}\}\ \otimes I_T\ P' = P\Sigma\otimes I_T P'.$$

By definition 2.1,(2.46) is a feasible Aitken estimator for β and by theorem 2.1, β^* is consistent, or, once more by Slutsky's theorem:

$$(2.51)\quad \operatorname*{plim}_{T\to\infty} \beta^* = \operatorname*{plim}_{T\to\infty}\left(\frac{X'\hat{\Omega}^{-1}X}{T}\right)^{-1} \operatorname*{plim}_{T\to\infty} \frac{X'\hat{\Omega}^{-1}y}{T}$$

$$= \left(\frac{X'\operatorname*{plim}_{T\to\infty} \hat{\Omega}^{-1}X}{T}\right)^{-1} \operatorname*{plim}_{T\to\infty} \frac{X'\operatorname*{plim}_{T\to\infty} \hat{\Omega}^{-1}(X\beta+\varepsilon)}{T}$$

$$= \beta + \left(\frac{X'\Omega^{-1}X}{T}\right)^{-1} \operatorname*{plim}_{T\to\infty} \frac{X'\Omega^{-1}\varepsilon}{T} = \beta. \qquad \Delta$$

<u>Note</u> Since $\hat{\Sigma}$ and hence $\hat{\Omega}$ is an even function of η (see (2.49) and (2.50)) and, therefore, also of ε, β^* is in general an unbiased estimator of β because of theorem 2.2

<u>Definition 2.2</u>

A sequence $\{x(t)\}$ of random vectors is called <u>n-dependent</u> if there exists a non negative integer n such that any finite subset $\{x(t_1), x(t_2),\dots,x(t_p)\}$ is stochastically independent of any other subset of vectors $\{x(\tau_1),x(\tau_2),\dots,x(\tau_q)\}$, provided the index sets $\{t_i\}_{i=1,2,\dots,p}$ and $\{\tau_j\}_{j=1,2,\dots,q}$

are chosen so as to satisfy $\min_{i=1,2,\ldots,p}\{t_i\}-\max_{j=1,2,\ldots,q}\{\tau_j\}>n$ (*).

Theorem 2.5

Consider model (1.34) with

1. $E(\varepsilon\varepsilon')=\Omega=PE(\eta\eta')P'=P\Sigma\otimes I_T P'$

2. the matrix $\lim_{T\to\infty}\left(\frac{X'\Omega^{-1}X}{T}\right)^{-1}=V$ exists and is positive definite

3. the matrices X_i $(i=1,2,\ldots,M)$ are non stochastic

4. the error vectors $\eta(t)=\left[\eta_1(t),\eta_2(t),\ldots,\eta_M(t)\right]'$ are assumed to be mutually independent with $E\left[\eta(t)\right]=0$ and $E\left[\eta(t)\eta'(t)\right]=\Sigma$ while the error vectors $\varepsilon(t)=\left[\varepsilon_1(t),\varepsilon_2(t)\ldots,\varepsilon_M(t)\right]'(t=1,2,\ldots,T)$ are allowed to be n-dependent for any $T=1,2,\ldots$ with zero first order and finite second order moments.
5. the matrices X_i are uniformly bounded such that the Lindeberg condition for finite second order moments of the $\varepsilon(t)$ implies:

$$(2.52)\quad \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\int_{|\phi_t|>\delta\sqrt{T}}|\phi_t|^2 dF_t(\phi)=0 \text{ is satisfied for any } \delta>0,$$

then, $\sqrt{T}(\beta^*-\beta)$ and $\sqrt{T}(\tilde{\beta}-\beta)$, β^* being the feasible and $\tilde{\beta}$ the usual Aitken estimator of β in the autocorrelated model, have the same asymptotic normal distribution with mean zero and variance covariance matrix V.

Proof 1. Asymptotically $\sqrt{T}(\tilde{\beta}-\beta)\sim N(0,V)$

From theorem 2.3 (2.15), the $(K\times 1)$ vector $\frac{X'\Omega^{-1}\varepsilon}{\sqrt{T}}$ can be written as:

(*) Or simply, if q-p>n implies that the two sets $\{x(1), x(2),\ldots x(p)\}$ and $\{x(q), x(q+1),\ldots,x(T)\}$ are stochastically independent, the sequence $\{x(t)\}$ is said to be n-dependent.

$$(2.53)\quad \frac{X'\Omega^{-1}\varepsilon}{\sqrt{T}} = \frac{X'P'^{-1}(\Sigma^{-1}\otimes I_T)P^{-1}\varepsilon}{\sqrt{T}} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} A(t)\varepsilon(t)$$

$$= \frac{1}{\sqrt{T}}\sum_{t=1}^{T} s(t) = s_T,$$

where the $(K \times M)$-matrices $A(t)$ $(t=1,2,\ldots,T;\ T=1,2,\ldots)$ are non stochastic satisfying the following inequality, due to the boundedness condition of X_i and Σ:

$$(2.54)\quad ||A(t)|| \leq C < \infty$$

where C is a positive constant and $||\ ||$ may indicate any "consistent" matrix norm for all $A(t)$, such as e.g. the maximum of all absolute values of the elements of $A(t)$. Since, in general, the error vectors $\varepsilon(t)$ are assumed to be n-dependent for any sample size T, with mean $E\left[\varepsilon(t)\right]=0$ and equal variance-covariance matrix for any t:

$$(2.55)\quad \begin{aligned} E\left[\varepsilon(t)\varepsilon'(t+\tau)\right] &= \Phi_\tau \quad \text{if } |\tau|\leq n \\ &= 0 \quad \text{if } |\tau|>n \end{aligned}$$

the first and second order moments of the mutually dependent s_T-vectors $(T=1,2,\ldots)$ are computed as:

$$(2.56)\quad \frac{E(X'\Omega^{-1}\varepsilon)}{\sqrt{T}} = E(s_T) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} A(t)E\left[\varepsilon(t)\right] = 0 \quad \text{and}$$

$$(2.57)\quad \frac{1}{T}E\left[X'\Omega^{-1}\varepsilon\varepsilon'\Omega^{-1}X\right] = \frac{1}{T}(X'\Omega^{-1}X)$$

$$= \frac{1}{T}\left\{\sum_{t=1}^{T} A(t)\Phi_o A'(t) + \sum_{\tau=1}^{\theta}\sum_{t=1}^{T-\tau}\left[A(t)\Phi_\tau A'(t+\tau)+(A(t)\Phi_\tau A'(t+\tau))'\right]\right\}$$

with $\theta = \min(n, T-1)$,

from which it is easly verified that if the $\varepsilon(t)$ were mutually independent distributed, the result (2.18) would be obtained with $A(t) = W(t)$ and $\Phi_o = \Sigma$. (*)

The composite 2nd term on the r h s of (2.57) specifies the covariance structure between the dependent random vectors $\{\varepsilon(t), \varepsilon(t+\tau)\}$.

The sequence of K-dimensional vectors $s_T = \frac{X'\Omega^{-1}\varepsilon}{\sqrt{T}}$ converges now to a random variable, say s, which is <u>normally</u> distributed with mean zero and variance covariance matrix

$\lim_{T\to\infty} \frac{1}{T}(X'\Omega^{-1}X)$, the proof of which follows similar lines as outlined in theorem 2.3, where this time a central limit theorem for dependent univariate random variables has to be applied. (**)

It develops along following ideas:
Reducing to the univariate case with a scalar vector λ:

$$(2.58) \quad \xi_T = \lambda' s_T = \sum_{t=1}^{T} \frac{\lambda' s(t)}{\sqrt{T}} = \sum_{t=1}^{T} \xi(t),$$

we may partition the observations $\xi_1, \xi_2, \xi_3, \ldots, \xi_T$, whose partial sums are stochastically independent (hence also for $s_1 s_2, \ldots, s_T$ defined e.g. as

$\frac{1}{\sqrt{T}} \sum_{t=t'}^{t''} s(t) = \frac{1}{\sqrt{T}} \sum_{t=t'}^{t''} A(t)\varepsilon(t)$), performed under, say,

additive representation such as:

$$(2.59) \quad s_T = u_{Tk} + v_{Tk}, \qquad T=1,2,3,\ldots \quad k=1,2,3,\ldots,K_T \ (K_T \to \infty \text{ as } T\to\infty)$$

(*) Hence, zero dependence, i.e. $n=0$ and so $\theta = o$, is equivalent to independence.

(**) See e.g. W. Hoeffding and H. Robbins [9], theorems 1-3 pp. 774-776.

where the u_{Tk} are stochastically independent variables with zero mean and finite variance for each element so that the analogon of the Lindeberg condition (2.52) may be utilized on the u_{Tk} or also on the independent parts of ξ (t)(see 2.22).

Hence, asymptotically

(2.60) $\sqrt{T}(\tilde{\beta}-\beta) \sim N(0,V)$ (*)

2. Asymptotically: both $\sqrt{T}(\beta^{*}-\beta)$ and $\sqrt{T}(\tilde{\beta}-\beta)$ are N(0,V)

Since from lemma 2.1:

$$(2.62)\quad \underset{T\to\infty}{\text{plim}}\ \sqrt{T}(\beta^{*}-\beta)=\underset{T\to\infty}{\text{plim}}\left(\frac{X'\hat{\Omega}^{-1}X}{T}\right)^{-1}\underset{T\to\infty}{\text{plim}}\ \frac{X'\hat{\Omega}^{-1}\varepsilon}{\sqrt{T}}$$

$$=\lim_{T\to\infty}\left(\frac{X'\Omega^{-1}X}{T}\right)^{-1}\underset{T\to\infty}{\text{plim}}\ \frac{X'(\underset{T\to\infty}{\text{plim}}\ \hat{\Omega}^{-1})\varepsilon}{\sqrt{T}} =$$

$$=\lim_{T\to\infty}\left(\frac{X'\Omega^{-1}X}{T}\right)^{-1}\underset{T\to\infty}{\text{plim}}\ \frac{X'\Omega^{-1}\varepsilon}{\sqrt{T}} = \underset{T\to\infty}{\text{plim}}\ \sqrt{T}(\tilde{\beta}-\beta)$$

and from part 1 of this theorem, both $\sqrt{T}(\beta^{*}-\beta)$ and $\sqrt{T}(\tilde{\beta}-\beta)$ have the same limiting normal distribution with zero mean and $V=\lim_{T\to\infty}\left(\frac{X'\Omega^{-1}X}{T}\right)^{-1}$ as covariance matrix. Δ

(ii) Model II (A2)

This model, guaranteeing the time invariancy of both variances and covariances σ_{ij}, may be estimated in the same way as the parameters of model A1 are estimated.

(*) It is a conjecture of us that the property (2.60) can also directly be derived from (2.53) as:

$$(2.61)\quad \frac{X'\Omega^{-1}\varepsilon}{\sqrt{T}} = \frac{X'P'^{-1}(\Sigma^{-1}\otimes I_T)\eta}{\sqrt{T}} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} B(t)\eta(t),\ \text{with}\ \eta(t)$$

M-dimensional mutually independent random vectors (see theorem 2.3).The sole remaining difficulty is to specify B(t).

There is only one slight complication since premultiplication of system (1.34) by $R=P^{-1}$ does not reduce the transformed system:

(2.63) $Ry = R X \beta + \eta$

to the classical S U R-model with variance covariance matrix $\Sigma \otimes I_T$ because the off-diagonal blocks of the new covariance matrix $E(\eta\eta')$ have the slightly altered form (1.43) instead of $\sigma_{ij} I_T$.

Therefore, the estimation procedure runs like:

1. Estimation of the autoregressive parameters ρ_i by OLS as in (2.41);
2. The variance elements σ_{ii} are estimated from OLS on the transformed equations (2.43) as in (2.44), while the covariance elements σ_{ij} $(i \neq j)$ are estimated from the estimated residuals of the transformed eq. (2.43), with the modification that the first element of each residual vector is discarded. Then each covariance block $E(\eta_i \eta_j')$ $(i \neq j)$ is estimated as:

(2.64) $$\hat{\sigma}_{ij} \begin{bmatrix} \dfrac{(1-\hat{\rho}_i^2)^{\frac{1}{2}}(1-\hat{\rho}_j^2)^{\frac{1}{2}}}{1-\hat{\rho}_i\hat{\rho}_j} & 0 \\ 0 & I_{T-1} \end{bmatrix}$$ and a consistent estimate

of Ω is directly obtained as $\hat{\Omega} = \hat{P}\, E(\hat{\eta\eta'})\, \hat{P}'$.
3. Obtain a feasible estimator β^* as in (2.46).

(iii) Model III (A3)

This model is estimated in a similar way as model A1. The classical SUR-model is obtained for (T-1) observations. So, given consistent estimates for ρ_i (say by (2.41)), consistent estimators of Ω and β are easily obtained utilising ((T-1)xT)-transformation matrices $\hat{R}_i^*$ (see (1.45)).

Note Most statistical properties, presented for model A1, equally apply on models A2 and A3.

2.22 Heteroscedasticity

(i) Model I (H1)

A feasible Aitken estimator is obtained by the following three step procedure.

1. Apply OLS on equations (1.25) yielding $\hat{y}_i = X_i\hat{\beta}_i$ $(i=1,2,\ldots,M)$
2. Apply classical SUR-estimation on the transformed system (1.52) i.e. on:

$$(2.65)\quad \hat{S}y = \hat{S}X\beta + u \qquad \text{with } E(uu') = \Sigma \otimes I_T,$$

$$\text{where } \hat{S} = \begin{bmatrix} \hat{Q}_1^{-1} & 0 & \cdots & 0 \\ 0 & \hat{Q}_2^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \hat{Q}_M^{-1} \end{bmatrix} \quad \text{and } \hat{Q}_i = (\text{diag } \hat{y}_i)^{\frac{r_i}{2}} = \hat{S}_i^{-1} (i=1,2,\ldots,M),\text{(*)}$$

i.e., in first instance, the σ_{ij}'s are estimated as:

$$(2.66)\quad \hat{\sigma}_{ij} = \frac{\hat{u}_i'\hat{u}_j}{T} = \frac{(\hat{S}_i y_i - \hat{S}_i X_i \hat{\hat{\beta}}_i)'(\hat{S}_j y_j - \hat{S}_j X_j \hat{\hat{\beta}}_j)}{T} \quad \text{so that } \Omega \text{ is}$$

estimated as:

$$(2.67)\quad \hat{\Omega} = \hat{Q}\,(\hat{\Sigma} \otimes I_T)\,\hat{Q}$$

3. Finally, the β-vector in model (1.52) is estimated as:

$$(2.68)\quad \beta^{*} = (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}y = (X'\hat{Q}^{-1}\hat{\Sigma}^{-1} \otimes I_T\hat{Q}^{-1}X)^{-1}X'\hat{Q}^{-1}\hat{\Sigma}^{-1} \otimes I_T\hat{Q}^{-1}y$$

(*) We considered $\hat{y}_i$ since, in the case of heteroscedasticity, the OLS-estimator of β remains unbiased.

Theorem 2.6

β^* is (generally) an unbiased and consistent estimator of β, and $\sqrt{T}(\beta^*-\beta)$ has the same limiting normal distribution as $\sqrt{T}(\tilde{\beta}-\beta)$ with mean 0 and variance covariance matrix

$$\lim_{T\to\infty} \left(\frac{X'\Omega^{-1}X}{T}\right)^{-1} = V.$$

Proof

Under general conditions, $\varepsilon(t)$ has a multivariate symmetric and continuous pdf so that from theorem (2.2) β^* is an unbiased estimator of β.

β^* is a consistent estimator of β, because in (2.68), $\hat{\Omega}^{-1}$ is a consistent estimator of Ω which is proved as follows:

$$(2.69)\quad \underset{T\to\infty}{\text{plim}}\ \hat{\sigma}_{ij} = \underset{T\to\infty}{\text{plim}} \frac{u_i'u_j}{T} - \underset{T\to\infty}{\text{plim}} \left(\frac{u_i'\hat{S}_iX_i}{T}\right)\left(\frac{X_i'\hat{S}_i\hat{S}_iX_i}{T}\right)^{-1} \frac{X_i'\hat{S}_iu_j}{T}$$

$$- \underset{T\to\infty}{\text{plim}} \frac{u_i'\hat{S}_jX_j}{T} \left(\frac{X_j'\hat{S}_j\hat{S}_jX_j}{T}\right)^{-1} \frac{X_j'\hat{S}_ju_j}{T} + \underset{T\to\infty}{\text{plim}} \frac{u_i'\hat{S}_iX_i}{T}$$

$$\left(\frac{X_i'\hat{S}_i\hat{S}_iX_i}{T}\right)^{-1} \left(\frac{X_i'\hat{S}_i\hat{S}_jX_j}{T}\right)\left(\frac{X_j'\hat{S}_j\hat{S}_jX_j}{T}\right)^{-1}\left(\frac{X_j'\hat{S}_j\eta_j}{T}\right),$$

and since $\hat{\beta}_i$ is a consistent estimator for β_i and applying Slutsky's theorem:

$$(2.70)\quad \underset{T\to\infty}{\text{plim}}\ \hat{\sigma}_{ij} = \sigma_{ij} \quad \text{or}$$

$$(2.71)\quad \underset{T\to\infty}{\text{plim}}\ \hat{\Omega} = \underset{T\to\infty}{\text{plim}} (\hat{Q}\hat{\Sigma} \otimes I_T\hat{Q}) = Q\Sigma \otimes I_T Q = \Omega \quad \text{and}$$

$$(2.72)\quad \underset{T\to\infty}{\text{plim}}\ \beta^* = \beta.$$

Following arguments similar to those set forth in the proof of theorem 2.3 (but with matrices with different contents):

$$(2.73)\quad \frac{X'\Omega^{-1}\varepsilon}{\sqrt{T}} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} W(t)\varepsilon(t) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} z(t) = z_T$$

converges to a random variable, say z, which is K-variate normally distributed with mean zero and variance-covariance matrix:

$$(2.74)\quad \lim_{T\to\infty} \frac{1}{T} (X'\Omega^{-1}X) = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} W(t)\Sigma_t W'(t), \text{with } \Sigma_t = E\left[\varepsilon(t)\varepsilon'(t)\right],$$

so that $\tilde{\beta}$ is asymptotically normally distributed with zero mean and variance-covariance matrix $\lim_{T\to\infty} \left(\frac{X'\Omega^{-1}X}{T}\right)^{-1}$, and since:

$$(2.75)\quad \operatorname*{plim}_{T\to\infty} \sqrt{T}\,(\beta^{*}-\beta) = \operatorname*{plim}_{T\to\infty} \sqrt{T}\,(\tilde{\beta}-\beta) \text{ or } \operatorname*{plim}_{T\to\infty} \sqrt{T}\,(\beta^{*}-\tilde{\beta}) = 0,$$

β^{*} has the same limiting normal distribution as $\tilde{\beta}$. Δ

(ii) Model II (H2)

The parameters are estimated in a 2-step procedure:

1. Estimate directly the transformed model (1.52), with Q_i equal to the expression (1.58), as a classical SUR-problem to yield a consistent estimate of Ω.
2. Compute the feasible Aitken estimator β^{*}.

The statistical properties of β^{*} are similar to those of the feasible Aitken estimator in model H1.

Note

Combinations of heteroscedastic and autocorrelated models are generally estimated by a 3-step procedure, obtained as a combination of the procedures described above (with covariance specifications (1.59-60)).

2.3 Singular error variance-covariance matrix and/or X-matrix of incomplete rank.

If the true error covariance matrix $\Omega=\Sigma\otimes I_T$ is unknown, the last p=M-s eigenvalues of $\hat{\Sigma}$, being of "preliminary" rank s and based on OLS-estimates of $\hat{\varepsilon}_i(i=1,2,\ldots,M)$, may be tested on their (significant) departure from a preassigned small value λ_o (given $\varepsilon(t)$ is assumed to be M-dimensionally normally distributed) by the following χ^2-test statistic (see app.B of part I:theorem B1):

$$(2.76)\quad \chi_q^2=\left\{T-1-s-\frac{1}{6}\left(2p+1-\frac{2}{p+1}\right)-\frac{1}{p+1}\left[\sum_{i=1}^{s}\left(\frac{\hat{\lambda}_i}{\hat{\lambda}_i-\lambda_o}\right)\right]^2+\lambda_o^2\sum_{i=1}^{s}\frac{1}{(\hat{\lambda}_i-\lambda_o)^2}\right\}$$

$$\left\{p\ \ln\lambda_o-\ln\left(\frac{|\hat{\Sigma}|}{\prod_{i=1}^{s}\hat{\lambda}_i}\right)+\frac{(\mathrm{tr}\ \hat{\Sigma}-\sum_{i=1}^{s}\hat{\lambda}_i)}{\lambda_o}-p\right\}$$

with the d.f. $q=\frac{1}{2}\ p(p+1)$

Once the "real rank" of $\hat{\Sigma}$, and hence of $\hat{\Omega}=\hat{\Sigma}\otimes I_T$ (and/or of $(X'\hat{\Omega}^+X)$) is determined, feasible and consistent Aitken estimators of β may be obtained by suitable substitution into the expressions (1.77-80) and (1.119).

Due to the singularity of the moment or covariance matrices, asymptotic normality in the sense of theorems 2.3,2.5 and 2.6 is not obtained. Despite the degeneracy in the pdf of the $\varepsilon(t)$'s, umbiasedness in the sense of theorem 2.2 may still be proved for several generalized models.

2.4 Feasible Aitken estimation of autoregressive models.

A feasible Aitken estimator of β in model (1.150) may be derived if an initial consistent estimator for Ω can be found. This may be obtained by several methods depending upon the possible presence of autocorrelation in the disturbance vectors ε_i (i=1,2,...,M).

2.41 If no autocorrelation of the disturbance terms

with $E\left[\varepsilon_i(t)\ \varepsilon_i(t-1)\right]=0$, then Ω is consistently estimated by OLS (by ML if $\varepsilon_i(t)$ are NID $(0,\sigma_{ii})$).

In the presence of autocorrelation, however, the OLS estimate of β is no longer consistent (see appendix B2).

2.42 If autocorrelation of the error terms and form of autocorrelation known.

Say first order autocorrelation:

$$(2.77)\quad \varepsilon_i(t) = \rho_i \varepsilon_i(t-1) + \eta_i(t)$$

Then:

- in a first stage, the ρ_i's may be estimated by OLS(see 2.41)
- in a second stage, the Cochran-Orcutt procedure is used to obtain a consistent estimator $\hat{\Omega}$, i.e. OLS is recursively applied on:

$$(2.78)\quad y_i(t)-\hat{\rho}_i y_i(t-1) = \sum_{k=1}^{k_i} \alpha_{ik}\left[z_{ik}(t)-\hat{\rho}_i z_{ik}(t-1)\right] +$$

$$\sum_{\tau=1}^{\theta_i} \gamma_{i\tau}\left[y_i(t-\tau)-\hat{\rho}_i y_i(t-\tau-1)\right] + \eta_i(t)$$

or the $\hat{\rho}_i$-estimates are substituted in an equation per equation covariance matrix to obtain Aitken's generalized least squares equation per equation estimates. From these second round parameter estimates, a consistent $\hat{\Omega}$ is derived.

2.43 If autocorrelation of the error terms but with unknown form.

Proposition 2.3

If the form of autocorrelation is unknown, the parameters of an autoregressive model are consistently estimated by instrumental variables.

Proof

The purpose of instrumental variable estimation is to replace the lagged dependent variables in equations (1.149) by those linear combinations of all explanatory variables z_{ik} which are most strongly correlated with the corresponding lagged explained variable but uncorrelated with the error vector ε_i, or the "best choice" instrumental variables for $y_i(t-\tau)$ are the lagged values of:

(2.79) $y_i^*(t) = \sum_{k=1}^{k_i} \hat{\lambda}_{ik} z_{ik}(t)$ with $\hat{\lambda}_{ik}$ the OLS-coefficients from regression of y_i on Z_i.

Then the instrumental variable estimator for β is given by:

(2.80) $\beta_i^* = (X_i^{*'} X_i)^{-1} \; X_i^{*'} y_i$ with $X_i^* = (Z_i, y_{\theta_i}^*)$ $(i=1,2,\ldots,M)$

where $\underset{T\to\infty}{\text{plim}} \frac{1}{T} X_i^{*'} \varepsilon_i = 0$ and $\underset{T\to\infty}{\text{plim}} \left(\frac{X^{*'}X}{T}\right)$ exists and is non singular.

The estimated parameter vectors (2.80) are consistent because:

(2.81) $\underset{T\to\infty}{\text{plim}}\, \beta_i^* = \underset{T\to\infty}{\text{plim}}\, (X_i^{*'} X_i)^{-1} X_i^{*'} (X_i \beta_i + \varepsilon_i) = \beta_i + \underset{T\to\infty}{\text{plim}} (X_i^{*'} X_i)^{-1} X_i^{*'} \varepsilon_i$

$$= \beta_i + \underset{T\to\infty}{\text{plim}} \left(\frac{X_i^{*'} X_i}{T}\right)^{-1} \underset{T\to\infty}{\text{plim}} \frac{X_i^{*'} \varepsilon_i}{T} = \beta_i \;. \qquad \Delta$$

To estimate the variance-covariance matrix Ω, one has to introduce restrictions about the form of autocorrelation (and/or heteroscedasticity), so that Ω can directly be estimated from the consistent instrumental variables estimator β_i^* $(i=1,2,\ldots,M)$. Many authors however (see e.g. T.Amemiya and W.Fuller [1] and K.Wallis [24]) propose to follow up the instrumental variables estimation by an equation per equation Aitken estimation of the parameters in the equation. Although the resulting parameter estimation also yields consistent estimates, there emerges a loss of asymptotic efficiency because of the joint occurrence of two factors:

- the use of an estimated variance-covariance matrix
- the presence of lagged dependent variables (see Appendix B3).

So, in fact, there is no fundamental reason to make the job more complicated by further applying GLS after instrumental variables estimation of ρ_i, the more since only consistent estimators are needed.

Given an initial consistent estimator $\hat{\Omega}$, the feasible Aitken estimator of β is consistent since the estimator:

$$(2.82)\quad \beta^{*}_{(n)} = \left[X'_{(n)}\hat{\Omega}^{-1}_{(n)}X_{(n)}\right]^{-1}X'_{(n)}\hat{\Omega}^{-1}_{(n)}y_{(n)}.$$

tends i.p. to the ML estimator for known Ω and increasing $n=MT$ (such as $\frac{M}{T} \to o$ as $n \to \infty$, e.g. if M remains fixed) obtained from:

$$(2.83)\quad \max_{R^K} L(y|\beta,\Omega^{-1}) = (2\pi)^{-\frac{1}{2}n}|\Omega^{-1}|^{\frac{1}{2}} \exp\left(-\frac{1}{2}\varepsilon'\Omega^{-1}\varepsilon\right)$$

If $E\left[y(n)X(n)\right]'\left[y(n)X(n)\right] = A(n)$ are finite and $\lim_{n\to\infty}\frac{A(n)}{n} = A$ exists and is finite, where $X(n)$ and $y(n)$ are the observations on the dependent and explanatory variables written so as to depend explicitly upon the number of observations.

III An Application: A stochastic model for the generation production coefficients.

3.1 In economics, interindustrial analysis is dominated by the famous Leontief model, based on the assumption of constant production coefficients, being defined as:

$$(3.1) \quad a_{ij,t}=q_{ij,t}/q_{j,t}$$

with $q_{ij,t}$ representing the input of industry j at time t of commodities produced by industry i and $q_{j,t}$ representing the corresponding production at time t of industry j.

However it is generally recognized, that the assumption of constant production coefficients (or zero elasticity of factor substitution) can only be valid as a first (and very rough) approximation. So, the problem arises whether it is possible to build a model for the generation of production coefficients themselves.

Economic reasoning involving market behaviour and production theory on industry level (*), may lead, under profit maximization, to the following simple stochastic specification of production coefficients:

$$(3.2) \quad a_{ij,t}=a_{ij,0}\left(\frac{p_{i,t}}{p_{n,t}}\right)^{\lambda_i}\left(\frac{p_{j,t}}{p_{n,t}}\right)^{\mu_j} u_{ij,t} \quad (i,j=1,2,\ldots,n-1) .$$

In this equation $p_{i,t}$ and $p_{j,t}$ stand for the price of the production, at time t, of industries i and j, where as $p_{n,t}$ symbolizes the wage level (as a weighted average of all prices). The factor $u_{ij,t}$ refers to the disturbance term and λ_i and μ_j are unknown parameters.To some extent, this model specification follows the Walrasian theory in which production coefficients are ultimately explained by relative prices.

(*) See R.A.van Straelen: <u>Prijsontwikkeling en Productiestructuur</u>,

Ph.D.Thesis, Louvain, 1970 (Dutch-unpublished).

As can be remarked, only 2(n-1) estimations of parameters λ_i and μ_j are needed for explaining $(n-1)^2$ coefficients of production. This limited number of parameters may be considered as a very appealing feature of the model. Another attractive feature of the model consists in its possible relationship to the wellknown RAS-method for generating production coefficients:

$$(3.3) \quad \hat{a}_{ij,t}=r_{i,t}a_{ij,0}s_{j,t}$$

where $\hat{a}_{ij,t}$ represents the generated production coefficient (i,j) by means of the corresponding production coefficient $a_{ij,0}$ of the base period and the RAS-multipliers $r_{i,t}$ and $s_{j,t}$. If time series on $a_{ij,t}$ are not available, one can use generated production coefficients $\hat{a}_{ij,t}$, provided that marginal totals of input-output tables are known for each period t in order to deduce the RAS-multipliers. We will concentrate on this case by studying the estimation aspects of the basic model (3.2).

32. One of the problems we meet in estimating the parameters of model (3.2) is the multicollinearity problem mainly caused by the division of all production prices by a common factor viz. the wage level. We can get rid of this difficulty by stating the estimation procedure in terms of the RAS-multipliers (*). It has to be noticed that these multipliers are only defined for each period considered up to a constant multiple. Therefore, we have to read equation (3.3) as follows:

$$(3.4) \quad \hat{a}_{ij,t}=r_{i,t}\left(\frac{1}{v_t}\right)a_{ij,0}s_{j,t}v_t.$$

Random disturbances are assumed to represent the discrepancy between $a_{ij,t}$ and $\hat{a}_{ij,t}$. One can write:

$$(3.5) \quad a_{ij,t}=\hat{a}_{ij,t}v_{ij,t}.$$

(*) The problem of multicollinearity could be tackled in other ways, e.g. by a generalized inverse estimation under certain parameter constraints. However, the procedure would seem less efficient.

Rewriting (3.2) in terms of the generated production coefficients leads to:

$$(3.6)\quad \hat{a}_{ij,t}=a_{ij,0}\left(\frac{p_{i,t}}{p_{n,t}}\right)^{\lambda_i}\left(\frac{p_{j,t}}{p_{n,t}}\right)^{\mu_j}\frac{u_{ij,t}}{v_{ij,t}}$$

and in accordance to the particular role of the RAS-multipliers r_i and s_j we assume that:

$$(3.7)\quad \frac{u_{ij,t}}{v_{ij,t}} = w_{i,t}z_{j,t}$$

where $w_{i,t}$ are random error terms standing for "disturbances" over the rows (input structure) and $z_{j,t}$ random error terms standing for "disturbances" over the columns (output structure).

Then it becomes possible, by combining (3.4) and (3.6), to write:

$$(3.8)\quad r_{i,t}=\left(\frac{p_{i,t}}{p_{n,t}}\right)^{\lambda_i} w_{i,t}v_t \qquad \text{and}$$

$$(3.9)\quad s_{j,t}=\left(\frac{p_{j,t}}{p_{n,t}}\right)^{\mu_j} \frac{z_{j,t}}{v_t}\ .$$

Expressing (3.8) and (3.9) for simplicity in (natural) logarithms, we have:

$$(3.10)\quad r'_{i,t}=\lambda_i\pi_{i;t}+w'_{i,t}+v'_t \qquad \text{and}$$

$$(3.11)\quad s'_{j,t}=\mu_j\pi_{j,t}+z'_{j,t}-v'_t \qquad ,$$

where the accent refers to the operation of taking (natural) logarithms of magnitudes involved and

$$(3.12)\quad \pi_{i,t}=\ln\left(p_{i,t}/p_{n,t}\right) \text{ and } \pi_{j,t}=\ln\left(p_{j,t}/p_{n,t}\right)\ .$$

In this way, we obtain a rather simple system of equations that allows us to estimate the unknown parameters λ_i and μ_j. By using RAS-multipliers which are function of only one explanatory variable, the problem of multicollinearity disappears. We now turn to the estimation procedure.

33. The multiples v_t and their reciprocals, which are the same for all RAS-multipliers of period t, can be interpreted as additional parameters to be estimated or can be considered as stochastic factors belonging to the disturbance terms of the model. From the view point of estimation, both ways of thinking lead to results which are asymptotically equivalent (*). However, the first way is more complicated. Therefore, we are only proceeding along the second way.

By considering v_t stochastically we obtain an interesting application of a SUR-model, because the same residual component appears in all equations of the same period. Using matrix and vector notation we can write the whole system of equations for an arbitrary period t in the usual manner as follows:

(3.13) $y_t = X_t \beta + u_t$

in which y_t is the 2(n-1)-vector $\begin{pmatrix} r_t \\ s_t \end{pmatrix}$, X_t the diagonal matrix $\begin{bmatrix} \hat{\pi}_t & 0 \\ 0 & \hat{\pi}_t \end{bmatrix}$ of order 2(n-1)×2(n-1) and u_t the 2(n-1)-vector $\begin{pmatrix} w_t + iv_t \\ z_t - iv_t \end{pmatrix}$, omitting accents in order to avoid confusion with the transpose symbol. However, all variables which we have considered remain expressed in (natural) logarithms. Vector i represents a vector with all elements equal to unity.

In principle, the variance-covariance matrix can be estimated by using the residuals of the first round ordinary least squares. Or

(3.14) $$\hat{\Sigma} = \frac{1}{T-1} \sum_{t=1}^{T} \begin{bmatrix} r_t - \hat{r}_t \\ s_t - \hat{s}_t \end{bmatrix} \begin{bmatrix} r_t - \hat{r}_t \\ s_t - \hat{s}_t \end{bmatrix}'$$

which implies temporal independency of the disturbance terms (neither autocorrelation nor heteroscedasticity). Therefore, the variance-covariance matrix enters the Aitken estimator as

(3.15) $$\hat{\Sigma}^{-1} \otimes I_{2(n-1)} \qquad \text{(classical SUR-model)} .$$

(*) We are due to Prof.A.P.Barten for this conclusion. See also R.A. Van Straelen, <u>o.c.</u>, Chapter 6.

However, as explained above, it is possible to modify (3.15) in order to take temporal dependencies among the disturbance terms into account. Obviously, the necessary condition to be satisfied for non-singularity of (3.15) is

$$(3.16)\quad T \geq 2(n-1)\ .$$

If this condition is violated, additional assumptions are to be made concerning the pdf of the disturbance terms and their covariance matrix. An other possibility consists in defining a generalized SUR-model with the help of $\hat{\Sigma}^{+}$. This procedure can also be followed when the variance-covariance matrix turns out to be "nearly singular". Test statistic (2.76) may be utilized then for determining the "real significant rank" of $\hat{\Sigma}$. Alternatively, a consistent minimax norm as e.g. $\xi\sqrt{2\sum_{i=1}^{(n-1)}\lambda_i^2}$, with ξ the maximum relative error in the eigenvalues, may be subtracted from each eigenvalue of $\hat{\Sigma}$. The rank is then equal to the number of positive corrected eigenvalues.

34. A dynamic version of the model can be obtained by introducing lagged variables. Demand and supply often react not only to present but also to previous prices. Up to now we know very little about the precise form of lag structures. However, an important and useful structure is that of Koyck. As it is well known, the Koyck lag structure rests upon the basic assumption of a geometrically declining effect. Applying the Koyck transformation leads to a very simple dynamic specification. Our model formed by the equations (3.10) and (3.11) now becomes:

$$(3.17)\quad \begin{bmatrix} r_t \\ s_t \end{bmatrix} = \begin{bmatrix} \hat{\pi}_t & 0 \\ 0 & \hat{\pi}_t \end{bmatrix} \begin{bmatrix} \lambda^{*} \\ \mu^{*} \end{bmatrix} + \begin{bmatrix} \hat{r}_{t-1} & 0 \\ 0 & \hat{s}_{t-1} \end{bmatrix} \begin{bmatrix} \rho \\ \rho \end{bmatrix} + u_t$$

As can be noticed the number of parameters increases only with (n-1) resulting from assuming the same lag parameters for the corresponding row and column multipliers. This makes sense by considering the argument that if there exists some lagged behaviour in an industry, it is very likely to happen at all levels more or less in the same manner. Obviously, the content of the disturbance term u_t differs completely from the residual

u_t defined in (3.13). The price parameters of both models are related to each other as follows:

(3.18) $\lambda^{*}=\lambda(1-\rho)$ and $\mu^{*}=\mu(1-\rho)$

The estimation of the parameter vector ρ can be done in an easy way by taking the sum of the vectors of row and column multipliers. By means of this sum we are getting rid of the constant multiple v_t. Estimating ρ in this way no longer forms an application of an autoregressive SUR-model. So, autoregressive constrained and unconstrained SUR-estimates (2(n-1) in number) will be compared in the next paragraph. After obtaining consistent first round estimates, they can be substituted in (3.17) (*) which allows us to obtain consistent second round estimates either of $\lambda^{*},\mu^{*},\rho_1$ and ρ_2 in the autoregressive SUR-model or of λ^{*} and μ^{*} in the classical SUR-models with redefined dependent variables

(3.19) $$\begin{bmatrix} r_t \\ s_t \end{bmatrix} - \begin{bmatrix} \hat{r}_{t-1} & 0 \\ 0 & \hat{s}_{t-1} \end{bmatrix} \begin{bmatrix} \rho \\ \rho \end{bmatrix}$$

We now turn to some briefly commented numerical results.

35. Some experiments have been performed for the Belgian economy during the period 1953-1967. Basic data were:annual relative prices, the input-output table for 1959 and annual marginal totals on 12 aggregate industries.

A complete report on the numerical results does not meet the objectives of this memorandum (**). So, we have limited ourselves to the statement of some main results. Estimates are given for two industries:one for which the performance of the model was relatively poor (building industry) and another one for which the performance of the model was relatively good (energy sector). Five models have been retained:

(*) With obviously a vector ρ_1 and ρ_2 for row and column multipliers in the case of a real autoregressive SUR-model.

(**) See R.A.Van Straelen, o.c., for a more detailed description of the numerical experiments.

1. the classical SUR-model
2. the classical SUR-model with redefined dependent variables (autoregressive nature)
3. the real autoregressive SUR-model with 24 ρ_i-parameters
4. a "static" generalized SUR-model with positive semi-definite variance covariance matrix $\hat{\Omega}$ (see 1)
5. a "dynamic" generalized SUR-model with positive semi-definite $\hat{\Omega}$ (see 2).

The results are presented in the following tables, where OLS and SUR estimates are given for the unknown parameter values and their standard deviations.

Table 1 : Energy Sector

	Classical SUR-model	Classical SUR-model (auroregressive nature)	Real autoregressive SUR-model
OLS	$\hat{\lambda}$ = -0.3621 (0.1021) $\hat{\mu}$ = 0.6407 (0.1252)	$\hat{\rho}$ = 0.1948 (ML) $\hat{\lambda}$ = -0.2867 (0.0914) $\hat{\mu}$ = 0.5313 (0.1094)	$\hat{\lambda}$ = 0.0413 $\hat{\rho}_\lambda$ = 1.1041 (0.1476) (0.4285) $\hat{\mu}$ = -0.0293 $\hat{\rho}_\mu$ = 1.1211 (0.1531) (0.2687)
SUR	$\hat{\lambda}$ = -0.2613 (0.0948) $\hat{\mu}$ = 0.5163 (0.1164)	$\hat{\lambda}$ = -0.2137 (0.0852) $\hat{\mu}$ = 0.4371 (0.1019)	$\hat{\lambda}$ = -0.0230 $\hat{\rho}_\lambda$ = 0.8226 (0.0751) (0.1370) $\hat{\mu}$ = 0.0084 $\hat{\rho}_\mu$ = 1.1558 (0.0657) (0.0660)

	"Static" generalized SUR-model	"Dynamic" generalized SUR-model
	$\hat{\lambda}$ = -0.2659 (0.1042) $\hat{\mu}$ = 0.5335 (0.1185)	$\hat{\rho}$ = 0.1948 (ML) $\hat{\lambda}$ = -1.1945 (1.1556) $\hat{\mu}$ = 0.9189 (0.9506)

Table 2 : Building Industry

	Classical SUR-model	Classical SUR-model (autoregressive nature)	Real autoregressive SUR-model	
OLS		$\hat{\rho}$ = 0.8271 (ML)		
	$\hat{\lambda}$ = -6.9024	$\hat{\lambda}$ = -1.9700	$\hat{\lambda}$ = -0.8715	$\hat{\rho}_\lambda$ = 1.0097
	(1.6636)	(0.6565)	(1.5773)	(0.2760)
	$\hat{\mu}$ = 2.9459	$\hat{\mu}$ = 0.9974	$\hat{\mu}$ = 1.2619	$\hat{\rho}_\mu$ = 0.7155
	(0.6142)	(0.3324)	(1.3439)	(0.5976)
SUR	$\hat{\lambda}$ = -4.4845	$\hat{\lambda}$ = -1.3107	$\hat{\lambda}$ = -0.8883	$\hat{\rho}_\lambda$ = 0.9316
	(1.3994)	(0.5522)	(0.4897)	(0.0935)
	$\hat{\mu}$ = 2.1162	$\hat{\mu}$ = 0.8035	$\hat{\mu}$ = 1.1226	$\hat{\rho}_\mu$ = 0.7399
	(0.5140)	(0.2785)	(0.1229)	(0.0802)

	"Static" generalized SUR-model	"Dynamic" generalized SUR-model
SUR		$\hat{\rho}$ = 0.8271
	$\hat{\lambda}$ = -4.5582	$\hat{\lambda}$ = -1.3702
	(1.3998)	(0.5526)
	$\hat{\mu}$ = 2.1724	$\hat{\mu}$ = 0.8080
	(0.5166)	(0.3009)

As can be observed, differences between OLS and SUR are quite important. Judged against the usual standards of the t-test for determiming the significance of individual parameter estimates, model 3 (the real autoregressive model) gives no satisfactory results compared with the first two models.

By analyzing the eigenvalues of the estimated variance covariance matrix $\hat{\Sigma}$, used in $\hat{\Omega}_i = \hat{\Sigma}_i \otimes I_{2(n-1)}$ (i=1,2) for models 1 and 2, we observed that, taking account of a limited error on the accuracy of the variance-covariance elements, the "real rank" of $\hat{\Sigma}_1$ could be fixed at 21 and of $\hat{\Sigma}_2$ at 20. Also, the sole break in the evolution of the eigenvalues occurred at those places.

Table 3 Smallest 6 eigenvalues of $\hat{\Sigma}_1$ and $\hat{\Sigma}_2$.

Nr.	$\hat{\Sigma}_1$	Nr.	$\hat{\Sigma}_2$
19	0.001385	19	0.001733
20	0.001303	20	0.001615
21	0.001253	21	0.000566
22	0.000556	22	0.000515
23	0.000488	23	0.000452
24	0.000436	24	0.000369

Therefore, pseudo-inverses of $\hat{\Sigma}_1$ and $\hat{\Sigma}_2$ were computed, resp. with ranks 20 and 21 (*). The estimates of models 4 and 5 given in tables 1 and 2 are obtained from:

(3.20) $\beta^{*}=(X'\hat{\Omega}^{+}X)^{-1}X'\hat{\Omega}^{+}y$ (see (1.65)) and

(3.21) $V^{*}(\beta^{*})=(X'\hat{\Omega}^{+}X)^{-1}$ (see (1.66)).

Notice, however that (3.20) is not asymptotically most efficient, which follows immediately from proposition 1.1, but <u>is</u> consistent if $\hat{\Omega}$ is a consistent estimate of Ω.

To judge the relevance of the parameter estimates, some simultaneous tests on a priori restrictions of the parameters have been performed:

a. for models 1,2,4 and 5:

- $Ho:\lambda_1=\lambda_2=.....=\lambda_{12}=\mu_1=\mu_2=....=\mu_{12}=0$ (zero restrictions)
- $Ho:\lambda_1=\lambda_2=.....=\lambda_{12}=\mu_1=\mu_2=.....=\mu_{12}$
- $Ho:\lambda_1=\mu_1;\lambda_2=\mu_2;.....,\lambda_{12}=\mu_{12}$.

(*) Since, in fact, the pseudo-inverse of an "approximate matrix" (corresponding to the"postulated rank") is computed by the method retained, it was interesting to notice that in both cases the elements of the approximant differed with only 2% at the maximum from the original elements in $\hat{\Sigma}_1$ and $\hat{\Sigma}_2$, which is largely within the range of the allowed inaccuracy.

b. for model 3:

- $Ho: \begin{pmatrix} \lambda_1 \\ \rho_{\lambda_1} \end{pmatrix} = \begin{pmatrix} \lambda_2 \\ \rho_{\lambda_2} \end{pmatrix} = \ldots\ldots = \begin{pmatrix} \lambda_{12} \\ \rho_{\lambda_{12}} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ (zero restrictions)

- $Ho: \begin{pmatrix} \mu_1 \\ \rho_{\mu_1} \end{pmatrix} = \begin{pmatrix} \mu_2 \\ \rho_{\mu_2} \end{pmatrix} = \ldots\ldots = \begin{pmatrix} \mu_{12} \\ \rho_{\mu_{12}} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ (" ") (*)

- $Ho: \begin{pmatrix} \lambda_1 \\ \rho_{\lambda_1} \end{pmatrix} = \begin{pmatrix} \lambda_2 \\ \rho_{\lambda_2} \end{pmatrix} = \ldots\ldots = \begin{pmatrix} \lambda_{12} \\ \rho_{\lambda_{12}} \end{pmatrix}$ (vector equality)

- $Ho: \begin{pmatrix} \mu_1 \\ \rho_{\mu_1} \end{pmatrix} = \begin{pmatrix} \mu_2 \\ \rho_{\mu_2} \end{pmatrix} = \ldots\ldots = \begin{pmatrix} \mu_{12} \\ \rho_{\mu_{12}} \end{pmatrix}$ (" ").

The test statistic on zero restrictions can simply be written as a simultaneous significance test of the parameters:

$$(3.22)\quad F^{*}_{K,s-K} = \frac{s-K}{K}\,\frac{y^{*'}\hat{\Omega}^{+}y^{*}}{y'\hat{\Omega}^{+}y - y^{*'}\hat{\Omega}^{+}y^{*}} = \frac{s-K}{K}\,\frac{y'\hat{\Omega}^{+}y^{*}}{y'\hat{\Omega}^{+}y - y'\hat{\Omega}^{+}y^{*}} ,$$

with $y^{*} = X\beta^{*} = X(X'\hat{\Omega}^{+}X)^{-1}X'\hat{\Omega}^{+}y$, which is asymptotically F-distributed with K and s-K degrees of freedom (s being the rank of $\hat{\Omega}$).

The remaining a priori restrictions may all be expressed as linear homogeneous restrictions on the parameter vector β, written as:

(*) The equality hypothesis $\begin{pmatrix} \lambda_i \\ \rho_{\lambda_i} \end{pmatrix} = \begin{pmatrix} \mu_j \\ \rho_{\mu_j} \end{pmatrix}$ for $\forall_{i,j}$ $i,j=1,2,\ldots,12$ could not be tested since the complete SUR model (24 equations with 48 explanatory variables) was too large for the dimensions of the present programming system available at the Tilburg University ICL-installation. Therefore we had to split the problem into 2 submodels, both of 12 equations (row, viz. column multipliers) and 24 explanatory variables, so that the $\hat{\lambda}_i$- and $\hat{\mu}_j$-coefficients of model 3 are not "really global" SUR-results.

(3.23) $C\beta=0$

with C a known (q×K) matrix, q being the number of restrictions.

Following propositions C1 and C2 of the appendix we observe that:

$$(3.24)\quad F^{\star}_{q,s-K}=\frac{s-K}{q}\;\frac{\beta^{\star\prime}C'\left[C(X'\hat{\Omega}^{+}X)^{-1}C'\right]^{-1}C\beta^{\star}}{y'\hat{\Omega}^{+}y-\;y'\hat{\Omega}^{+}y^{\star}}$$

is asymptotically distributed as $F_{q,s-K}$.

The results are given in the underlying table:

Table 4 Asymptotic F-tests on a priori restrictions

Model 1	Model 2	Model 3	Model 4	Model 5
		$F^{\star}_{24,144}=167.2$		
$F^{\star}_{24,336}=44.8$	$F^{\star}_{24,312}=29.8$	$F^{\star}_{24,144}=472.7$	$F^{\star}_{24,291}=42.1$	$F^{\star}_{24,256}=22,6$
$F^{\star}_{23,336}=19.4$	$F^{\star}_{23,312}=15.6$	$F^{\star}_{22,144}=16.6$	$F^{\star}_{23,291}=18.0$	$F^{\star}_{23,256}=10.7$
$F^{\star}_{12,336}=19.2$	$F^{\star}_{12,312}=20.8$	$F^{\star}_{22,144}=92.2$	$F^{\star}_{12,291}=20.0$	$F^{\star}_{12,312}=11.8$

Comparing the results contained in the above table with the critical values of an F-table, we immediately see that all zero hypotheses are strongly rejected (even at a significance level of 99%).

Finally, the performance of the alternative models is compared by computing performance indices indicating the fitting degree of the model to the data. So, if z_t are the observations for time periods 1,2,...,T and $z^{\star}_t$ the corresponding calculated values, the performance index is defined as:

$$(3.25)\quad P.I.=\frac{\sqrt{\frac{1}{T}\sum_t (z_t-z^{\star}_t)^2}}{\sqrt{\frac{1}{T}\sum_t z_t^2}}\,,$$

which commonly indicates a good performance if it is smaller than 0.4. Such indices have been calculated for the observed and estimated marginal totals. Some results are presented in table 5. For both sectors, energy and building, the first figure refers to the performance index of the row total, where as the second figure refers to the performance index of the column total; for all 12 sectors together the figure mentioned refers to the overall performance regarding all marginal totals.

Table 5 Performance Indices.

	Energy Sector		Building Industry		All Sectors
	Row P.I.	Column P.I.	Row P.I.	Column P.I.	
Model 1	0.047	0.052	0.353	0.137	0.066
Model 2	0.041	0.034	0.145	0.069	0.051
Model 3	0.044	0.307	0.207	0.287	0.273
Model 4	0.137	0.170	0.605	0.531	0.561
Model 5	0.129	0.120	0.445	0.351	0.502

The above table indicates that the "generalized models" 4 and 5 give inferior performance compared with the first three, except for the performance of the column totals of the energy sector which is worst for the real autoregressive model 3. Indeed, it strikes immediately that the column totals are rather badly predicted by model 3, although this model does not give a very poor global performance (overall P.I. is smaller than 0.4 and row totals are even better predicted than with model 1). Also a substantial improvement of the building industry prediction capacity is noticed by using the lagged model 2 instead of the unlagged model 1. In general, this model 2 has the best performance of all models retained.

Appendix A Analysis of a classical SUR-two equation model (*)

1. The role of dependency between different sets of expl. atory variables w.r.t. efficiency.

Proposition A1

If the variance covariance matrix Ω is known, Aitken estimation of β in model

$$\text{(A.1)} \quad y=\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ o & X_2 \end{bmatrix}\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}$$

where $X_i\,(i=1,2)$ are non-stochastic $(T\times k_i)$-matrices of explanatory variables, yields maximum gain in efficiency w.r.t. OLS if $X_1' X_2 = 0$ and the disturbance vectors of the 2 equations are highly correlated.

Proof (see also theorem 2.4 for general M-equation systems)

The Aitken estimator of $\beta=(\beta_1,\beta_2)'$ in model (A.1) where it is assumed that $E\begin{pmatrix}\varepsilon_1\\ \varepsilon_2\end{pmatrix} = 0$ and

$$E\left[\begin{pmatrix}\varepsilon_1\\ \varepsilon_2\end{pmatrix}(\varepsilon_1' \ \ \varepsilon_2')\right]=\Omega=\begin{bmatrix} \sigma_{11}I_T & \sigma_{12}I_T \\ \sigma_{21}I_T & \sigma_{22}I_T \end{bmatrix}, \text{ is equal to:}$$

$$\text{(A.2)} \quad \tilde{\beta}=\begin{bmatrix} \bar{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} \sigma^{11}X_1'X_1 & \sigma^{12}X_1'X_2 \\ \sigma^{21}X_2'X_1 & \sigma^{22}X_2'X_2 \end{bmatrix}^{-1}\begin{bmatrix} \sigma^{11}X_1'y_1 + \sigma^{12}X_1'y_2 \\ \sigma^{21}X_2'y_1 + \sigma^{22}X_2'y_2 \end{bmatrix}$$

with variance covariance matrix (see 1.7)

$$\text{(A.3)} \quad V(\tilde{\beta})=E\left[(\tilde{\beta}-\beta)(\tilde{\beta}-\beta)\right]' = \begin{bmatrix} \sigma^{11}X_1'X_1 & \sigma^{12}X_1'X_2 \\ \sigma^{21}X_2'X_1 & \sigma^{22}X_2'X_2 \end{bmatrix}^{-1} = \begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix}^{-1}$$

(*) The exposition in this appendix is based upon A.Zellner [28], A.Zellner and D.Huang [27].

and leading submatrix for the first equation's coefficients:

$$(A.4)\quad V(\tilde{\beta}_1)=(W_{11}-W_{12}W_{22}^{-1}W_{21})^{-1}=\left[\sigma^{11}X_1'X_1 - (\sigma^{12})^2X_1'X_2(\sigma^{22})^{-1}(X_2'X_2)^{-1}X_2'X_1\right]^{-1}.$$

Remembering that the simple correlation coefficient between the disturbances of the 2 equations is defined as $\rho=\frac{\sigma_{12}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}}$ and $\sigma^{11}=(\sigma_{11}-\sigma_{12}\sigma_{22}^{-1}\sigma_{21})^{-1}=\sigma_{11}^{-1}(1-\rho^2)^{-1}$,

the variance covariance matrix (A.4) of the $\tilde{\beta}_1$-vector may be written as:

$$(A.5)\quad V(\tilde{\beta}_1)=\left[\frac{1}{\sigma_{11}(1-\rho^2)}(X_1'X_1)-\frac{\rho^2}{\sigma_{11}(1-\rho^2)}(X_1'X_2)(X_2'X_2)^{-1}(X_2'X_1)\right]^{-1}$$

and its generalized variance as:

$$(A.6)\quad |V(\tilde{\beta}_1)|=(1-\rho^2)^{k_1}\sigma_{11}|(X_1'X_1)^{-1}||I_{k_1}-\rho^2D|^{-1} \quad \text{with}$$

$$(A.7)\quad D=(X_1'X_1)^{-1}(X_1'X_2)(X_2'X_2)^{-1}X_2'X_1 \quad \text{and}$$

$$(A.8)\quad |I_{k_1}-\rho^2D|=\prod_{i=1}^{k_1}\lambda_i,$$ where λ_i are the eigenvalues of

$I_{k_1}-\rho^2D$, satisfying the characteristic determinantal equation:

$$(A.9)\quad |(I_{k_1}-\rho^2D)-\lambda I_{k_1}| = |D-\frac{(1-\lambda)}{\rho^2}I_{k_1}| = 0.$$

So, the values of $\frac{1-\lambda}{\rho^2}$ are the characteristic roots of D, being equal to the squared canonical correlation coefficients r_i^2 for the sets X_1 and X_2 or

$$(A.10)\quad r_i^2=\frac{1-\lambda_i}{\rho^2} \text{ and } \lambda_i=1-\rho^2r_i^2,$$ and the generalized variance (A.6) becomes, taking account of (A.8) and (A.10):

(A.11) $|V(\tilde{\beta}_1)| = \dfrac{(1-\rho^2)^{k_1}\sigma_{11}|(X_1'X_1)^{-1}|}{\prod_{i=1}^{k_1}(1-\rho^2 r_i^2)}$, and since $0 \leq r_i^2 \leq 1$

(A.12) $|V(\tilde{\beta}_1)| \leq \sigma_{11}|(X_1'X_1)^{-1}|$, from which it is clear that the equality sign only holds when <u>all</u> canonical correlation coefficients are equal to unity (i.e.if $X_1=X_2$). If the column vectors of X_1 and X_2 are mutually orthogonal, i.e. if $X_1'X_2=0$, then <u>all</u> canonical correlation coefficients are equal to zero and maximum gain in efficiency is obtained w.r.t. OLS, because then (A.11) is minimum for a certain $\rho^2 \neq 0$ (denominator=1). In this case ($X_1'X_2=0$) the higher the correlation among the disturbance terms amounts to, the more is gained , relative to OLS, by estimating model (A.1) by Aitken's method . Δ

2. The efficiency of Aitken relative to OLS-estimation concerning the unexplained variation

<u>Proposition A 2</u>

The unexplained variation (generalized unexplained variance) of the OLS estimation of model (A.1) with $X_1'X_2=0$ will be greater than that associated with the Aitken estimation of β in (A.1)

unless $T \leq \dfrac{k_1^2+k_2^2+k_1k_2\rho^2}{k_1+k_2}$, where ρ is the simple correlation coefficient between the disturbance vectors ε_1 and ε_2(assumption: $\Omega=\Sigma\otimes I_T$ is known).

Proof.

Since:

(A.13) $$\tilde{\varepsilon}=\begin{pmatrix}\tilde{\varepsilon}_1\\ \tilde{\varepsilon}_2\end{pmatrix}=\left[I-X(X'\Sigma^{-1}\otimes I_T X)^{-1}X'\Sigma^{-1}\otimes I_T\right]\varepsilon$$ and $X_1'X_2=0$,

the residual vector of system (A.1) amounts to:

(A.14) $$\begin{pmatrix}\tilde{\varepsilon}_1\\ \tilde{\varepsilon}_2\end{pmatrix}=\left[\begin{pmatrix}I_T & 0\\ 0 & I_T\end{pmatrix}-\begin{pmatrix}\frac{1}{\sigma^{11}}X_1(X_1'X_1)^{-1}X_1' & 0\\ 0 & \frac{1}{\sigma^{22}}X_2(X_2'X_2)^{-1}X_2'\end{pmatrix}\right.$$

$$\begin{pmatrix} \sigma^{11} I_T & \sigma^{12} I_T \\ \sigma^{21} I_T & \sigma^{22} I_T \end{pmatrix} \Bigg] \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} \quad \text{or}$$

(A.15) $\tilde{\varepsilon}_1 = \left[I_T - X_1(X_1'X_1)^{-1}X_1'\right]\varepsilon_1 - \frac{\sigma^{12}}{\sigma^{11}} X_1(X_1'X_1)^{-1}X_1'\varepsilon_2$ and since $\frac{\sigma^{12}}{\sigma^{11}} = -\sigma_{22}^{-1}\sigma_{21} = -\rho\sqrt{\frac{\sigma_{11}}{\sigma_{22}}}$ and taking variance-covariance elements:

(A.16) $E(\tilde{\varepsilon}_i'\tilde{\varepsilon}_i) = \sigma_{ii}(T - k_i + \rho^2 k_i)$ $(i=1,2)$ (see also (2.38-39)) and

(A.17) $$E(\tilde{\varepsilon}_1'\tilde{\varepsilon}_2) = E\Big\{\varepsilon_1'\left[I_T - X_1(X_1'X_1)^{-1}X_1'\right]\left[I_T - X_2(X_2'X_2)^{-1}X_2'\right]\varepsilon_2 + \rho\sqrt{\frac{\sigma_{11}}{\sigma_{22}}}\varepsilon_2'X_1(X_1'X_1)^{-1}X_1'\left[I_T - X_2(X_2'X_2)^{-1}X_2'\right]\varepsilon_2 + \rho\sqrt{\frac{\sigma_{22}}{\sigma_{11}}}\varepsilon_1'\left[I_T - X_1(X_1'X_1)^{-1}X_1'\right]X_2(X_2'X_2)^{-1}X_2'\varepsilon_1 + \rho^2\varepsilon_2'X_1(X_1'X_1)^{-1}X_1'X_2(X_2'X_2)^{-1}X_2'\varepsilon_1\Big\} = T\sigma_{12}$$ (see (2.38)).

or the generalized unexplained variance of the Aitken estimators of the disturbances is given by:

(A.18) $$V_A = \begin{vmatrix} E(\tilde{\varepsilon}_1'\tilde{\varepsilon}_1) & E(\tilde{\varepsilon}_1'\tilde{\varepsilon}_2) \\ E(\tilde{\varepsilon}_2'\tilde{\varepsilon}_1) & E(\tilde{\varepsilon}_2'\tilde{\varepsilon}_2) \end{vmatrix} = \begin{vmatrix} \sigma_{11}(T - k_1 + k_1\rho^2) & \sigma_{12}T \\ \sigma_{21}T & \sigma_{22}(T - k_2 + k_2\rho^2) \end{vmatrix} =$$

$$= \sigma_{11}\sigma_{22}\left[T - k_1(1-\rho^2)\right]\left[T - k_2(1-\rho^2)\right] - \sigma_{12}^2 T^2$$

while the generalized unexplained variance of the OLS-residuals is (see (2.39) with $r_H^2 = 0$):

(A.19) $V_{OLS} = \sigma_{11}\sigma_{22}(T-k_1)(T-k_2) - \sigma_{12}^2(T-k_1-k_2)^2$ or

(A.20) $$V_{OLS} - V_A = \sigma_{11}\sigma_{22}\Big\{\left[(T-k_1)(T-k_2) - \rho^2(T-k_1-k_2)^2\right] - \left[(T-k_1(1-\rho^2))(T-k_2(1-\rho^2)) - \rho^2T^2\right]\Big\}$$

$$= \sigma_{11}\sigma_{22}\rho^2\left[T(k_1+k_2) - (k_1^2 + k_2^2 + k_1k_2\rho^2)\right],$$

from which it is seen that $V_{OLS} > V_A$ if

$$\text{(A.21)} \quad T > \frac{k_1^2+k_2^2+k_1k_2\rho^2}{k_1+k_2} \quad ,$$

so that ,certainly, the unexplained variation through OLS-estimation of model (A.1) is larger than the unexplained variation through Aitken estimation as long as the number of observations is at least as large as the total number of explanatory variables in the system under consideration ($T \geq k_1+k_2$). Δ

3. Exact finite sample properties of the feasible Aitken estimator in the classical SUR-model.

We consider the following model (see (A.1)):

$$\text{(A.22)} \quad \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}$$

with (i) X_i non stochastic $(T \times k_i)$-matrices $(i=1,2)$ of explanatory variables;

(ii) $X_1'X_2 = X_2'X_1 = 0$: pairwise orthogonality of the explanatory variables in the 2 equations (*) ;

(iii) $E\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} = 0$ and $E\left[\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}(\varepsilon_1' \varepsilon_2')\right] = \Omega = \begin{bmatrix} \sigma_{11}I_T & \sigma_{12}I_T \\ \sigma_{21}I_T & \sigma_{22}I_T \end{bmatrix}$

and Ω is unknown;

(*) In proposition A1 it was shown that there is maximum gain in efficiency if $X_1'X_2 = X_2'X_1 = 0$ for certain $\rho \neq 0$. This result only holds asymptotically for the feasible Aitken estimator. Furthermore, the condition of pairwise orthogonality is assumed in this section for ease of derivation. Hence, this assumption is not essential to the results obtained.

(iv) $\varepsilon(t)=\left[\varepsilon_1(t),\varepsilon_2(t)\right]'$ are assumed to be bivariate normal for each t $(t=1,2,\dots,T)$ with $E\left[\varepsilon(t)\right]=0$ and $E\left[\varepsilon(t)\varepsilon'(t)\right]=\Sigma$, so that the $\varepsilon(t)$ are assumed to be mutually independent identically normally distributed with zero mean and Σ variance-covariance matrix.

3.1 Exact sample moments of the feasible Aitken estimator β^*

Theorem A1.

The feasible Aitken estimator β^* of β in model (A.22) has the following finite sample moments:

(A.23) $E(\beta^*)=\beta$ (*) and

(A.24) $$V(\beta^*)=E\left[(\beta^*-\beta)(\beta^*-\beta)'\right]=(1-\rho^2)\frac{n-1}{n-2}\begin{bmatrix}\sigma_{11}(X_1'X_1)^{-1} & 0\\ 0 & \sigma_{22}(X_2'X_2)^{-1}\end{bmatrix}$$

where $n=T-k_1-k_2>0$

Proof

The feasible Aitken estimator β^* of model (A.22) is given by:

(A.25) $$\beta^*=\begin{pmatrix}\beta_1^*\\ \beta_2^*\end{pmatrix}=\begin{bmatrix}\hat{\sigma}^{11}X_1'X_1 & 0\\ 0 & \hat{\sigma}^{22}X_2'X_2\end{bmatrix}^{-1}\begin{bmatrix}\hat{\sigma}^{11}X_1'y_1+\hat{\sigma}^{12}X_1'y_2\\ \hat{\sigma}^{21}X_2'y_1+\hat{\sigma}^{22}X_2'y_2\end{bmatrix}$$

$$=\begin{bmatrix}(X_1'X_1)^{-1}X_1'y_1+\dfrac{\hat{\sigma}^{12}}{\hat{\sigma}^{11}}(X_1'X_1)^{-1}X_1'y_2\\ \dfrac{\hat{\sigma}^{21}}{\hat{\sigma}^{22}}(X_2'X_2)^{-1}X_2'y_1+(X_2'X_2)^{-1}X_2'y_2\end{bmatrix}$$

(*) From theorem 2.2, normality of $\varepsilon(t)$ about zero is a sufficient condition for unbiasedness of β^*.

$$= \begin{bmatrix} (X_1'X_1)^{-1}X_1'y_1 - \frac{\hat{\sigma}_{12}}{\hat{\sigma}_{22}} (X_1'X_1)^{-1}X_1'y_2 \\ \\ (X_2'X_2)^{-1}X_2'y_2 - \frac{\hat{\sigma}_{12}}{\hat{\sigma}_{11}} (X_2'X_2)^{-1}X_2'y_1 \end{bmatrix}$$

with variance covariance matrix:

$$\text{(A.26)}\quad V(\beta^*) = \begin{bmatrix} \frac{(X_1'X_1)^{-1}}{\hat{\sigma}^{11}} & 0 \\ 0 & \frac{(X_2'X_2)^{-1}}{\hat{\sigma}^{22}} \end{bmatrix} = (1-\hat{\rho}^2) \begin{bmatrix} \hat{\sigma}_{11}(X_1'X_1)^{-1} & 0 \\ 0 & \hat{\sigma}_{22}(X_2'X_2)^{-1} \end{bmatrix}$$

where $\hat{\Sigma}=\{\hat{\sigma}_{ij}\}=\{\frac{\hat{\varepsilon}_i'\hat{\varepsilon}_j}{T}\}(i,j=1,2)$ based on the maximum likelihood (OLS) estimation of the parameter vectors of the equations involved in (A.22) and $\hat{\rho}$ the correlation coefficient between $\hat{\varepsilon}_1$ and $\hat{\varepsilon}_2$: $\hat{\rho}=\frac{\hat{\sigma}_{12}}{\sqrt{\hat{\sigma}_{11}\hat{\sigma}_{22}}}$

To establish the exact finite sample moments of (A.25), consider the model:

$$\text{(A.27)}\quad \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \left[\begin{array}{cc|cc} X_1 & X_2 & 0 & 0 \\ \hline 0 & 0 & X_1 & X_2 \end{array}\right] \begin{bmatrix} \beta_1 \\ \beta_{10} \\ \hline \beta_{20} \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} ,$$

which is equivalent to model (A.22) if $\beta_{10}=\beta_{20}=0$.

Under the assumption that $\varepsilon(t)=[\varepsilon_1(t),\varepsilon_2(t)]'$ are bivariate normal with mean 0 and variance covariance matrix Σ for all t, the ML estimator of the parameter vector in (A.27) is:

$$(A.28)\quad \hat{\beta}^{+}=\begin{bmatrix}\hat{\beta}_{1}\\ \hat{\beta}_{10}\\ \hat{\beta}_{20}\\ \hat{\beta}_{2}\end{bmatrix}=\begin{bmatrix}(X_1'X_1)^{-1}X_1'y_1\\ (X_2'X_2)^{-1}X_2'y_1\\ (X_1'X_1)^{-1}X_1'y_2\\ (X_2'X_2)^{-1}X_2'y_2\end{bmatrix}$$ with the Σ-matrix,

consistently estimated as:

$$(A.29)\quad \hat{\Sigma}=\begin{bmatrix}\hat{\sigma}_{11} & \hat{\sigma}_{12}\\ \hat{\sigma}_{21} & \hat{\sigma}_{22}\end{bmatrix}=\frac{1}{T}\begin{bmatrix}\hat{\varepsilon}_1'\hat{\varepsilon}_1 & \hat{\varepsilon}_1'\hat{\varepsilon}_2\\ \hat{\varepsilon}_2'\hat{\varepsilon}_1 & \hat{\varepsilon}_2'\hat{\varepsilon}_2\end{bmatrix}$$

Following T.Anderson [2], p.183, H.Cramèr [6], p.185 and P. Dhrymes [7], p. 166 the maximum likelihood estimator (A.28) is normally distributed with mean $\beta^{+}=(\beta_1,0,0,\beta_2)'$ and variance-covariance matrix (since $X_1'X_2=X_2'X_1=0$):

$$(A.30)\quad E\left[(\hat{\beta}^{+}-\beta^{+})(\hat{\beta}^{+}-\beta^{+})'\right]=$$

$$=\begin{bmatrix}\sigma_{11}(X_1'X_1)^{-1} & 0 & \sigma_{12}(X_1'X_1)^{-1} & 0\\ 0 & \sigma_{11}(X_2'X_2)^{-1} & 0 & \sigma_{12}(X_2'X_2)^{-1}\\ \sigma_{12}(X_1'X_1)^{-1} & 0 & \sigma_{22}(X_1'X_1)^{-1} & 0\\ 0 & \sigma_{12}(X_2'X_2)^{-1} & 0 & \sigma_{22}(X_2'X_2)^{-1}\end{bmatrix},$$

while the random matrix $T\hat{\Sigma}$ is independently distributed according to Wishart with parameters Σ and $n=T-k_1-k_2$ (*)

(*) This is derived from the (assumed) property that the vectors $\varepsilon(t)=\left[\varepsilon_1(t),\varepsilon_2(t)\right]'$ are mutually independent normally distributed for all t, with mean zero and variance covariance matrix Σ or the joint likelihood function of the sample considered is:

$$(A.31)\ \mathcal{L}(\varepsilon|\Sigma^{-1})=(2\pi)^{-T}|\Sigma^{-1}|^{\frac{T}{2}}\exp\left[-\frac{1}{2}\sum_{t=1}^{T}\varepsilon(t)'\Sigma^{-1}\varepsilon(t)\right]$$

$$=(2\pi)^{-T}|\Sigma^{-1}|^{\frac{T}{2}}\exp\left[-\frac{1}{2}\text{ tr }\varepsilon'\Sigma^{-1}\varepsilon\right]$$

$$=(2\pi)^{-T}|\Sigma^{-1}|^{\frac{T}{2}}\exp\left[-\frac{1}{2}\text{ tr }\Sigma^{-1}\varepsilon\varepsilon'\right]$$ with

$\varepsilon=\{\varepsilon(t)\}=(2\times T)$ matrix of error terms or transforming to the dependent variables (Jacobian = 1):

$$\mathcal{L}(Y|\Sigma^{-1},\beta^{+})=(2\pi)^{-T}|\Sigma^{-1}|^{\frac{T}{2}}\exp\left[-\frac{1}{2}\text{tr}\Sigma^{-1}(Y-XB)'(Y-XB)\right]$$

with $Y=XB+\varepsilon'$, $Y=(y_1,y_2)$, $X=(X_1,X_2)$ and $B=\begin{bmatrix}\beta_1 & \beta_{20}\\ \beta_{10} & \beta_2\end{bmatrix}$ or

$$(A.32)\ \mathcal{L}(Y|\Sigma^{-1},B)=(2\pi)^{-T}|\Sigma^{-1}|^{\frac{T}{2}}\exp\{-\frac{1}{2}\text{tr}\Sigma^{-1}\left[(Y-X\hat{B})+X(\hat{B}-B)\right]'\left[(Y-X\hat{B})+X(\hat{B}-B)\right]\}$$

$$=(2\pi)^{-T}|\Sigma^{-1}|^{\frac{T}{2}}\exp\{-\frac{1}{2}\text{tr}\Sigma^{-1}\left[(Y-X\hat{B})'(Y-X\hat{B})+(\hat{B}-B)'(X'X)(\hat{B}-B)\right]\}$$

$$=(2\pi)^{-T}|\Sigma^{-1}|^{\frac{k_1+k_2}{2}}\exp\{-\frac{1}{2}\text{tr}\Sigma^{-1}\left[(\hat{B}-B)'(X'X)(\hat{B}-B)\right]\}$$

$$|\Sigma^{-1}|^{\frac{n}{2}}\exp\left(-\frac{1}{2}\text{tr}\Sigma^{-1}A\right)$$

where $\hat{B}$ is the ML estimator of B, i.e. $\hat{B}=(X'X)^{-1}X'Y$ and $A=T\hat{\Sigma}$. So, it is verified that the kernel of the joint likelihood of the original 2-variate normal density in $\varepsilon(t)$ may be written as the product between the kernel of a multivariate pdf in $\hat{B}$ with mean B and variance-covariance matrix $\Sigma\otimes(X'X)^{-1}$ (see (A.30)) and the kernel of a Wishart pdf in $A=T\hat{\Sigma}$ ....

Then the feasible Aitken estimator (A.25) becomes:

$$(A.33)\begin{pmatrix}\beta_1^*\\ \beta_2^*\end{pmatrix}=\begin{bmatrix}\hat\beta_1-\frac{\hat\sigma_{12}}{\hat\sigma_{22}}\hat\beta_{20}\\ \hat\beta_2-\frac{\hat\sigma_{12}}{\hat\sigma_{11}}\hat\beta_{10}\end{bmatrix}$$ with mean:

$$(A.34)\ E(\beta^*)=\begin{bmatrix}E(\hat\beta_1)-E\left(\frac{\hat\sigma_{12}}{\hat\sigma_{22}}\right)E(\hat\beta_{20})\\ \\ E(\hat\beta_2)-E\left(\frac{\hat\sigma_{12}}{\hat\sigma_{11}}\right)E(\hat\beta_{10})\end{bmatrix}=\begin{bmatrix}\beta_1\\ \\ \beta_2\end{bmatrix}\quad (*),$$

which is in accordance with theorem 2.2 (see also footnote on p.6), and variance-covariance matrix:

$$(A.35)\ V(\beta^*)=\begin{bmatrix}E(\beta_1^*-\beta_1)(\beta_1^*-\beta_1)' & E(\beta_1^*-\beta_1)(\beta_2^*-\beta_2)'\\ \\ E(\beta_2^*-\beta_2)(\beta_1^*-\beta_1)' & E(\beta_2^*-\beta_2)(\beta_1^*-\beta_1)'\end{bmatrix},$$

where the occurring block matrices are defined as follows (utilizing (A.32), $X_1'X_2=X_2'X_1=0$ and the mutually independent distribution of the regression coefficients and the $\hat\sigma_{ij}$'s):

.... with variance covariance matrix Σ and degrees of freedom $n=T-k_1-k_2>0$. Hence, $\hat B$ or vec $(\hat B)=\beta^+$ and $\hat\Sigma$ are independently distributed and they are jointly sufficient for the parameters β^+ and Σ (Fisher-Neyman criterion; see e.g. P.Dhrymes [7], p.131-133).

(*) By the above mentioned independence property, also $\frac{\hat\sigma_{12}}{\hat\sigma_{22}}$ and $\hat\beta_{20}$, resp $\frac{\hat\sigma_{12}}{\hat\sigma_{11}}$ and $\hat\beta_{10}$ are mutually independently distributed, since by a change of variable, the normal-Wishart form (A.31) can be expressed as $N(\hat\beta_1,\hat\beta_{10},\hat\beta_{20},\hat\beta_2).W(\hat\sigma_{11},\hat\sigma_{22},\frac{\hat\sigma_{12}}{\hat\sigma_{22}})$ or $N.W(\hat\sigma_{11},\hat\sigma_{22},\frac{\hat\sigma_{12}}{\sigma_{11}})$ (see also (A.41)).

(A.36) $V(\beta_1^*)=E(\beta_1^*-\beta_1)(\beta_1^*-\beta_1)'=E(\hat\beta_1-\beta_1)(\hat\beta_1-\beta_1)'-$

$$E\left(\frac{\hat\sigma_{12}}{\hat\sigma_{22}}\right)E\left[(\hat\beta_1-\beta_1)\hat\beta_{20}'\right]-E\left(\frac{\hat\sigma_{12}}{\hat\sigma_{22}}\right)E\left[\hat\beta_{20}(\hat\beta_1-\beta_1)'\right]+E\left(\frac{\hat\sigma_{12}^2}{\hat\sigma_{22}^2}\right)E(\hat\beta_{20}\beta_{20}')$$

$$=\sigma_{11}(X_1'X_1)^{-1}\left[1-2E\left(\frac{\hat\sigma_{12}}{\hat\sigma_{22}}\right)\frac{\sigma_{12}}{\sigma_{11}}+E\left(\frac{\hat\sigma_{12}}{\hat\sigma_{22}}\right)^2\frac{\sigma_{22}}{\sigma_{11}}\right],$$

(A.37)
$V(\beta_2^*)=E(\beta_2^*-\beta_2)(\beta_2^*-\beta_2)'=\sigma_{22}(X_2'X_2)^{-1}\left[1-2E\left(\frac{\hat\sigma_{12}}{\hat\sigma_{11}}\right)\frac{\sigma_{12}}{\sigma_{22}}+E\left(\frac{\hat\sigma_{12}}{\hat\sigma_{11}}\right)^2\frac{\sigma_{11}}{\sigma_{22}}\right]$ and

(A.38)
$Cov(\beta_1^*,\beta_2^*)=E(\beta_1^*-\beta_1)(\beta_2^*-\beta_2)'=E(\hat\beta_1-\beta_1)(\hat\beta_2-\beta_2)'-E\left(\frac{\hat\sigma_{12}}{\hat\sigma_{22}}\right)E\left[\hat\beta_{20}(\hat\beta_2-\beta_2)'\right]$

$$-E\left(\frac{\hat\sigma_{12}}{\hat\sigma_{11}}\right)E\left[(\hat\beta_1-\beta_1)\hat\beta_{10}'\right]+E(\hat\rho^2)E(\hat\beta_{20}\hat\beta_{10}')$$

$$=E\left[(\beta_2^*-\beta_2)(\beta_1^*-\beta_1)'\right]'=\left[Cov(\beta_2^*,\beta_1^*)\right]'=0,$$

To evaluate (A.36) and (A.37), the first and second moments of the ratio of random variables $\hat\sigma_{12}$ and $\hat\sigma_{22}$, resp $\hat\sigma_{12}$ and $\hat\sigma_{11}$, have to be determined. This may be carried out by deriving the density function, say of $v=\frac{\hat\sigma_{12}}{\hat\sigma_{22}}$

Since the (2x2) positive definite matrix of random variables $a_{11}=T\hat\sigma_{11}$, $a_{12}=T\hat\sigma_{12}$ and $a_{22}=T\hat\sigma_{22}$ is Wishart distributed with covariance matrix Σ and degrees of freedom $n=T-k_1-k_2$:

$$\text{(A.39)}\quad f(A)dA=W(\Sigma,n)=\frac{|A|^{\frac{n-3}{2}}\exp(-\frac{1}{2}tr\Sigma^{-1}A)}{2^n|\Sigma|^{\frac{n}{2}}\Gamma\frac{1}{2}\Gamma\frac{n}{2}\Gamma\frac{n-1}{2}}\,dA \qquad (\star)$$

$$=c(a_{11}a_{22}-a_{12}^2)^{\frac{1}{2}(n-3)}\exp\left[\frac{1}{2(1-\rho^2)}\left(\frac{a_{11}}{\sigma_{11}}-\frac{2\rho a_{12}}{\sqrt{\sigma_{11}\sigma_{22}}}+\frac{a_{22}}{\sigma_{22}}\right)\right]da_{11}da_{22}da_{12}$$

with $c=\dfrac{1}{2^n|\Sigma|^{\frac{1}{2}n}\sqrt{\pi}\left(\frac{n-2}{2}\right)!\left(\frac{n-3}{2}\right)!}$

Rewriting the Wishart p.d.f. in terms of a_{11}, a_{22} and $v=\frac{\hat{\sigma}_{12}}{\hat{\sigma}_{22}}=\frac{a_{12}}{a_{22}}$ $(\star\star)$ so that the Jacobian of the integrand transformation is :

$$\text{(A.40)}\quad J=\frac{\partial(a_{11}a_{22}a_{12})}{\partial(a_{11}a_{22}\;v\;)}=a_{22},$$ we find:

$$\text{(A.41)}\quad W(\Sigma,n)=ca_{11}^{\frac{1}{2}(n-3)}a_{22}^{\frac{1}{2}(n-1)}\left(1-v^2\frac{a_{22}}{a_{11}}\right)^{\frac{1}{2}(n-3)}$$

$$\exp\left[-\frac{1}{2(1-\rho^2)}\left(\frac{a_{11}}{\sigma_{11}}-\frac{2\rho v a_{22}}{\sqrt{\sigma_{11}\sigma_{22}}}+\frac{a_{22}}{\sigma_{22}}\right)\right]\quad da_{11}da_{22}dv$$

$(\star)$ T.Anderson [2], p.67 and p.154.

$(\star\star)$ The following reasoning is that of the appendix in A.Zellner [28], pp.989-992.

Since $v^2\frac{a_{22}}{a_{11}}=\frac{a_{12}^2}{a_{11}a_{22}}=\frac{\sigma_{12}^2}{\sigma_{11}\sigma_{22}}=\rho^2\,(0\leq\rho^2<1)$, a binomial expansion (Newton) can be utilized:

$$(A.42)\quad \left(1-\frac{v^2 a_{22}}{a_{11}}\right)^{\frac{n-3}{2}} = \sum_{i=0}^{\infty}\frac{(-1)^i v^{2i} a_{22}^i}{a_{11}^i}\,\frac{\left(\frac{n-3}{2}\right)!}{i!\left(\frac{n-3}{2}-i\right)!}\ ,$$

or the probability density (A.41) becomes:

$$(A.43)\quad W(\Sigma,n)=c\sum_{i=0}^{\infty}\frac{(-1)^i v^{2i} a_{22}^{\,i+\frac{n-1}{2}}\left(\frac{n-3}{2}\right)!}{a_{11}^{\,i-\frac{(n-3)}{2}}\; i!\left(\frac{n-3}{2}-i\right)!}\exp\left(-\frac{a_{11}}{2(1-\rho^2)\sigma_{11}}\right)$$

$$\exp\left[-\frac{a_{22}}{2(1-\rho^2)\sigma_{22}}\left(1-\frac{2\rho v\sqrt{\sigma_{22}}}{\sqrt{\sigma_{11}}}\right)\right]\ da_{11}da_{22}dv$$

and setting $s=\frac{a_{11}}{2(1-\rho^2)\sigma_{11}}$

$$(A.44)\quad W(\Sigma,n)=c\left[2(1-\rho^2)\sigma_{11}\right]^{\frac{n-1}{2}}\left[\sum_{i=0}^{\infty}\frac{(-1)^i v^{2i} a_{22}^i}{i!\left[2(1-\rho^2)\sigma_{11}\right]^i}\right]a_{22}^{\frac{n-1}{2}}\left(\frac{n-3}{2}\right)!$$

$$\exp\left[-\frac{a_{22}}{2(1-\rho^2)\sigma_{22}}\left(1-\frac{2\rho v\sqrt{\sigma_{22}}}{\sqrt{\sigma_{11}}}\right)\right]\ da_{22}dv\ ,$$

where use has been made of the property that the Jacobian of the integrand transformation from a_{11} to s is $2\,(1-\rho^2)\sigma_{11}$ and the gamma-integral:

$$\left(\frac{n-3}{2}-i\right)! = \int_0^{\infty} s^{-(i-\frac{n-3}{2})} e^{-s}ds\ .$$

Since $\sum_{i=0}^{\infty} \frac{(-1)^i v^{2i} a_{22}^i}{i!\left[2(1-\rho^2)\sigma_{11}\right]^i} = \exp\left[-\frac{v^2 a_{22}}{2(1-\rho^2)\sigma_{11}}\right]$, we get, after

integrating out a_{22}, the marginal density of v:

$$\text{(A.45)}\; W(\Sigma,n) = c\frac{\left[2(1-\rho^2)\sigma_{11}\right]^{\frac{n-1}{2}}\left(\frac{n-3}{2}\right)!\left(\frac{n-1}{2}\right)!\left[2(1-\rho^2)\sigma_{22}\right]^{\frac{n+1}{2}}}{\left[1-2\rho v\sqrt{\frac{\sigma_{22}}{\sigma_{11}}}+v^2\frac{\sigma_{22}}{\sigma_{11}}\right]^{\frac{n+1}{2}}}\,dv$$

and substituting for c, we find:

$$\text{(A.46)}\; W(\Sigma,n) = f(v)dv = (1-\rho^2)^{\frac{n}{2}}\sqrt{\frac{\sigma_{22}}{\sigma_{11}}}\,\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n}{2}\right)}\,\frac{dv}{\frac{\sigma_{22}^{\frac{n+1}{2}}}{\sigma_{11}^{\frac{n+1}{2}}}\left(v^2-2\rho v\sqrt{\frac{\sigma_{11}}{\sigma_{22}}}+\frac{\sigma_{11}}{\sigma_{22}}\right)^{\frac{n+1}{2}}}$$

$$= (1-\rho^2)^{\frac{n}{2}}\sqrt{\frac{\sigma_{22}}{\sigma_{11}}}\,\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n}{2}\right)}\,\frac{dv}{\left(\frac{\sigma_{22}}{\sigma_{11}}\right)^{\frac{n+1}{2}}\left[\left(v-\sqrt{\frac{\sigma_{11}}{\sigma_{22}}}\rho\right)^2+\frac{\sigma_{11}}{\sigma_{22}}(1-\rho^2)\right]^{\frac{n+1}{2}}}$$

$$= \sqrt{\frac{\sigma_{22}}{(1-\rho^2)\sigma_{11}}}\,\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n}{2}\right)}\,\frac{dv}{\left[1+\frac{\left(v-\rho\sqrt{\frac{\sigma_{11}}{\sigma_{22}}}\right)^2}{(1-\rho^2)\frac{\sigma_{11}}{\sigma_{22}}}\right]^{\frac{n+1}{2}}}$$, and if

v is transformed into $z = \frac{\left(v-\rho\sqrt{\frac{\sigma_{11}}{\sigma_{22}}}\right)}{\sqrt{(1-\rho^2)\frac{\sigma_{11}}{\sigma_{22}}}}$ with Jacobian $J = \left((1-\rho^2)\frac{\sigma_{11}}{\sigma_{22}}\right)^{\frac{1}{2}}$,

the pdf of z, and hence of $v = \frac{\hat{\sigma}_{12}}{\hat{\sigma}_{22}}$, is clearly in the standardized

student t-form with n degrees of freedom:

$$(A.47)\ G(z)dz = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n}{2}\right)} \frac{dz}{(1+z^2)^{\frac{n+1}{2}}}$$ and first two moments (*)

$$(A.48)\ E(z)=0 \text{ and } E(z^2) = \frac{\Gamma\left(\frac{3}{2}\right)\Gamma\left(\frac{n-2}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n}{2}\right)} = \frac{\frac{1}{2}\Gamma\left(\frac{n-2}{2}\right)}{\frac{n-2}{2}\Gamma\left(\frac{n-2}{2}\right)} = \frac{1}{n-2}$$ or in

terms of the original variables v:

$$(A.49)\ E(v) = \rho\sqrt{\frac{\sigma_{11}}{\sigma_{22}}} \text{ and } E(v^2) = \frac{(1-\rho^2)\sigma_{11}}{(n-2)\ \sigma_{22}} + \rho^2 \frac{\sigma_{11}}{\sigma_{22}}$$

So that the exact second order block-diagononal matrix $V(\beta_1^*)$ is found from substitution of (A.49) into (A.36):

$$(A.50)\ V(\beta_1^*) = E(\beta_1^* - \beta_1)(\beta_1^* - \beta_1)' = \sigma_{11}(X_1'X_1)^{-1}$$

$$\left[1 - 2\rho\sqrt{\frac{\sigma_{11}}{\sigma_{22}}} \cdot \frac{\sigma_{12}}{\sigma_{11}} + \frac{\sigma_{11}}{\sigma_{22}}\left(\frac{1-\rho^2}{n-2} + \rho^2\right) \cdot \frac{\sigma_{22}}{\sigma_{11}}\right]$$

$$= \sigma_{11}(X_1'X_1)^{-1}(1-\rho^2)\ \frac{n-1}{n-2}\ ,$$

where $\sigma_{11}(X_1'X_1)^{-1}$ is the covariance matrix of the ML (OLS) estimator $\hat{\beta}_1$ and $\sigma_{11}(1-\rho^2)(X_1'X_1)^{-1}$ the covariance matrix of the Aitken estimator $\tilde{\beta}_1$ (see (A.26)).

Taking account of (A.50), (A.37) and (A.38), the exact variance covariance matrix of β^* becomes:

$$(A.51)\ V(\beta^*) = (1-\rho^2)\frac{n-1}{n-2} \begin{bmatrix} \sigma_{11}(X_1'X_1)^{-1} & 0 \\ 0 & \sigma_{22}(X_2'X_2)^{-1} \end{bmatrix} .$$

Computing the values of $(1-\rho^2)\ \frac{n-1}{n-2}$ for various values of n and ρ, it is seen that there emerges a considerable gain in efficiency , when deriving a feasible Aitken estimator of β in

(*) See M.Kendall and A.Stuart, [11], vol I, pp. 59-60 stead of its OLS-estimator, if $\rho>0.30$ (obviously, the gain becomes more considerable if $n = T-k_1-k_2$ increases) (*) Δ

3.2 Exact sample distribution of the feasible Aitken estimator β^*

Theorem A2

The finite sample pdf of the feasible Aitken estimator β_1^* of β_1 satisfies:

$$(A.52)\quad h(\beta_1^*)d\beta_1^* \approx h(z_1)dz_1 = \frac{1}{\left[2\pi\sigma_{11}(1-\rho^2)\right]^{\frac{k_1}{2}}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}$$

$$\sum_{i=0}^{\infty} \frac{(-1)^i}{i!}\left[\frac{z_1' z_1}{2(1-\rho^2)\sigma_{11}}\right]^i \frac{\Gamma\left(i+\frac{n+k_1}{2}\right)}{\Gamma\left(i+\frac{n+k_1+1}{2}\right)} \, d\, z_1$$

with the $(k_1\times 1)$-vector z_1 equal to: $z_1=\hat{\beta}_1-\beta_1-\frac{\hat{\sigma}_{12}}{\hat{\sigma}_{22}}\hat{\beta}_{20}=\beta_1^*-\beta_1$ and $n=T-k_1-k_2$ and the matrix X_1 is assumed to consist of a set of mutually orthonormal vectors or $X_1'X_1=I_{k_1}$.

Proof

From (A.32) and (A.46), the joint pdf of $\hat{\beta}_1,\hat{\beta}_{20}$ and $\frac{\hat{\sigma}_{12}}{\hat{\sigma}_{22}}$ can be written as:

$$(A.53)\quad h\left(\hat{\beta}_1,\hat{\beta}_{20},\frac{\hat{\sigma}_{12}}{\hat{\sigma}_{22}}\right)d\hat{\beta}_1\, d\hat{\beta}_{20}\, d\frac{\hat{\sigma}_{12}}{\hat{\sigma}_{22}}=g(\hat{\beta}_1,\hat{\beta}_{20})\, f\left(\frac{\hat{\sigma}_{12}}{\hat{\sigma}_{22}}\right)d\hat{\beta}_1\, d\hat{\beta}_{20}\, d\frac{\hat{\sigma}_{12}}{\hat{\sigma}_{22}} ,$$

where $g(.)$ is $2k_1$-variate normal and $f\left(\frac{\hat{\sigma}_{12}}{\hat{\sigma}_{22}}\right)=f(v)$ is given by (A.46).

Transforming to the random variables.

$z_1=\hat{\beta}_1-\beta_1-v\hat{\beta}_{20}$, $z_2=\hat{\beta}_{20}$ and $z_3=v$, where z_1 and z_2 are k_1-vectors and z_3 is a scalar random variable and the Jacobian is 1, (A.53) becomes:

(*) See W.Vandaele [23], §4.3. If ρ is very small ($\rho\leq 0.10$ à 0.20) OLS has to be preferred relative to feasible Aitken estimation. Only for large samples, the covariance matrices are(approximately) equal (asymptotic equality for $\rho=0$).

(A.54) $h(z_1,z_2,z_3)dz_1dz_2dz_3=g(z_1+z_2z_3,z_2)f(z_3)dz_1dz_2dz_3$, where

(A.55) $$g(z_1+z_2z_3,z_2)=g(\hat\beta_1-\beta_1,\hat\beta_{20})=\frac{\left|E\begin{pmatrix}\hat\beta_1-\beta_1\\ \hat\beta_{20}\end{pmatrix}\begin{pmatrix}\hat\beta_1-\beta_1\\ \hat\beta_{20}\end{pmatrix}'\right|^{-\frac{1}{2}}}{(2\pi)^{k_1}}$$

$$\exp\left\{-\frac{1}{2}\begin{pmatrix}\hat\beta_1-\beta_1\\ \hat\beta_{20}\end{pmatrix}'\left[E\begin{pmatrix}\hat\beta_1-\beta_1\\ \hat\beta_{20}\end{pmatrix}\begin{pmatrix}\hat\beta_1-\beta_1\\ \hat\beta_{20}\end{pmatrix}'\right]^{-1}\begin{pmatrix}\hat\beta_1-\beta_1\\ \hat\beta_{20}\end{pmatrix}\right\}$$ where

(A.56) $$\left[E\begin{pmatrix}\hat\beta_1-\beta_1\\ \hat\beta_{20}\end{pmatrix}\begin{pmatrix}\hat\beta_1-\beta_1\\ \hat\beta_{20}\end{pmatrix}'\right]^{-1}=\begin{bmatrix}\sigma_{11}(X_1'X_1)^{-1} & \sigma_{12}(X_1'X_1)^{-1}\\ \sigma_{12}(X_1'X_1)^{-1} & \sigma_{22}(X_1'X_1)^{-1}\end{bmatrix}^{-1}=$$

$\Sigma^{-1}\otimes(X_1'X_1)$ (see (A.30)), with determinant $|\Sigma|^{-k_1}|X_1'X_1|^2$.

Without loss of generality, we may assume that the columm vectors of X_1 are mutually orthonormal or $X_1'X_1=I_{k_1}$ so that substituting (A.56)and(A.46)into(A.54):

(A.57) $$h(z_1,z_2,z_3)dz_1dz_2dz_3=\frac{c\ \exp\left\{-\frac{1}{2}\left[\sigma^{11}(z_1+z_2z_3)'(z_1+z_2z_3)+2\sigma^{12}z_2'(z_1+z_2z_3)+\sigma^{22}z_2'z_2\right]\right\}}{\left(1-2\rho\sqrt{\frac{\sigma_{22}}{\sigma_{11}}}\ z_3+\frac{\sigma_{22}}{\sigma_{11}}\ z_3^2\right)^{\frac{n+1}{2}}}dz_1dz_2dz_3$$

with c equal to $c=\frac{1}{(2\pi)^{k_1}}\left(\sigma_{11}\sigma_{22}-\sigma_{12}^2\right)^{\frac{-k_1}{2}}(1-\rho^2)^{\frac{n}{2}}\sqrt{\frac{\sigma_{22}}{\sigma_{11}}}\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n}{2}\right)}$

and rewiting as:

(A.58) $h(z_1, z_2, z_3)dz_1dz_2dz_3 =$

$$\frac{c\ \exp\left(-\frac{1}{2}z_1'z_1\sigma^{11}\right)\ \exp\{-\frac{1}{2}\left[z_2'z_2(z_3^2\sigma^{11}+2z_3\sigma^{12}+\sigma^{22})+2z_2'(z_3\ z_1\sigma^{11}+z_1\sigma^{12})\right]\}}{\left(1-2\rho\sqrt{\frac{\sigma_{22}}{\sigma_{11}}}z_3+\frac{\sigma_{22}}{\sigma_{11}}z_3^2\right)^{\frac{1}{2}(n+1)}}$$

$dz_1dz_2dz_3$,

we can integrate out z_2, by rewriting the part of the above pdf containing z_2 with substitutions $a_1 = z_3^2\ \sigma^{11}+2z_3\sigma^{12}+\sigma^{22}$ and $a_2 = z_3z_1\sigma^{11}+z_1\sigma^{12}$, where a_1 is a scalar random variable and a_2 a k_1- variate random vector as:

$$\text{(A.59)} \int_{-\infty}^{+\infty} \exp\left\{-\frac{a_1}{2}\left(z_2'z_2+\frac{2z_2'a_2}{a_1}\right)\right\}dz_2 = \int_{-\infty}^{+\infty} \exp\left\{-\frac{a_1}{2}\left[\left(z_2+\frac{a_2}{a_1}\right)'\left(z_2+\frac{a_2}{a_3}\right)-\frac{a_2'a_2}{a_1^2}\right]\right\}dz_2 =$$

$$= \left(\frac{2\pi}{a_1}\right)^{\frac{k_1}{2}} \exp\left(\frac{a_2'a_2}{2a_1}\right)$$

So that the joint pdf of z_1 and z_3 becomes:

(A.60) $h(z_1, z_3)dz_1dz_3 =$

$$\frac{c(2\pi)^{\frac{k_1}{2}}\ \sigma_{11}^{\frac{1}{2}(n+1)}\ \exp\left(\frac{-z_1'z_1\sigma^{11}}{2}\right)\exp\left[\frac{z_1'z_1(z_3\sigma^{11}+\sigma^{12})^2}{2(z_3^2\sigma^{11}+2z_3\sigma^{12}+\sigma^{22})}\right]}{\left(z_3^2\sigma^{11}+2z_3\sigma^{12}+\sigma^{22}\right)^{\frac{k_1}{2}}\ (z_3^2\sigma_{22}-2z_3\sigma_{12}+\sigma_{11})^{\frac{1}{2}(n+1)}}\ dz_1\ dz_3$$

$$= \frac{c(2\pi)^{\frac{k_1}{2}}\ \sigma_{11}^{\frac{1}{2}(n+1)}\ \exp\{\frac{-z_1'z_1}{2(z_3^2\sigma_{22}-2z_3\ \sigma_{12}+\sigma_{11})}\}}{(\sigma_{11}\sigma_{22}-\sigma_{12}^2)^{\frac{-k_1}{2}}(z_3^2\ \sigma_{22}-2z_3\sigma_{12}+\sigma_{11})^{\frac{1}{2}(n+k_1+1)}}\ dz_1\ dz_3$$

Expanding the exponent term in a Mac Laurin series and integrating term by term w.r.t. z_3, utilizing:

$$\text{(A.61)} \int_{-\infty}^{\infty} \frac{dz_3}{(z_3^2\sigma_{22}-2z_3\sigma_{12}+\sigma_{11})^{i+\frac{1}{2}(n+k_1+1)}} = \frac{1}{\sigma_{22}^{\frac{1}{2}}\left[\sigma_{11}(1-\rho^2)\right]^{i+\frac{1}{2}(n+k_1)}} \cdot$$

$$\frac{\Gamma(\frac{1}{2})\Gamma\left(i+\frac{n+k_1}{2}\right)}{\Gamma\left(i+\frac{n+k_1+1}{2}\right)} \qquad (\text{ since } \sigma_{12}^2<\sigma_{22}\sigma_{11})$$

we obtain the following joint pdf for the elements of

$$z_1=\hat{\beta}_1-\beta_1-\frac{\hat{\sigma}_{12}}{\hat{\sigma}_{22}}\hat{\beta}_{20}=\beta_1^{*}-\beta_1:$$

$$\text{(A.62)}\ h(z_1)dz_1=$$

$$=\frac{c(2\pi)^{\frac{k_1}{2}}\sigma_{11}^{\frac{1}{2}(n+1)}}{\left(\sigma_{11}\sigma_{22}-\sigma_{12}^2\right)^{\frac{-k_1}{2}}}\sum_{i=0}^{\infty}\frac{(-1)^i}{i!}\left(\frac{z_1'z_1}{2}\right)^i\int_{-\infty}^{\infty}\frac{dz_3}{(z_3^2\sigma_{22}-2z_3\sigma_{12}+\sigma_{11})^{i+\frac{1}{2}(n+k_1+1)}}dz_1$$

$$=\frac{c(2\pi)^{\frac{k_1}{2}}\sigma_{11}^{\frac{1}{2}(n+1)}\sigma_{22}^{-\frac{1}{2}}\Gamma(\frac{1}{2})}{(\sigma_{11}\sigma_{22}-\sigma_{12}^2)^{\frac{-k_1}{2}}\left[\sigma_{11}(1-\rho^2)\right]^{\frac{1}{2}(n+k_1)}}\sum_{i=0}^{\infty}\frac{(-1)^i}{i!}\left(\frac{z_1'z_1}{2\sigma_{11}(1-\rho^2)}\right)^i\frac{\Gamma\left(i+\frac{n+k_1}{2}\right)}{\Gamma\left(i+\frac{n+k_1+1}{2}\right)}dz_1$$

or substituting for the constant c, given in (A.57), we find as exact pdf of z_1:

$$\text{(A.63)}\ h(z_1)dz_1=$$

$$\frac{1}{\left[2\pi\sigma_{11}(1-\rho^2)\right]^{\frac{k_1}{2}}}\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}\sum_{i=0}^{\infty}\frac{(-1)^i}{i!}\left(\frac{z_1'z_1}{2\sigma_{11}(1-\rho^2)}\right)^i\frac{\Gamma\left(i+\frac{n+k_1}{2}\right)}{\Gamma\left(i+\frac{n+k_1+1}{2}\right)}dz_1$$

where $\sigma_{11}(1-\rho^2)$ is the asymptotic variance of each element of z_1 (see (A.26)), which are all equal since it is assumed that $X_1'X_1=I_{k_1}$.

As n increases, the ratio of the Gamma functions involved in (A.63) will rapidly disappear so that the pdf of z_1 approaches the following normal density:

$$\text{(A.64)}\ h(z_1)\,dz_1=c'\exp\left\{\frac{-z_1'z_1}{2\sigma_{11}(1-\rho^2)}\right\}dz_1 \qquad \Delta$$

Appendix B

B.1 Finite sample bias of the ML estimator of the serial correlation coefficient.

Theorem B.1

The finite sample bias of the ML estimator for the autoregression parameter β in the model:

(B.1) $y_t=\beta y_{t-1}+\varepsilon_t$ with $\varepsilon_t \sim NID(0,\sigma^2)$ $(t=1,2,3,...,T)$ (*)

is equal to:

(B.2a) $E(\hat{\beta})-\beta=\frac{-2}{T}\beta+0(\beta T^{-2})$ (**) if the initial value y_o is fixed at zero or

(B.2b) $E(\hat{\beta})-\beta=-\frac{2}{T+1+c^2}\beta+0(\beta T^{-2})$ if the initial value y_o is fixed at a constant value $c \neq 0$ and

(B.3) $E(\hat{\beta})-\beta=-\frac{2}{T+1}\beta+0(\beta T^{-2})$ if the initial value y_o is a random variable with the same mean and variance as the other y_t - variables.

Proof

Following J.White [25], the expansions for the mathematical expectation will be given up to terms of order T^{-3} and β^4.

Model 1 The initial value y_o is assumed constant: $y_o=c$

Then, under the above assumptions, the ML-estimator of β in (B.1) results from the unconstrained maximization of the logarithmic transformation of the joint likelihood density of **y**:

(B.4) $L_1(y|\beta,\sigma^2) = (2\pi\sigma^2)^{-\frac{1}{2}T} \exp\{-\frac{1}{2\sigma^2}\sum_{t=1}^{T}(y_t-\beta y_{t-1})^2\}$,

It is clear that for this model the ML-estimator is equal to the OLS estimator:

(*) The discussion of this theorem follows J.White [25]. For simplicity of notation and discussion, we assume that the initial value is y_o and not y_1.

(**) See H.Cramèr [6], p.122, for the determination of the order of magnitude in probability of different functions.

(B.5) $$\hat{\beta}=\frac{\sum_{t=1}^{T} y_t y_{t-1}}{\sum_{t=1}^{T} y_{t-1}^2}$$

1.A $y_o=c=0$

Setting $U=\sum_{t=1}^{T} y_t y_{t-1}$ and $V=\sum_{t=1}^{T} y_{t-1}^2$ the joint moment generating function of U and V is:

(B.6) $$E_1\left[\exp(Uu+Vv)\right]=M_1(u,v)=\int_{-\infty}^{+\infty}\exp(Uu+Vv)\; L_1(y)dy$$

$$=(2\pi)^{-\frac{1}{2}T}\exp\int_{-\infty}^{+\infty}(-\tfrac{1}{2}y'Dy)dy=|D|^{-\frac{1}{2}} \qquad (\star) \quad ,$$

where D is a (T×T)-matrix with determinant

$$|D|=\begin{vmatrix} p & q & 0 & 0 & \cdots & 0 & 0 & 0\\ q & p & q & 0 & \cdots & 0 & 0 & 0\\ 0 & q & p & q & \cdots & 0 & 0 & 0\\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots\\ 0 & 0 & 0 & 0 & \cdots & q & p & q\\ 0 & 0 & 0 & 0 & \cdots & 0 & q & 1 \end{vmatrix}$$

with $p=1+\beta^2-2v$ and $q=-(\beta+u)$.

Expanding

(B.7) $$\frac{E_1(\hat{\beta})}{\beta}=\frac{1}{\beta}\int_{-\infty}^{o}\left.\frac{\partial M_1(u,v)}{\partial u}\right|_{u=o}dv \;=-\tfrac{1}{2}\int_{-\infty}^{o}\frac{1}{\beta}|D|^{-\frac{3}{2}}\left.\frac{\partial D}{\partial u}\right|_{u=o}dv$$

in a Mac Laurin series and setting $\alpha=\beta^2$:

(B.8) $$\frac{E_1(\hat{\beta})}{\beta}=Q_1(\alpha)=Q_1(o)+Q_1'(o)\alpha+Q_1''(o)\frac{\alpha^2}{2}+\ldots\ldots,$$ with

(B.9a) $$Q_1'(o)=-\frac{1}{2}\int_{-\infty}^{o}\frac{\partial}{\partial\alpha}\left(\frac{1}{\beta}|D|^{-\frac{3}{2}}\left.\frac{\partial D}{\partial u}\right|_{u=o}\right)dv$$ and

(B. 9b) $$Q_1''(o)=-\frac{1}{2}\int_{-\infty}^{o}\frac{\partial^2}{\partial\alpha^2}\left(\frac{1}{\beta}|D|^{-\frac{3}{2}}\left.\frac{\partial D}{\partial u}\right|_{u=o}\right)dv$$

($\star$) Where, without loss of generality, σ^2 is set equal to 1, since $\hat{\beta}$ is independent of σ^2.

From (B.6), we may denote the $(T\times T)$ matrix $D=D(T)$ and the right-lower submatrices with $D(T-1), D(T-2)$ etc.. so that $D(T)$ satisfies the second order difference equation:

(B.10) $D(T)=pD(T-1)-q^2 D(T-2)$ with $D(1)=1$ and $D(2)=p-q^2$,
and solution:

$$\text{(B.11)}\; D=D(T)=p^{T-2}\sum_{\tau=0}^{T-1}(-1)^{\tau}\left[p\binom{T-\tau-1}{\tau}-q^2\binom{T-\tau-2}{\tau}\right]p^{-2\tau}q^{2\tau}$$

$$=\sum_{\tau=0}^{T-1}(-1)^{\tau}\left[(z+\alpha)\binom{T-\tau-1}{\tau}-x\binom{T-\tau-2}{\tau}\right](z+\alpha)^{T-2-2\tau}x^{\tau} \quad \text{with}$$

$$z=1-2v=p-\beta^2 \qquad \text{and } x=(\beta+u)^2=q^2 ,$$

so that the derivatives involved in (B.9) and (B.10) may be evaluated by means of the various values of $D(i,j)$:

$$\text{(B.12)}\; D(i,j)=D(T,i,j)=\frac{\partial}{\partial\alpha^j}\left(\frac{\partial D(T)}{\partial x^i}\bigg|_{x=\alpha}\right)\bigg|_{\alpha=0} \qquad \text{since}$$

$$\text{(B.13)}\; \frac{1}{\beta}\frac{\partial D(T)}{\partial u}\bigg|_{u=0}=\frac{1}{\beta}\frac{\partial D(T)}{\partial x}\frac{\partial x}{\partial u}\bigg|_{u=0}=\frac{2(\beta+u)}{\beta}\frac{\partial D(T)}{\partial x}\bigg|_{u=0}=2\frac{\partial D(T)}{\partial x}\bigg|_{x=\alpha}$$

$$\text{and } \quad D(T)\big|_{u=0} = D(T)\big|_{x=\alpha=\beta^2}$$

So,

$$Q_1'(0)=\int_{-\infty}^{0}\{\tfrac{3}{2}D(T,0,0)^{-\frac{5}{2}}D(T,0,1)D(T,1,0)-2D(T,0,0)^{\frac{3}{2}}D(T,1,1)\}\,dv$$

$$=\int_{-\infty}^{0}\{\tfrac{3}{2}z^{-\frac{5}{2}(T-1)}(T-2)z^{T-3}(z-1)z^{T-3}(2-T-z)-2z^{\frac{3}{2}(T-1)}z^{(T-5)}\left[(T-3)(T-4)-(T-2)z^2-(T-3)(T-4)z\right]\}\,dv$$

$$\text{(B.14)}=\int_{-\infty}^{0}\{-\frac{(T-2)}{2}z^{-\frac{1}{2}(T+3)}-\frac{(T+2)(T-3)}{2}z^{-\frac{1}{2}(T+5)}+\frac{T^2+2T-12}{2}z^{-\frac{1}{2}(T+7)}\}\,dv$$

$$=\frac{12}{(T+1)(T+3)(T+5)}=\frac{12}{T^3}+0(T^{-4}) \quad \text{and similarly,}$$

$$(B.15)\ Q_1(0)=\frac{T^2-2T+3}{(T-1)(T+1)}=1-\frac{2}{T}+\frac{4}{T^2}-\frac{2}{T^3}+0(T^{-4})$$ and

$$(B.16)\ Q_1''(0)=\frac{36(T+8)}{(T+3)(T+5)(T+7)(T+9)}=\frac{36}{T^3}+0(T^{-4})$$ or substituting into (B.8):

$$(B.17)\ E(\hat{\beta})=\left(1-\frac{2}{T}+\frac{4}{T^2}-\frac{2}{T^3}\right)\beta+\frac{12}{T^3}\beta^3+\frac{18}{T^3}\beta^5+....=(1-\frac{2}{T})\beta+0(\beta T^{-2}) \quad \Delta$$

1B $y_0=c\neq 0$ (c is known)

Then the joint moment generating function (B.6) of the composite variates U and V becomes:

$$(B.18)\ M(u,v)=|D(T)|^{-\frac{1}{2}}\exp\left\{\frac{c^2}{2}\left[1-\frac{D(T+1)}{D(T)}\right]\right\}$$ with the first term

of the Mac Laurin series expansion (B.8) of $\frac{E(\hat{\beta})}{\beta}=Q_c(\alpha)$:

$$(B.19)\ Q_c(0)=\int_{-\infty}^{0}\frac{1}{\beta}\left.\frac{\partial M(u,v)}{\partial u}\right|_{u=0;\,\beta=0}dv$$ (and integrand transformation)

$$=\frac{e^{\frac{1}{2}c^2}}{2}\int_1^{\infty}e^{-\frac{1}{2}c^2z}\left[z^{-\frac{1}{2}(T+1)}+(T-2+c^2)z^{-\frac{1}{2}(T+3)}\right]dz \quad (z=1-2v)$$

If $a=\frac{1}{2}(T+1)$ and $x=\frac{1}{2}c^2$, we can integrate part by part to obtain:

$$(B.20)\ Q_c(0)=\tfrac{1}{2}e^x\left[x^{a-1}\Gamma(1-a,x)-\frac{x^a}{a}(2a+2x-3)\Gamma(1-a,x)+(2a+2x-3)\frac{e^{-x}}{a}\right],$$

with $\Gamma(1-a,x)$ the incomplete Gamma function (*)

$$(B.21)\ \Gamma(1-a,x)=\frac{e^{-x}x^{1-a}}{x+a}\left[1-\frac{a}{(x+a)^2}+\frac{2a}{(x+a)^3}+0(a^2(x+a)^{-4})\right]$$ or (B.20) becomes:

(*) This asymptotic expansion of the incomplete Gamma function $\Gamma(1-a,x)=\int_x^{\infty}e^{-u}u^{-a}du$ (if x=0, usual Gamma function) can be found in Erdélyi, Higher Transcendental Functions, 2, New York, Mc.Graw Hill, 1953,p.140.

$$(B.22)\ Q_c(0)=\tfrac{1}{2}\left(\frac{2a+2x-2}{x+a}\right)+0\left(\frac{T^2}{(T+c^2)^4}\right)=1-\frac{2}{T+1+c^2}+0\left(\frac{T^2}{(T+c^2)^4}\right)\quad \text{or}$$

$$(B.23)\ E(\hat{\beta})=\left(1-\frac{2}{T+1+c^2}\right)\beta+\dots\dots$$

From which it is seen that irrespectrive of the remaining terms of $Q_c(\alpha)$, the bias of $\hat{\beta}$ vanishes if the initial known constant $|y_o|=|c|$ is large. Δ

<u>Model 2</u> <u>Stationarity condition: y_o is random with same marginal distribution as y_t.</u>

If (B.1) is assumed to satisfy an infinite stationary process:

$$(B.24)\ y_t=\beta y_{t-1}+\varepsilon_t=\sum_{\tau=0}^{\infty}\beta^{\tau}\varepsilon_{t-\tau}\ \text{ with }\ \varepsilon_{t-\tau}\approx NID(0,\sigma^2)\ \text{ and }\ -1<\beta<1,$$

then

$$(B.25)\ \operatorname{var}(y_t)=\sum_{\tau=0}^{\infty}\beta^{2\tau}\operatorname{var}(\varepsilon_{t-\tau})=\frac{\sigma^2}{1-\beta^2}\ ,\quad \text{so}$$

$$(B.26)\ y_t\approx N\left(0,\frac{\sigma^2}{(1-\beta^2)}\right)\ \text{ and }\ y_o\approx N\left(0,\frac{\sigma^2}{(1-\beta^2)}\right)\qquad (*)$$

or the probability of obtaining a y_o-variable is equal to:

$$(B.27)\ L_2(y_o|\beta,\sigma^2)=\frac{(1-\beta^2)^{\frac{1}{2}}}{\sigma\sqrt{2\pi}}\exp\left[-\frac{(1-\beta^2)}{2\sigma^2}y_o^2\right]$$

and since the ε_t's (t=1,...,T), are mutually independently normally distributed, the joint sample likelihood function is (Jacobian of transformation of ε_t-variables to y_t variables is unity):

$$(B.28)\ L_2(y|\beta,\sigma^2)=(2\pi\sigma^2)^{-\frac{1}{2}(T+1)}(1-\beta^2)^{\frac{1}{2}}\exp\left\{-\frac{1}{2\sigma^2}\left[(1-\beta^2)y_o^2+\sum_{t=1}^{T}(y_t-\beta y_{t-1})^2\right]\right\}$$

(*) For model 1 ($y_o=c$), the marginal probability distribution of each observation y_t depends critically on the complete history of the stochastic process if $|\beta|>1$ (always possible since convergence is not postulated for model 1 and estimator (B.5) can easily satisfy: $|\hat{\beta}|>1$). For this model 2 $y_o=\dfrac{\varepsilon_o}{(1-\beta^2)^{\frac{1}{2}}}$ with $|\beta|<1$.

Logarithmizing (B.28), necessary (first order) conditions for the maximum of the likelihood function are provided by:

$$(B.29)\quad \left.\frac{\partial \ln L_2(y'|\beta,\sigma^2)}{\partial\beta}\right|_{\beta=\hat{\beta}} = -\frac{\hat{\beta}}{(1-\hat{\beta}^2)} + \frac{1}{\hat{\sigma}^2}\left[y_o^2\hat{\beta} + \sum_{t=1}^{T}(y_t-\hat{\beta}y_{t-1})y_{t-1}\right] = 0 \text{ and}$$

$$(B.30)\quad \left.\frac{\partial \ln L_2(y'|\beta,\sigma^2)}{\partial\sigma^2}\right|_{\sigma^2=\hat{\sigma}^2} = -\frac{(T+1)}{\hat{\sigma}^2} + \frac{1}{\hat{\sigma}^4}\left[(1-\hat{\beta}^2)y_o^2 + \sum_{t=1}^{T}(y_t-\hat{\beta}y_{t-1})^2\right] = 0$$

or

$$(B.31)\quad \sum_{t=1}^{T} y_t y_{t-1} = \hat{\beta}\sum_{t=1}^{T} y_{t-1}^2 - y_0^2\hat{\beta} + \frac{\hat{\sigma}^2\hat{\beta}}{(1-\hat{\beta}^2)} \quad \text{and}$$

$$(B.32)\quad \hat{\sigma}^2 = \frac{1}{T+1}\left[(1-\hat{\beta}^2)y_0^2 + \sum_{t=1}^{T}(y_t-\hat{\beta}y_{t-1})^2\right] = \frac{1}{T+1}(A-2\hat{\beta}B+\hat{\beta}^2C),$$

with

$$(B.33)\quad A = \sum_{t=0}^{T} y_t^2 \;,\quad B = \sum_{t=1}^{T} y_t y_{t-1} \quad \text{and} \quad C = \sum_{t=2}^{T} y_{t-1}^2 \;.$$

Substituting (B.32) into (B.31), and taking account of (B.33), yields:

$$(B.34)\quad B = \hat{\beta}C + \frac{1}{T+1}(A-2\hat{\beta}B+\hat{\beta}^2C)\,\frac{\hat{\beta}}{(1-\hat{\beta}^2)}\quad,$$

and rearranging terms according to the power of $\hat{\beta}$:

$$(B.35)\quad g(\hat{\beta}) = \hat{\beta}^3 - \frac{(T-1)}{T}\,\frac{B}{C}\hat{\beta}^2 - \frac{[A+(T+1)C]}{TC}\hat{\beta} + \frac{(T+1)}{T}\,\frac{B}{C} = 0 \;.$$

One root of this cubic equation in $\hat{\beta}$ is the real maximum likelihood estimator, say $\hat{\hat{\beta}}$, of model 2. We may try to investigate it by locating the three roots of $g(\hat{\beta})$ (*). Therefore, from (B.35), we easily determine values of $g(\hat{\beta})$ for the points $\hat{\beta} = \pm \infty$ and $\hat{\beta} = \pm 1$ successively:

(*) The roots may be determined analytically, but this is very cumbersome and does not produce much contribution for understanding. Indeed, denoting the coefficients of the cubic equations (B.35) by $a_1 = -\frac{T-1}{T}\frac{B}{C}$,

$a_2 = -\frac{[A+(T+1)C]}{TC}$ and $a_3 = \frac{T+1}{T}\frac{B}{C}$, the roots, say $\hat{\beta}_1, \hat{\beta}_2$ and $\hat{\beta}_3$ may be given by:

(B.36) $\hat{\beta}_1 = l_1^{1/3} + l_2^{1/3}$; $\hat{\beta}_2 = a_1 l_1^{1/3} + a_2 l_2^{1/3}$ and $\hat{\beta}_3 = a_2 l_1^{1/3} + a_1 l_2^{1/3}$,

where

(B.37) $l_{1,2} = -\frac{k_1}{2} \pm \sqrt{\frac{1}{4}k_2^2 + \frac{1}{27}k_1^3}$

with

(B.38a) $k_1 = -\frac{1}{3}a_1^2 + a_2 = -\frac{1}{3}\frac{(T-1)^2}{T^2}\frac{B^2}{C^2} - \frac{A+(T+1)C}{TC}$ and

(B.38b) $k_2 = \frac{2}{27}a_1^3 - \frac{1}{3}a_1 a_2 + a_3 = \frac{2}{27}\left(\frac{T+1}{T}\right)^3 \frac{B^3}{C^3} - \frac{1}{3}\frac{(T-1)[A+(T+1)C]}{T^2}\frac{B}{C^2} + \frac{T+1}{T}\frac{B}{C}$.

Clearly, $\hat{\beta}_{1,2}$ are real if $\frac{1}{4}k_2^2 + \frac{1}{27}k_1^3 \geq 0$, but as it is implied by (B.36-38), exact computation does not gain much comprehensing about the approximate numerical value of the roots.

$$g(\infty) = \infty \quad \text{and} \quad g(-\infty) = -\infty$$

(B.39)

$$g(1) = -\frac{(A+C)}{TC} < 0 \text{ and } g(-1) = \frac{A+C}{TC} > 0.$$

or the three roots of (B.35) satisfy:

(B.40) $\hat{\beta}_1 < -1 < \hat{\beta}_2 < +1 < \hat{\beta}_3$ ($\hat{\beta}_{1,2,3}$ in different notation as (B.36)),

where, due to the presupposed stationarity condition for this model 2, $\hat{\beta}_2 = \hat{\hat{\beta}}$ (unique maximum likelihood estimator lying in feasible region $(-1,+1)$).
For $T \to \infty$, the cubic equation (B.35) tends to:

(B.41) $g(\hat{\beta}) \approx \hat{\beta}^3 - \frac{B}{C}\hat{\beta}^2 - \hat{\beta} + \frac{B}{C} = (\hat{\beta} - \frac{B}{C})(\hat{\beta}^2 - 1) = 0$, so that the three

roots of $g(\hat{\beta})$ are asymptotically $b_1 = 1, b_2 = -1$ and $b_3 = \frac{B}{C}$, or in comparison with (B.40), the ML estimator for model 2 is asymptotically:

(B.42) $\hat{\hat{\beta}} = b_3 = \frac{B}{C} = \frac{\sum_{t=1}^{T} y_t y_{t-1}}{\sum_{t=2}^{T} y_{t-1}^2}$, so that in conjunction with (B.5),

it is found that the ML-estimators of both models 1 and 2 only differ by a term y_o^2 in the denominator (*).

For $U = \sum_{t=1}^{T} y_t y_{t-1}$ and $V = \sum_{t=2}^{T} y_{t-1}^2$, the joint moment

generating function (B.6) becomes for this model:

(B.43) $E_2\left[\exp(Uu+Vv)\right] = M_2(U,V) = \int_{-\infty}^{+\infty} \exp(Uu+Vv) L_2(y)\, dy$

(*) Note that the ML estimator (B.42) can directly be derived as a weighted least squares estimator minimizing

$$Q_2 = (1-\beta^2) y_0^2 + \sum_{t=1}^{T} (y_t - \beta y_{t-1})^2 = A - 2\beta B + \beta^2 C \quad \text{w.r.t. } \beta \text{ ((B.5)}$$

is the unweighted least squares estimator).

$$=(2\pi)^{-\frac{1}{2}(T+1)}(1-\beta^2)^{\frac{1}{2}}\int_{-\infty}^{+\infty}\exp(-\tfrac{1}{2}y'D^{\star}y)\,dy$$

$$=(1-\beta^2)^{\frac{1}{2}}|D^{\star}|^{-\frac{1}{2}} \quad \text{(see (B.28) for } \sigma^2=1) ,$$

with $D^{\star}$ a $(T+1)\times(T+1)$-matrix with determinant

$$\text{(B.44)} \quad D^{\star}=D(T+1)=\begin{vmatrix} p-\alpha & q & 0 & \cdots & 0 & 0 & 0 \\ q & p & q & \cdots & 0 & 0 & 0 \\ 0 & q & p & \cdots & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdots & q & p & q \\ 0 & 0 & 0 & \cdots & 0 & q & 1 \end{vmatrix}$$

with $p=1+\beta^2-2v$, $q=-(\beta+u)$ and $\alpha=\beta^2$.
Expanding, as for (B.7), $D^{\star}(T+1)$ by the elements of the first row:

(B.45) $D^{\star}(T+1)=(p-\alpha)D(T)-q^2D(T-1)$, with $D(T)$ defined as in model 1 (see (B.11)). Combining (B.45) and (B.11) with $T'=T+1$, we find:

(B.46) $D^{\star}(T+1)=D^{\star}(T')=(p-\alpha)D(T'-1)-q^2D(T'-2)=D(T')-\alpha D(T'-1)$,
and defining:

$$\text{(B.47)} \quad D^{\star}(T',i,j)=\frac{\partial}{\partial\alpha^j}\left(\left.\frac{\partial D(T')}{\partial x^i}\right|_{x=\alpha}\right)\Bigg|_{\alpha=0} ,$$

while expanding the integrands in their Mac Laurin series as in model 1, we obtain:

$$\text{(B.48)} \quad \frac{E_2(\hat{\beta})}{\beta}=Q_2(0)+Q_2'(0)\alpha+Q_2''(0)\frac{\alpha^2}{2}+\ldots\ldots\ \text{or}$$

$$\text{(B.49)} \quad E_2(\hat{\hat{\beta}})=\left(1-\frac{2}{(T+1)}+\frac{4}{(T+1)^2}-\frac{2}{(T+1)^3}\right)\beta+\frac{2\beta^3}{(T+1)^2}+\frac{2\beta^5}{(T+1)^2}+\ldots.=$$

$$(1-\frac{2}{T+1})\beta+O(\beta T^{-2}) \qquad \Delta$$

B.2 The bias and inconsistency of the O.L.S. autoregression estimators in autoregressive models.

B.2.1 First order autoregression and autocorrelation without exogenous variables

Theorem B.2

If the i'th equation of an M-equation model satisfies:

$$(B.50)\quad y_i(t)=\gamma_{i1}y_i(t-1)+\varepsilon_i(t) \text{ with } \varepsilon_i(t)=\rho_i\varepsilon_i(t-1)+\eta_i(t)\ \forall_t\ 1,2,\ldots,T$$

with $\eta_i(t)\tilde{}(0,\sigma_{ii})$ and $y_i(0)=\varepsilon_i(0)=0$ or stochastic, then the OLS estimators of γ_{i1} and ρ_i are inconsistent.

Proof 1. The OLS-estimator of γ_{i1} is neither unbiased nor consistent:

From (B.5), the OLS-estimator of γ_{i1} is given by:

$$(B.51)\quad \hat{\gamma}_{i1}=\frac{\sum_{t=1}^{T} y_i(t)y_i(t-1)}{\sum_{t=1}^{T} y_i^2(t-1)}=\frac{\sum_{t=2}^{T} y_i(t)y_i(t-1)}{\sum_{t=2}^{T} y_i^2(t-1)} \qquad \text{(see also (B.42))}.$$

Elimination of the serially correlated disturbances in (B.50):

$$(B.52)\quad y_i(t)=\gamma_{i1}y_i(t-1)+\rho_i y_i(t-1)-\gamma_{i1}\rho_i y_i(t-2)+\eta_i(t)\ ,$$

which multiplied with $y_i(t-1)$ and summed w.r.t.tgives:

$$(B.53)\quad \sum_{t=2}^{T} y_i(t)y_i(t-1)=(\gamma_{i1}+\rho_i)\sum_{t=2}^{T} y_i^2(t-1)-\gamma_{i1}\rho_i\sum_{t=2}^{T} y_i(t-1)y_i(t-2)$$

$$+\sum_{t=2}^{T}\eta_i(t)y_i(t-1)$$

or from (B.51)

$$(B.54)\quad \hat{\gamma}_{i1}=\gamma_{i1}+\rho_i-\gamma_{i1}\rho_i\frac{\sum_{t=2}^{T} y_i(t-1)y_i(t-2)}{\sum_{t=2}^{T} y_i^2(t-1)}+\frac{\sum_{t=2}^{T}\eta_i(t)y_i(t-1)}{\sum_{t=2}^{T} y_i^2(t-1)}\ ,$$

from which it is seen that the bias of the O.L.S.estimator $\hat{\gamma}_{i1}$ does not tend to zero if the number of observations grows indefinitely.

Indeed, the first ratio in (B.54) has the same probability limit as $\hat{\gamma}_{i1}$ and the second ratio tends in probability to zero, so that:

(B.55) $\underset{T\to\infty}{\text{plim}}\ \hat{\gamma}_{i1} = \dfrac{\gamma_{i1}+\rho_i}{1+\gamma_{i1}\rho_i}$, which does not tend to γ_{i1}, even

if ρ_i is small (unless $\rho_i=0$). Therefore, as long as the disturbances of an autoregressive model are autocorrelated, the OLS-estimator of γ_{i1} is inconsistent Δ

2. **The OLS-estimator of** ρ_i is biased and inconsistent.

Since the disturbance terms $\varepsilon_{i(t)}$ are in fact unobservable, the autocorrelation parameter is estimated from

$\hat{\varepsilon}_i(t)=r_i\hat{\varepsilon}_i(t-1)+\eta_i(t)$ with OLS-estimator:

(B.56) $\hat{\rho}_i=\hat{r}_i= \dfrac{\sum_{t=1}^{T}\hat{\varepsilon}_i(t)\hat{\varepsilon}_i(t-1)}{\sum_{t=1}^{T}\hat{\varepsilon}_i^2(t-1)}$

and assuming that the autoregressive process of the residuals satisfies an infinite stationary process ($|r_i|<1$):

(B.57) $\underset{T\to\infty}{\text{plim}}\ \frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_i^2(t-1)=\underset{T\to\infty}{\text{plim}}\ \frac{1}{T-1}\sum_{t=2}^{T}\hat{\varepsilon}_i^2(t-1)$

$$=\underset{T\to\infty}{\text{plim}}\ \frac{1}{T-1}\sum_{t=2}^{T}y_i^2(t-1)-2\underset{T\to\infty}{\text{plim}}(\hat{\gamma}_{i1})\underset{T\to\infty}{\text{plim}}\frac{1}{T-1}\sum_{t=2}^{T}y_i(t-1)y_i(t-2)+$$

$$\underset{T\to\infty}{\text{plim}}\frac{1}{T-1}\Sigma y_i^2(t-2)\ \underset{T\to\infty}{\text{plim}}(\hat{\gamma}_{i1}^2)$$

$$=\left[1-\underset{T\to\infty}{\text{plim}}(\hat{\gamma}_{i1})\right]^2\sigma_y^2 \quad \text{and}$$

(B.58) $\sum_{t=1}^{T}\hat{\varepsilon}_i(t)\hat{\varepsilon}_i(t-1)=\sum_{t=1}^{T}y_i(t)y_i(t-1)-\hat{\gamma}_{i1}\sum_{t=1}^{T}y_i^2(t-1)-\hat{\gamma}_{i1}\sum_{t=2}^{T}y_i(t)y_i(t-2)$

$$+\ \hat{\gamma}_{i1}^2\sum_{t=2}^{T}y_i(t-1)y_i(t-2)$$

$$=-\hat{\gamma}_{i1}\sum_{t=2}^{T}y_i(t)y_i(t-2)+\hat{\gamma}_{i1}^2\sum_{t=2}^{T}y_i(t-1)y_i(t-2)\ .$$

Proceeding in the same way as for determining the plim of $\hat{\gamma}_{i1}$, equation (B.52) may be multiplied with $y_i(t-2)$ and summed w.r.t.t (and divided by T-1):

$$(B.59)\ \frac{1}{T-1}\sum_{t=2}^{T} y_i(t)y_i(t-2)=(\gamma_{i1}+\rho_i)\frac{1}{T-1}\sum_{t=2}^{T} y_i(t-1)y_i(t-2)\ -$$

$$\gamma_{i1}\rho_i\frac{1}{T-1}\sum_{t=2}^{T} y_i^2(t-2)+\frac{1}{T-1}\sum_{t=2}^{T}\eta_i(t)y_i(t-2)$$

so that subtracting $\hat{\gamma}_{i1}\frac{1}{T-1}\sum_{t=2}^{T} y_i(t-1)y_i(t-2)$ from both sides of (B.59) and transforming to plim's:

$$(B.60)\ \operatorname*{plim}_{T\to\infty}\frac{1}{T-1}\sum_{t=2}^{T} y_i(t)y_i(t-2)-\operatorname*{plim}_{T\to\infty}\frac{1}{T-1}\hat{\gamma}_{i1}\sum_{t=2}^{T} y_i(t-1)y_i(t-2)$$

$$=\left[(\gamma_{i1}+\rho_i-\operatorname*{plim}_{T\to\infty}\hat{\gamma}_{i1})\operatorname*{plim}_{T\to\infty}\hat{\gamma}_{i1}-\gamma_{i1}\rho_i\right]\sigma_y^2$$

or, from (B.58) and (B.60):

$$(B.61)\ \operatorname*{plim}_{T\to\infty}\frac{1}{T-1}\sum_{t=1}^{T}\hat{\varepsilon}_i(t)\hat{\varepsilon}_i(t-1)=-\operatorname*{plim}_{T\to\infty}\hat{\gamma}_{i1}\left[(\gamma_{i1}+\rho_i-\operatorname*{plim}_{T\to\infty}\hat{\gamma}_{i1})\operatorname*{plim}_{T\to\infty}\hat{\gamma}_{i1}-\gamma_{i1}\rho_i\right]\sigma_y^2,$$

so that substituting $\operatorname*{plim}_{T\to\infty}\hat{\gamma}_{i1}$ by (B.55) in (B.61) and (B.57):

$$(B.62)\ \operatorname*{plim}_{T\to\infty}\hat{\rho}_i=\operatorname*{plim}_{T\to\infty}\hat{r}_i=\frac{\rho_i\gamma_{i1}(\gamma_{i1}+\rho_i)}{(1+\gamma_{i1}\rho_i)}$$

, or (B.55) and (B.62) imply:

$$(B.63)\ \operatorname*{plim}_{T\to\infty}(\hat{\gamma}_{i1}+\hat{\rho}_i)=\gamma_{i1}+\rho_i\ .$$

Hence the estimation of ρ_i by $\hat{\rho}_i$ entails an inconsistency which is exactly apposite to that generated in the estimation of γ_{i1} by $\hat{\gamma}_{i1}$. Δ

The expressions (B.55), (B.62) and (B.63) imply that for large samples:

- $\hat{\gamma}_{i1}$ underestimates γ_{i1} for $\rho_i<0$ and overestimates γ_{i1} for $\rho_i>0$;
- $\hat{r}_i=\hat{\rho}_i$ underestimates r_i and so ρ_i (since the autocorre-

lation of the residuals $\hat{\varepsilon}_i(t)$ is more moderate than the autocorrelation of the error terms $\varepsilon_i(t)$) if $\rho_i>0$ and overestimates the negative autocorrelation.

The inconsistency is reduced in magnitude if one or more exogenous variables appear among the explanatory variables of the i'th equation. However, it only disappears if the disturbances $\varepsilon_i(t)$ are not generated by a stochastic autoregressive process. This will be the subject of the next paragraph.

B2.2 First order autoregression and autocorrelation with exogenous variables.

Theorem B.3

The OLS-estimator of the autoregressive parameter γ_{i1} in the equation:

(B.64) $y_i(t)=\alpha_{i1}z_{i1}(t)+\gamma_{i1}y_i(t-1)+\varepsilon_i(t)$ with $\varepsilon_i(t)=\rho_i\varepsilon_i(t-1)+\eta_i(t)$

is biased and inconsistent with inconsistency amounting to:

(B.65) $$\mathrm{Inc}(\hat{\gamma}_{i1})=\operatorname*{plim}_{T\to\infty}\hat{\gamma}_{i1}-\gamma_{i1}=\frac{\rho_i(1-\gamma_{i1}^2)}{(1+\gamma_{i1}\rho_i)}\cdot\frac{1}{1+\dfrac{\alpha_{i1}^2\sigma^2_{v_{i(-1)}\cdot z_{i1}}}{\sigma^2_{w_i}}},\text{ with}$$

$\sigma^2_{v_{i(-1)}\cdot z_{i1}}$ = that part of the variance of $v_i(t-1)=\sum_{\tau=0}^{\infty}\gamma_{i1}^{\tau}z_{i1}(t-1-\tau)$ $(\forall t)$ which is not associated with the variance of $z_{i1}(t)$ (i.e. that part of the variance being uncorrelated with z_{i1})

$\sigma^2_{w_i}$ = variance of $w_i(t)=\sum_{\tau=0}^{\infty}\gamma_{i1}^{\tau}\varepsilon_i(t-\tau)$ $(\forall t)$ (*)

Proof

Since (B.64) or

(B.67) $y_i(t)=\alpha_{i1}z_{i1}(t)+\gamma_{i1}y_i(t-1)+\rho_i\varepsilon_i(t-1)+\eta_i(t)$

(*) Or, from (B.64):

(B.66) $y_i(t)=\alpha_{i1}\sum_{\tau=0}^{\infty}\gamma_{i1}^{\tau}z_{i1}(t-\tau)+\sum_{\tau=0}^{\infty}\gamma_{i1}^{\tau}\varepsilon_i(t-\tau)=\alpha_{i1}v_i(t)+w_i(t)$

is the "true" equation and, erroneously,

(B.68) $y_i(t)=a_{i1}z_{i1}(t)+c_{i1}y_i(t-1)+u_i(t)$ is estimated by OLS. assuming $u_i(t)\approx(0,s_{ii})$, we may consider (B.68) as a "misspecified" equation of the "true" relationship (B.76) (*)

(*) In general, the formula for (finite or asymptotic) specification bias and specification inconsistency may be derived as follows:

Consider the true model:

(B.69) $y_i=X_i\beta_i+\varepsilon_i\ (i=1,2,\dots,M)$ with initial assumptions $E(\varepsilon_i)=0$, $E(\varepsilon_i\varepsilon_i')=\sigma_{ii}I_T$ and the columns of X_i statistically independent of ε_i (always if X_i non stochastic)

and the misspecified model:

(B.70) $y_i=\overline{X}_ib_i+e_i$ with $E(e_i)=0, E(e_ie_i')=s_{ii}I_T$ and the columns of $\overline{X}_i$ independent of ε_i.

Then:

(B.71) $E(\hat{b}_i)=E\left[(\overline{X}_i'\overline{X}_i)^{-1}\overline{X}_i'y_i\right]=E\left[(\overline{X}_i'\overline{X}_i)^{-1}\overline{X}_i'X_i\right]\beta_i=P_i\beta_i$,

with $(\overline{X}_i'\overline{X}_i)^{-1}\overline{X}_i'X_i$ a matrix of regression coefficients in the set of "auxiliary" OLS-regressions of each x in X_i on all the $\overline{x}$'s in $\overline{X}_i$ or

(B.72) bias $(\hat{b}_i)=E(\hat{b}_i)-\beta_i=(P_i-I_{k_i})\beta_i$ and

(B.73) $\text{inc}(\hat{b}_i)=\underset{T\to\infty}{\text{plim}}\ \hat{b}_i-\beta_i=\left[(\underset{T\to\infty}{\text{plim}}\frac{1}{T}(\overline{X}'\overline{X}))^{-1}\underset{T\to\infty}{\text{plim}}\frac{1}{T}\overline{X}'X-I_{k_i}\right]\beta_i$.

For the above problem (B.67-B.68), one relevant variable, say $x_{ik_i}(t)=\varepsilon_i(t-1)$ has been left out, so that there is only one non-trivial "auxiliary" regression, that of x_{i,k_i} on all the included variables:

(B.74) $x_{i,k_i}=p_{i1}x_{i1}+p_{i2}x_{i2}+\dots+p_{i,k_i-1}x_{i,k_i-1}+\nu_i$

(the other equations are identities), so that the $(\overline{X}_i'\overline{X}_i)^{-1}\overline{X}_i'X_i$-matrix can be partitioned into an identity matrix and a column vector of the p_i-values or $P_i=(I_{k_i-1},p_i)$ or

(B.75) $E(\hat{b}_i)=\beta_i^-+p_i\beta_{i,k_i}$ with $(\beta_i^-,\beta_{i,k_i})'=\beta_i',\beta_i^-$ having (k_i-1) elements or

(B.76) $\text{bias}(\hat{b}_i)=\beta_{i,k_i}p_i$, and due to the lagged dependent variable in X_i,

(B.77) $\text{inc}(\hat{b}_i)=\beta_{i,k_i}p_i$.

From (B.73)and(B.77), the plim of the OLS-estimator $\hat{c}_{i1}$ is:

(B.78) $\operatorname*{plim}_{T\to\infty} \hat{c}_{i1} = \gamma_{i1} + \rho_i \operatorname*{plim}_{T\to\infty} \hat{p}_{i2}$, where $\hat{p}_{i2}$ is the regression coefficient of $y_i(t-1)$ in the auxiliary regression of $\varepsilon_i(t-1)$ on $z_{i1}(t)$ and $y_i(t-1)$.

But since

(B.79) $\operatorname*{plim}_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \varepsilon_i(t-1) z_{i1}(t) = 0$ and $\hat{p}_{i2} = b_{\varepsilon_i(-1) y_i(-1) . z_{i1}} =$

$$\frac{b_{\varepsilon_i(-1) y_i(-1)} - b_{\varepsilon_i(-1) z_{i1}} b_{z_{i1} y_i(-1)}}{1 - r^2_{y_i(-1) z_{i1}}}$$

(B.80) $\operatorname*{plim}_{T\to\infty} \hat{p}_{i2} = \operatorname*{plim}_{T\to\infty} \left[\frac{b_{\varepsilon_i(-1) y_i(-1)}}{1 - r^2_{y_i(-1) z_{i1}}} \right] = \frac{1}{1 - r^2_{y_i(-1) z_{i1}}}$

$$\operatorname*{plim}_{T\to\infty} \frac{\frac{1}{T} \sum_{t=1}^{T} \varepsilon_i(t-1) y_i(t-1)}{\frac{1}{T} \sum_{t=1}^{T} y_i^2(t-1)}$$

and because

(B.81) $\operatorname*{plim}_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} z_{i1}(t) \varepsilon_i(t) = 0$ (see also (B.80)) and

$\operatorname*{plim}_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \varepsilon_i(t) \varepsilon_i(t-1) = \rho_i \sigma^2_{\varepsilon_i}$ and by (B.66):

(B.82) $\operatorname*{plim}_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \varepsilon_i(t) y_i(t) = \operatorname*{plim}_{T\to\infty} \sum_{t=1}^{T} \varepsilon_i(t-1) y_i(t-1) =$

$$\sum_{\tau=0}^{\infty} \gamma_{i1}^{\tau} \operatorname*{plim}_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \varepsilon_i(t) \varepsilon_i(t-\tau$$

$$= \sum_{\tau=0}^{\infty} \gamma_{i1}^{\tau} \rho_i^{\tau} \sigma^2_{\varepsilon_i} = \frac{\sigma^2_{\varepsilon_i}}{(1 - \gamma_{i1} \rho_i)} \quad \text{(stationarity)},$$

and since the probability limit of a ratio can be written as a ratio of probability limits (*), (B.80) becomes:

(*) See Slutsky's theorem, e.g. in H.Cramèr, [6],p.255.

$$(B.83)\ \underset{T\to\infty}{\text{plim}}\ \hat{p}_{i2} = \frac{1}{(1-r^2_{y_i(-1)z_{i1}})} \frac{\underset{T\to\infty}{\text{plim}}\frac{1}{T}\sum_{t=1}^{T}\varepsilon_i(t-1)y_i(t-1)}{\underset{T\to\infty}{\text{plim}}\frac{1}{T}\sum_{t=1}^{T}y_i^2(t-1)} =$$

$$\frac{\sigma^2_{\varepsilon_i}}{(1-\gamma_{i1}\rho_i)} \frac{1}{\sigma^2_{y_i(-1)}(1-r^2_{y_i(-1)z_{i1}})}$$

or substituted into (B.78):

$$(B.84)\ \underset{T\to\infty}{\text{plim}}\ \hat{c}_{i1} \equiv \underset{T\to\infty}{\text{plim}}\hat{\gamma}_{i1} = \gamma_{i1} + \frac{\rho_i}{(1-\gamma_{i1}\rho_i)} \frac{\sigma^2_{\varepsilon_i}}{\sigma^2_{y_i(-1)}(1-r^2_{y_i(-1)z_{i1}})},$$

or the magnitude of the inconsistency depends upon ρ_i and the relative importance of ε_i:

$$(B.85)\ \text{inc}(\hat{\gamma}_{i1}) = \underset{T\to\infty}{\text{plim}}\hat{\gamma}_{i1} - \gamma_{i1} = \frac{\rho_i}{(1-\gamma_{i1}\rho_i)} \frac{\sigma^2_{\varepsilon_i}}{\sigma^2_{y_i(-1)}(1-r^2_{y_i(-1)z_{i1}})} =$$

$$\frac{\rho_i}{(1-\gamma_{i1}\rho_i)} \frac{\sigma^2_{\varepsilon_i}}{\sigma^2_{y_i(-1)z_{i1}}},$$

$\sigma^2_{y_i(-1)z_{i1}}$ being the part of the variance of $y_i(-1)$ which is uncorrelated with the variation of z_{i1}, so that from equation (B.66) and the expression for the asymptotic variance of $w_i(t) = \sum_{\tau=0}^{\infty} \gamma_{i1}^{\tau}\varepsilon_i(t-\tau)$, the inconsistency (B.65) is obtained. Therefrom, it is clear that $\hat{\gamma}_{i1}$ will asymptotically overestimate γ_{i1} as long as $\rho_i > 0$ (see also opposite inconsistency in previous model:(B.63)). Δ

<u>Corollary B 1</u> The introduction of an exogenous variable z_{i1} reduces the absolute value of the inconsistency of the OLS estimator $\hat{\gamma}_{i1}$.

<u>Proof.</u> From the expression (B.55), the inconsistency of $\hat{\gamma}_{i1}$ for equation (B.50) is:

(B.86) $Inc(\hat{\gamma}_{i1})=\frac{\rho_i(1-\gamma_{i1}^2)}{(1+\gamma_{i1}\rho_i)}$, which is larger in absolute value than (B.65) (★). Δ

Corollary B 2 If the observations of the exogenous variable z_{i1} also follow a 1 st order Markov scheme with parameter R_i, the expression (B.65) for $inc(\hat{\gamma}_{i1})$ becomes:

$$(B.87)\ Inc(\hat{\gamma}_{i1})=\frac{\rho_i(1-\gamma_{i1}^2)}{(1+\gamma_{i1}\rho_i)}\cdot\frac{1}{\left(1+\frac{\alpha_{i1}^2\sigma_{z_{i1}}^2}{\sigma_{\varepsilon_i}^2}\right)\frac{(1-R_i^2)(1-\gamma_{i1}\rho_i)}{(1-\gamma_{i1}R_i)^2(1+\gamma_{i1}\rho_i)}}\qquad(★★)$$

B2.3 An additional specification error: estimation of a 1 st order i.s. of a 2 nd order model

Theorem B.4

If the true equation is the 2 nd order autoregressive lag scheme:

$$(B.88)\ y_i(t)=\alpha_{i1}z_{i1}(t)+\gamma_{i1}y_i(t-1)+\gamma_{i2}y_i(t-2)+\varepsilon_i(t)$$

but a first order model is estimated instead (by OLS):

$$(B.89)\ y_i(t)=a_{i1}z_{i1}(t)+c_{i1}y_i(t-1)+u_i(t)$$

where the $z_{i1}(t)$ form a stationary and serially uncorrelated process and the disturbances may be uncorrelated or correlated, stationary or instationary, then the probability limit of the O.L.S. estimator of c_{i1} is equal to:

$$(B.90)\ \underset{T\to\infty}{\text{plim}}\ \hat{c}_{i1}=\underset{T\to\infty}{\text{plim}}\ \hat{\gamma}_{i1}=\frac{\gamma_{i1}}{1-\gamma_{i2}}$$

Proof. From the stationarity and the serial uncorrelation of $z_{i1}(t)$: $\sigma_{y(-1)}^2=\sigma_{y_i}^2$ and

$$(B.91)\ \underset{T\to\infty}{\text{plim}}\ b_{y_i(-2)y_i(-1).z_{i1}}=\underset{T\to\infty}{\text{plim}}\ b_{y_i(-2)y_i(-1)}=\underset{T\to\infty}{\text{plim}}\ b_{y_i(-1)y_i(-2)}$$

(★) Only if $\alpha_{i1}=0$, i.e. if there is no exogenous variable in (B.64), both inconsistencies are equal.

(★★) The proof is left for the reader.

$= \underset{T\to\infty}{\text{plim}}\ b = \underset{T\to\infty}{\text{plim}}\ \hat{c}_{i1} = \underset{T\to\infty}{\text{plim}}\ \hat{\gamma}_{i1}$ or from (B.73)and(B.77):

$$\text{(B.92)}\ \underset{T\to\infty}{\text{plim}}\ \hat{\gamma}_{i1} = \gamma_{i1} + \gamma_{i2} \underset{T\to\infty}{\text{plim}}\ \hat{\gamma}_{i1} = \frac{\gamma_{i1}}{1-\gamma_{i2}}$$ (see also (B.78)) Δ

B.3 The consistency and asymptotic (in) efficiency of the feasible Aitken estimator in autoregressive autocorrelated models.

Theorem B 5.

If Aitken's G.L.S. is utilized (equation by equation) using a consistent estimate of the error variance-covariance matrix (say by instrumental variables), then the resulting estimates are consistent but not asymptotic efficient if (a) lagged dependent variable(s) occur(s) among the explanatory variables.

Proof. (*)

Consider the i'th equation:

$$\text{(B.93)}\ y_i(t) = \alpha_{i1} z_{i1}(t) + \gamma_{i1} y_i(t-1) + \varepsilon_i(t-1) \text{ with } \varepsilon_i(t) = \rho_i \varepsilon_i(t-1) + \eta_i(t)$$

with

$\underset{T\to\infty}{\text{plim}} \frac{1}{T} X_i' X_i$ a finite positive definite matrix with $X_i = (Z_i, y_{\theta_i}) = (z_{i1} y_i(-1))$ the matrix of observations of the explanatory variables in the i'th equation(i=1,2,...,M).

(i) If ρ_i and hence Ω_{ii} are known, the Aitken estimator of α_{i1} and γ_{i1}, being

$$\text{(B.94)} \begin{pmatrix} \tilde{\alpha}_{i1} \\ \tilde{\gamma}_{i1} \end{pmatrix} = \left[\begin{pmatrix} z_{i1}' \\ y_i'(-1) \end{pmatrix} \Omega_{ii}^{-1} (z_{i1}, y_i(-1)) \right]^{-1} \begin{pmatrix} z_{i1}' \\ y_i'(-1) \end{pmatrix} \Omega_{ii}^{-1} y_i$$

or

$$\text{(B.95)}\ \tilde{\beta}_i = (X_i' \Omega_{ii}^{-1} X_i)^{-1} X_i' \Omega_{ii}^{-1} y_i \qquad \text{with}$$

(*) See also T. **Amemiya** and W.Fuller [1], Section 5, pp 520-523 and K.Wallis [24], Appendix, pp. 566-567.

$$(B.96)\ \Omega_{ii}^{-1}=\frac{1}{\sigma_{ii}}\begin{bmatrix} 1 & -\rho_i & 0 & \cdots & 0 & 0 \\ -\rho_i & 1+\rho_i^2 & -\rho_i & \cdots & 0 & 0 \\ 0 & -\rho_i & 1+\rho_i^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1+\rho_i^2 & -\rho_i \\ 0 & 0 & 0 & \cdots & -\rho_i & 1 \end{bmatrix}$$

is BLUE, consistent and asymptotically efficient if the error term ε_i is assumed to be multivariate normally distributed. Then the asymptotic variance-covariance matrix of $\sqrt{T}(\tilde{\alpha}_{i1}-\alpha_{i1})$ and $\sqrt{T}(\tilde{\gamma}_{i1}-\gamma_{i1})$ is given by:

$$(B.97)\ V_i=\operatorname*{plim}_{T\to\infty} T(X_i'\Omega_{ii}^{-1}X_i)^{-1}=\operatorname*{plim}_{T\to\infty} T\left[\begin{pmatrix} z_{i1}' \\ y_i(-1)' \end{pmatrix}\Omega_{ii}^{-1}(z_{i1}\,y_i(-1))\right]^{-1}$$

(ii) If ρ_i (and/or σ_{ii}) is not known, then a consistent estimate $\hat{\Omega}_{ii}$ of Ω_{ii} may be used to obtain:

$$(B.98)\ \hat{\hat{\beta}}_i=\begin{pmatrix} \hat{\hat{\alpha}}_{i1} \\ \hat{\hat{\gamma}}_{i1} \end{pmatrix}=(X_i'\hat{\Omega}_{ii}^{-1}X_i)^{-1}X_i'\hat{\Omega}_{ii}^{-1}y_i$$

$$=\left[\begin{pmatrix} z_{i1}' \\ y_i'(-1) \end{pmatrix}\hat{\Omega}_{ii}^{-1}\left(z_{i1}\,y_i(-1)\right)\right]^{-1}\begin{pmatrix} z_{i1}' \\ y_i(-1) \end{pmatrix}\hat{\Omega}_{ii}^{-1}y_i\ ,$$

with sampling error:

$(B.99)\ \hat{\hat{\beta}}_i-\beta_i=(X_i'\hat{\Omega}_{ii}^{-1}X_i)^{-1}X_i'\hat{\Omega}_{ii}^{-1}\varepsilon_i$, and by Slutsky's theorem (*)

$$(B.100)\ \operatorname*{plim}_{T\to\infty}\sqrt{T}(\hat{\hat{\beta}}_i-\beta_i)=\operatorname*{plim}_{T\to\infty}\sqrt{T}\begin{pmatrix} \hat{\hat{\alpha}}_{i1}-\alpha_{i1} \\ \hat{\hat{\gamma}}_{i1}-\gamma_{i1} \end{pmatrix}$$

$$=\operatorname*{plim}_{T\to\infty}\left(\frac{X_i'\hat{\Omega}_{ii}^{-1}X_i}{T}\right)^{-1}\operatorname*{plim}_{T\to\infty}\frac{X_i'\hat{\Omega}_{ii}^{-1}\varepsilon_i}{\sqrt{T}}$$

Now, we shall evaluate the two probability limits of the r.h.s. of (B.100) upto order $O(T^{-\frac{1}{2}})$.

(1.) First, consider the second plim.

Expanding $\hat{\Omega}_{ii}^{-1}$ in a Taylor series about ρ_i yields:

(*)See footnote on p.82

$$(B.101)\ \hat{\Omega}_{ii}^{-1}=\Omega_{ii}^{-1}+\frac{\partial\Omega_{ii}^{-1}}{\partial\rho_i}(\hat{\rho}_i-\rho_i)+0(\hat{\rho}_i-\rho_i)^2$$

$$=\Omega_{ii}^{-1}+\frac{\partial\Omega_{ii}^{-1}}{\partial\rho_i}(\hat{\rho}_i-\rho_i)+0(\tfrac{1}{T}) \qquad (*) \quad \text{or}$$

$$(B.102)\ \frac{X_i'\hat{\Omega}_{ii}^{-1}\varepsilon_i}{\sqrt{T}}=\frac{X_i'\Omega_{ii}^{-1}\varepsilon_i}{\sqrt{T}}+\sqrt{T}(\hat{\rho}_i-\rho_i)\ \frac{X_i'\left(\frac{\partial\Omega_{ii}^{-1}}{\partial\rho_i}\right)\varepsilon_i}{T}+0(\tfrac{1}{\sqrt{T}}).$$

Since the (2x1)-matrix $X_i'\left(\frac{\partial\Omega_{ii}^{-1}}{\partial\rho_i}\right)X_i$ can be written as:

(B.103)

$$X_i'\left(\frac{\partial\Omega_{ii}^{-1}}{\partial\rho_i}\right)\varepsilon_i=\frac{1}{\sigma_{ii}}\begin{pmatrix} z_{i1}(1),z_{i1}(2),\ldots\ldots,z_{i1}(T) \\ y_i(o),y_i(1),\ \ldots\ldots,y_i(T-1)\end{pmatrix}$$

$$\cdot\begin{bmatrix} 0 & -1 & 0 & 0\ldots\ldots & 0 & 0 \\ -1 & 2\rho_i & -1 & 0\ldots\ldots & 0 & 0 \\ 0 & -1 & 2\rho_i & -1\ldots\ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0\ldots\ldots -2\rho_i & & -1 \\ 0 & 0 & 0 & 0\ldots\ldots -1 & & 0\end{bmatrix}\begin{bmatrix}\varepsilon_i(1) \\ \varepsilon_i(2) \\ \varepsilon_i(3) \\ \vdots \\ \varepsilon_i(T)\end{bmatrix}$$

$$=\frac{1}{\sigma_{ii}}\begin{pmatrix} z_{i1}(1),z_{i1}(2),\ldots\ldots,z_{i1}(T) \\ y_i(o),y_i(1),\ \ldots\ldots,y_i(T-1)\end{pmatrix}\begin{pmatrix} -\varepsilon_i(2) \\ -\varepsilon_i(1)+2\rho_i\varepsilon_i(2)-\varepsilon_i(3) \\ -\varepsilon_i(2)+2\rho_i\varepsilon_i(3)-\varepsilon_i(4) \\ \vdots \\ -\varepsilon_i(T-2)+2\rho_i\varepsilon_i(T-1)-\varepsilon_i(T) \\ \varepsilon_i(T-1)\end{pmatrix},$$

from which it is seen that the 2nd element involves terms as

(*) Because $\hat{\rho}_i$ is assumed to be a consistent estimator of ρ_i such that $\hat{\rho}_i-\rho_i=0(T^{-\frac{1}{2}})$

$y_i(t)\varepsilon_i(t), y_i(t)\varepsilon_i(t+1)$ and $y_i(t)\varepsilon_i(t+2)$ which are assumed to converge in probability to resp.

$$\frac{\sigma^2_{\varepsilon_i}}{(1-\gamma_{i1}\rho_i)}, \quad \frac{\rho_i\sigma^2_{\varepsilon_i}}{(1-\gamma_{i1}\rho_i)} \text{ and } \frac{\rho_i^2\sigma^2_{\varepsilon_i}}{(1-\gamma_{i1}\rho_i)}$$ (see (B.81) and (B.82): stationarity) or

(B.104)

$$\underset{T\to\infty}{\text{plim}}\frac{1}{T}X_i'\left(\frac{\partial\Omega_{ii}^{-1}}{\partial\rho_i}\right)\varepsilon_i$$

$$=\frac{1}{\sigma_{ii}}\underset{T\to\infty}{\text{plim}}\frac{1}{T}\begin{bmatrix}-z_{i1}(1)\varepsilon_i(2)+z_{i1}(2)(-\varepsilon_{i1}(1)+2\rho_i\varepsilon_i(2)-\varepsilon_i(3))+\ldots+z_{i1}(T)\varepsilon_i(T-1)\\ -y_i(o)\varepsilon_i(2)+y_i(1)(-\varepsilon_i(1)+2\rho_i\varepsilon_i(2)-\varepsilon_i(3))+\ldots+y_i(T-1)\varepsilon_i(T-1)\end{bmatrix}$$

$$=\frac{1}{\sigma_{ii}}\lim_{T\to\infty}\frac{1}{T}\begin{bmatrix} o \\ -\frac{\rho_i^2\sigma^2_{\varepsilon_i}}{(1-\gamma_{i1}\rho_i)}+(T-2)\frac{\left(\rho_i^2\sigma^2_{\varepsilon_i}-\sigma^2_{\varepsilon_i}\right)}{(1-\gamma_{i1}\rho_i)}-\frac{\sigma^2_{\varepsilon_i}}{(1-\gamma_{i1}\rho_i)}\end{bmatrix}$$

$$=\frac{1}{\sigma_{ii}}\lim_{T\to\infty}\frac{1}{T}\begin{bmatrix} o \\ \left[T(\rho_i^2-1)-3\rho_i^2+1\right]\sigma_{ii}(1-\rho_i^2)^{-1}(1-\gamma_{i1}\rho_i)^{-1}\end{bmatrix}$$

$$=\begin{bmatrix} o \\ \frac{-1}{(1-\gamma_{i1}\rho_i)}\end{bmatrix}$$, since

$$\underset{T\to\infty}{\text{plim}}\frac{1}{T}\sum_{\tau=1}^{T} z_{i1}(t)\varepsilon_i(t-\tau)=0 \qquad (\tau=0,1,2,\ldots) \qquad (\star)$$

(★) So, the expression (B.104) completely vanishes if the set of explanatory variables in the i'th equation consists only of variables which are asymptotically uncorrelated with the disturbance vector ε_i.

Or, from equations (B.102) and (B.104), we find that

$$\frac{X_i'\hat{\Omega}_{ii}^{-1}\varepsilon_i}{\sqrt{T}} \text{ and } \frac{X_i'\Omega_{ii}^{-1}\varepsilon_i}{\sqrt{T}}+\sqrt{T}(\hat{\rho}_i-\rho_i)\begin{pmatrix} o \\ \frac{-1}{(1-\gamma_{i1}\rho_i)}\end{pmatrix} \text{ have the same}$$

limiting probability distributions.

Applying the same Taylor-series expansion as in (B.101)-(B.102) for the covariance matrix, we find:

$$\text{(B.105)}\ \underset{T\to\infty}{\text{plim}}\ T(X_i'\hat{\Omega}_{ii}^{-1}X_i)^{-1}=\underset{T\to\infty}{\text{plim}}\ T(X_i'\Omega_{ii}^{-1}X_i)^{-1} = V_i\ .$$

Substituting into equation (B.100), we find that $\sqrt{T}(\hat{\hat{\beta}}_i-\beta_i)$ is asymptotically distributed as

$$\frac{1}{\sqrt{T}}V_iX_i'\Omega_{ii}^{-1}\varepsilon_i+V_i\begin{pmatrix} o \\ \frac{-1}{(1-\gamma_{i1}\rho_i)}\end{pmatrix}\sqrt{T}(\hat{\rho}_i-\rho_i) \qquad ,$$

the second term of which implies that $\hat{\hat{\beta}}_i$ is not asymptotically efficient. Thus, the magnitude of this asymptotic inefficiency depends upon the asymptotic distribution of $\sqrt{T}(\hat{\rho}_i-\rho_i)$ and it is a consequence of the joint occurrence of (a) lagged dependent variable(s) and the (consistent) estimation of the covariance matrix (otherwise, no Taylor series expansions;see also previous footnote). Δ

Example: First round consistent estimation by Instrumental Variables.

If $\hat{\beta}_i$ is the consistent instrumental variables estimator:

$$(B.106)\ \hat{\beta}_i = (X_i^{*'} X_i)^{-1} X_i^{*'} y_i \quad \text{and}$$

$$(B.107)\ \hat{\varepsilon}_i = y_i - X_i \hat{\beta}_i = \left[I_T - X_i (X_i^{*'} X_i)^{-1} X_i^{*'}\right] y_i = \left[I_T - X_i (X_i^{*'} X_i)^{-1} X_i^{*'}\right] \varepsilon_i$$

$$= \varepsilon_i - X_i (X_i^{*'} X_i)^{-1} X_i^{*'} \varepsilon_i$$

$$(B.108)\ \hat{\varepsilon}_i(-1) = y_i(-1) - X_i(-1)\hat{\beta}_i = \varepsilon_i(-1) - X_i(-1)(X_i^{*'} X_i)^{-1} X_i^{*'} \varepsilon_t \quad \text{with}$$

$\underset{T\to\infty}{\text{plim}} \frac{1}{T} X_i^{*'} X_i = H_i$ a finite and nonsingular matrix and

$\underset{T\to\infty}{\text{plim}} \frac{1}{T} X_i^{*'} \varepsilon_i = 0$, where X_i^{*} is a $(T\times k_i)$ matrix of observations on instrumental variables for X_i.

Then Ω_{ii}^{-1} is consistently estimated with the help of:

$$(B.109)\ \hat{\rho}_i = \frac{\sum_{t=1}^{T} \hat{\varepsilon}_i(t)\hat{\varepsilon}_i(t-1)}{\sum_{t=2}^{T} \hat{\varepsilon}_i^2(t-1)} \text{and } \hat{\sigma}_{ii} = \frac{\hat{\varepsilon}_i'\hat{\varepsilon}_i}{T}(1-\hat{\rho}_i^2)\ , \text{ with}$$

$$(B.110)\ \underset{T\to\infty}{\text{plim}} \frac{1}{T} \sum_{t=1}^{T} \hat{\varepsilon}_i^2(t-1) = \underset{T\to\infty}{\text{plim}} \frac{1}{T-1} \sum_{t=2}^{T} \hat{\varepsilon}_i^2(t) = \sigma_{\varepsilon_i}^2 \quad \text{and}$$

$$(B.111)\ \underset{T\to\infty}{\text{plim}} \frac{1}{T} \sum_{t=1}^{T} \hat{\varepsilon}_i(t)\hat{\varepsilon}_i(t-1) = \underset{T\to\infty}{\text{plim}} \frac{1}{T}\hat{\varepsilon}_i'\hat{\varepsilon}_i(-1) = \underset{T\to\infty}{\text{plim}} \frac{1}{T}\varepsilon_i'\varepsilon_i(-1)$$

$$-\underset{T\to\infty}{\text{plim}} \frac{1}{T}\varepsilon_i' X_i(-1)(X_i^{*'} X_i)^{-1} X_i^{*'} \varepsilon_i - \underset{T\to\infty}{\text{plim}} \frac{1}{T}\varepsilon_i' X_i^{*}(X_i' X_i^{*})^{-1} X_i' \varepsilon_i(-1)$$

$$+\underset{T\to\infty}{\text{plim}} \frac{1}{T}\varepsilon_i' X_i^{*}(X_i' X_i^{*})^{-1} X_i' X_i(-1)(X_i^{*'} X_i)^{-1} X_i^{*'} \varepsilon_i = \rho_i \sigma_{\varepsilon_i}^2$$

(or from (B.109-111), $\hat{\rho}_i$ is a consistent estimator of ρ_i) ,

and since

$$(B.112)\ \underset{T\to\infty}{\text{plim}} \frac{1}{T}\varepsilon_i' X_i(-1) = \underset{T\to\infty}{\text{plim}} \frac{1}{T}\varepsilon_i'(z_{i1}(-1), y_i(-2)) = \left(0,\ \frac{\rho_i^2\sigma_{\varepsilon_i}^2}{(1-\gamma_{i1}\rho_i)}\right) \text{and}$$

$$\text{(B.113)}\ \underset{T\to\infty}{\text{plim}}\frac{1}{T}X_i'\varepsilon_i(-1)=\underset{T\to\infty}{\text{plim}}\frac{1}{T}\begin{pmatrix} z_{i1}' \\ y_i'(-1)\end{pmatrix}\varepsilon_i(-1)=\begin{pmatrix} o \\ \dfrac{\sigma^2_{\varepsilon_i}}{(1-\gamma_{i1}\rho_i)}\end{pmatrix},$$

so that from (B.109-B.113):

$$\text{(B.114)}\ \underset{T\to\infty}{\text{plim}}\sqrt{T}(\hat{\rho}_i-\rho_i)=-\frac{1}{\sigma^2_{\varepsilon_i}}\left[\left(0,\frac{\rho_i^2\sigma^2_{\varepsilon_i}}{(1-\gamma_{i1}\rho_i)}\right)H_i^{-1}\underset{T\to\infty}{\text{plim}}\frac{X_i^{\star\prime}\varepsilon_i}{\sqrt{T}}\right.$$

$$\left.+\underset{T\to\infty}{\text{plim}}\frac{\varepsilon_i'X_i^{\star}}{\sqrt{T}}\ H_i'^{-1}\begin{pmatrix} o \\ \dfrac{\sigma^2_{\varepsilon_i}}{(1-\gamma_{i1}\rho_i)}\end{pmatrix}\right]$$

and since (B.114) is a scalar quantity:

$$\text{(B.115)}\ \underset{T\to\infty}{\text{plim}}\sqrt{T}(\hat{\rho}_i-\rho_i)=-\left(0,\frac{(1+\rho_i^2)}{(1-\gamma_{i1}\rho_i)}\right)H_i^{-1}\underset{T\to\infty}{\text{plim}}\frac{X_i^{\star\prime}\varepsilon_i}{\sqrt{T}}\ .$$

Substituting (B.115) into the asymptotic distribution of $\sqrt{T}(\hat{\hat{\beta}}_i-\beta_i)$:

$$\text{(B.116)}\ \underset{T\to\infty}{\text{plim}}\sqrt{T}(\hat{\hat{\beta}}_i-\beta_i)=V_i\underset{T\to\infty}{\text{plim}}\frac{X_i'\Omega_{ii}^{-1}\varepsilon_i}{\sqrt{T}}+V_iQ_iH_i^{-1}\underset{T\to\infty}{\text{plim}}\frac{X_i^{\star\prime}\varepsilon_i}{\sqrt{T}},$$

$$\text{with } Q_i=\begin{bmatrix} 0 & 0 \\ 0 & \dfrac{(1+\rho_i^2)}{(1-\gamma_{i1}\rho_i)^2}\end{bmatrix} \qquad \text{or}$$

$$\text{(B.117)}\ \underset{T\to\infty}{\text{plim}}\ T\left[(\hat{\hat{\beta}}_i-\beta_i)(\hat{\hat{\beta}}_i-\beta_i)'\right]=V_i\underset{T\to\infty}{\text{plim}}\frac{X_i'\Omega_{ii}^{-1}\varepsilon_i}{\sqrt{T}}\underset{T\to\infty}{\text{plim}}\frac{\varepsilon_i'\Omega_{ii}^{-1}X_i}{\sqrt{T}}V_i\ +$$

$$V_i\underset{T\to\infty}{\text{plim}}\frac{X_i'\Omega_{ii}^{-1}\varepsilon_i}{\sqrt{T}}\underset{T\to\infty}{\text{plim}}\frac{\varepsilon_i'X_i^{\star}}{\sqrt{T}}H_i'^{-1}Q_iV_i\qquad +$$

$$V_iQ_iH_i^{-1}\underset{T\to\infty}{\text{plim}}\frac{X_i^{*\prime}\varepsilon_i}{\sqrt{T}}\underset{T\to\infty}{\text{plim}}\frac{\varepsilon_i^{\prime}\Omega_{ii}^{-1}X_i}{\sqrt{T}}V_i \quad +$$

$$V_iQ_iH_i^{-1}\underset{T\to\infty}{\text{plim}}\frac{X_i^{*\prime}\varepsilon_i}{\sqrt{T}}\underset{T\to\infty}{\text{plim}}\frac{\varepsilon_i^{\prime}X_i^{*}}{\sqrt{T}}H_i^{\prime -1}Q_iV_i \quad \text{and by (B.97):}$$

$$=V_i+2V_iQ_iV_i+V_iQ_i\phi_iQ_iV_i \ ,$$

where, implicitly, asymptotic expectations have been taken and

$$\phi_i=H_i^{-1}\underset{T\to\infty}{\text{plim}}\left(\frac{X_i^{*\prime}\Omega_{ii}X_i^{*}}{T}\right)H_i^{\prime -1}$$ is the asymptotic

variance- covariance matrix of the consistent first round instrumental variable estimator $\hat{\beta}$. Since $2V_iQ_iV_i+V_iQ_i\phi_iQ_iV_i$ is positive definite, there is a loss in asymptotic efficiency when comparing $\hat{\hat{\beta}}_i$ and $\tilde{\beta}_i$, i.e. of the feasible w.r.t. the "usual" Aitken estimator. The same can be said w.r.t. the initial instrumental variable estimator since $V_i+2V_iQ_iV_i$ is positive definite. Only if $\rho_i=0$, there is no loss in asymptotic efficiency.

Appendix C A likelihood ratio test on vector equality with error variance covariance matrix of arbitrary rank.

Proposition C 1

The "test statistic" on the vector equality hypothesis ($H_o:\beta_1=\beta_2=\ldots\ldots=\beta_M$):

$$(C.1)\ \frac{s-Mk}{q}\ \cdot\ \frac{\tilde{\beta}'C'\left[C(X'\Omega^{+}X)^{-1}C'\right]^{-1}C\tilde{\beta}}{y'\Omega^{+}y-y'\Omega^{+}\tilde{y}}$$

is F distributed with q and s-Mk degrees of freedom, q being the number of restrictions, s the rank of the variance covariance matrix Ω and k the number of explanatory variables in each equation (or here: $q=(M-1)k$).

The known matrix of restrictions is defined as:

$$(C.2)\ C=\begin{bmatrix} I_k & -I_k & 0\ldots\ldots 0 \\ 0 & I_k & -I_k\ldots\ldots 0 \\ \vdots & \vdots & \vdots \\ 0 & 0\ldots\ldots I_k & -I_k \end{bmatrix}$$

Proof (*)

a) Under the null hypothesis, the system of linear equations becomes:

$$(C.3a)\ \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix}=\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_M \end{bmatrix}\beta_1+\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_M \end{bmatrix} \qquad \text{or}$$

$$(C.3b)\ y=Z\beta_1+\varepsilon$$

Following proposition 1.1 an $(s\times MT)$-transformation matrix G exists such that $E(G\varepsilon\varepsilon'G')=\sigma_1^2 G\Omega G'=\sigma_1^2 I_s$ and $G'G=\Omega^{+}$.

Putting $Gy=\dot{y}$, $GZ=\dot{Z}$ and $G\varepsilon=\dot{\varepsilon}$, the likelihood function under H_o becomes:

$$(C.4)\ L_1=\frac{1}{(2\pi\sigma_1^2)^{\frac{s}{2}}}\exp\left(-\frac{1}{2\sigma_1^2}\dot{\varepsilon}'\dot{\varepsilon}\right)$$ with concentrated likelihood:

(*) See also A.Zellner [26], Appendix A.

(C.5) $L_1^{*} = \frac{1}{(2\pi\hat{\sigma}_1^2)^{\frac{s}{2}}} \exp(-\frac{1}{2}s)$ where

(C.6) $\hat{\sigma}_1^2 = \frac{\hat{\dot{\varepsilon}}'\hat{\dot{\varepsilon}}}{s} = \frac{(\dot{y}-\dot{Z}\hat{\beta}_1)'(\dot{y}-\dot{Z}\hat{\beta}_1)}{s}$ with $\hat{\beta}_1 = (\dot{Z}'\dot{Z})^{-1}\dot{Z}'\dot{y}$

b) Under the hypothesis that there are <u>no</u> restrictions on the coefficients, we find, putting $GX=\dot{X}$ in model (1.64), the likelihood function:

(C.7) $L_2 = \frac{1}{(2\pi\sigma_2^2)^{\frac{s}{2}}} \exp(-\frac{1}{2\sigma_2^2}\dot{\varepsilon}'\dot{\varepsilon})$ with concentrated likelihood:

(C.8) $L_2^{*} = \frac{1}{(2\pi\hat{\sigma}_2^2)^{\frac{s}{2}}} \exp(-\frac{1}{2}s)$ where

(C.9) $\hat{\sigma}_2^2 = \frac{\hat{\dot{\varepsilon}}'\hat{\dot{\varepsilon}}}{s} = \frac{(\dot{y}-\dot{X}\hat{\beta})'(\dot{y}-\dot{X}\hat{\beta})}{s}$ with $\hat{\beta} = (\dot{X}'\dot{X})^{-1}\dot{X}'\dot{y}$.

From (C.5) and (C.8), the estimated likelihood ratio is then:

(C.10) $\ell = \frac{\max_{H_0} L_1}{\max_{H_1} L_2} = \frac{L_1^{*}}{L_2^{*}} = \left(\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2}\right)^{-\frac{1}{2}s}$ or

(C.11) $-2\ln \ell = s \ln \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2}$ $(0 \leq l \leq 1)$,

which is asymptotically distributed as $\chi^2_{(M-1)k}$, because by defining a singular multivariate normal distribution on each $\varepsilon(t)$ vector, we obtain, by the non singular transformation G of ε, a non singular multivariate normal distribution on each $\dot{\varepsilon}(\theta)$ vector for which the standard distribution properties for likelihood ratios can be applied.

Now in order to complete the proof, we have to show that:

(C.12) $\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} = 1 + \frac{q}{s-Mk} F_{q,s-Mk}$ or alternatively:

$$(C.13)\ \frac{q}{s-Mk}F_{q,s-Mk}=\frac{\hat{\sigma}_1^2-\hat{\sigma}_2^2}{\hat{\sigma}_2^2}=\frac{(\dot{y}-\dot{Z}\hat{\beta}_1)'(\dot{y}-\dot{Z}\hat{\beta}_1)-(\dot{y}-\dot{X}\hat{\beta})'(\dot{y}-\dot{X}\hat{\beta})}{(\dot{y}-\dot{X}\hat{\beta})'(\dot{y}-\dot{X}\hat{\beta})}$$

$$=\frac{\dot{y}'\dot{X}(\dot{X}'\dot{X})^{-1}\dot{X}'\dot{y}-\dot{y}'\dot{Z}(\dot{Z}'\dot{Z})^{-1}\dot{Z}'\dot{y}}{\dot{y}'\dot{y}-\dot{y}'\dot{X}(\dot{X}'\dot{X})^{-1}\dot{X}'\dot{y}}$$

$$=\frac{y'\Omega^+X(X'\Omega^+X)^{-1}X'\Omega^+y-y'\Omega^+Z(Z'\Omega^+Z)^{-1}Z'\Omega^+y}{y'\Omega^+y-y'\Omega^+X(X'\Omega^+X)^{-1}X'\Omega^+y}$$

$$=\frac{y'\Omega^+\left[X(X'\Omega^+X)^{-1}X'-Z(Z'\Omega^+Z)^{-1}Z'\right]\Omega^+y}{y'\Omega^+y-y'\Omega^+X(X'\Omega^+X)^{-1}X'\Omega^+y}$$

with $F_{q,s-Mk}$ equal to (C.1).

Since

$$(C.14)\ Z=\begin{bmatrix}X_1\\X_2\\\vdots\\X_M\end{bmatrix}=\begin{bmatrix}X_1 & 0 & \cdots & 0\\0 & X_2 & \cdots & 0\\\vdots & & \ddots & \vdots\\0 & 0 & \cdots & X_M\end{bmatrix}\begin{bmatrix}I_k\\I_k\\\vdots\\I_k\end{bmatrix}=XJ\quad,$$

with J an (Mk×k) matrix consisting of M (k×k) unitary matrices, the numerator of (C.13) may be written as:

$$(C.15)\ y'\Omega^+X\left[(X'\Omega^+X)^{-1}-J(J'X'\Omega^+XJ)^{-1}J'\right]X'\Omega^+y=$$

$$=y'\Omega^+X(X'\Omega^+X)^{-1}\left[(X'\Omega^+X)-(X'\Omega^+XJ)(J'X'\Omega^+XJ)^{-1}(J'X'\Omega^+X)\right](X'\Omega^+X)^{-1}X'\Omega^+y$$

or to proof (C.1) it has to be verified that:

$$(C.16)\ C'\left[C(X'\Omega^+X)^{-1}C'\right]^{-1}C=(X'\Omega^+X)-(X'\Omega^+XJ)(J'X\Omega^+XJ)^{-1}(J'X'\Omega^+X),$$

which is true since from premultiplication of (C.16) with $C(X'\Omega^{+}X)^{-1}$ and postmultiplication by J:

$$(C.17)\quad C=C-CJ(J'X'\Omega^{+}XJ)^{-1}(J'X'\Omega^{+}X) \qquad \text{or}$$

$$(C.18)\quad CJ=C(I_{Mk}J-J)=CJ-CJ$$

it follows from the definition (C.2) of the matrix C that both sides of (C.18) are equal to zero matrices. △

Corollary C 1

The quantities $s\ln\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2}$ and $qF_{q,s-Mk}=(M-1)kF_{(M-1)k,s-Mk}$ are both asymptotically distributed as χ^2_q .

Proof

Since (C.12) is

$$(C.19)\quad -2\ln\ell=s\ln\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2}=s\ln\left(1+\frac{q}{s-Mk}F_{q,s-Mk}\right)$$

$$=\frac{sq}{s-Mk}F_{q,s-Mk}-s\left(\frac{q}{s-Mk}\right)^2F^2_{q,s-Mk}+s\left(\frac{q}{s-Mk}\right)^3F^3_{q,s-Mk}=\dots$$

$$=qF_{q,s-Mk}+0(s^{-1}) \quad .$$

or from the convergence theorem in Cramèr H. [6],p. 254, the corollary is proved △

Proposition C 2

If Ω is unknown and a consistent estimate of it is employed, the resultant test statistic, say F^{*}, has the same asymptotic probability distribution as $F_{q,s-Mk}$.

Proof (*)

If it is shown that

(*) See A.Zellner [27], Appendix B for an alternative proof.

$$(C.20)\quad F^{*}=\frac{s-Mk}{q}\cdot\frac{\beta^{*'}C'\left[C(X'\hat{\Omega}^{+}X)^{-1}C'\right]^{-1}C\beta^{*}}{y'\hat{\Omega}^{+}y-y'\hat{\Omega}^{+}y^{*}}\quad\text{, with}$$

$y^{*}=X\beta^{*}=X(X'\hat{\Omega}^{+}X)^{-1}X'\hat{\Omega}^{+}y$ and $\hat{\Omega}$ a consistent estimate of Ω, has probability limit $F_{q,s-Mk}$ defined in (C.1), we may conclude that F^{*} and $F_{q,s-Mk}$ have the same limiting distribution.

This is easily established, utilizing the property that

(C.21) $\operatorname{plim}_{T\to\infty}\hat{\Omega}=\Omega$ **and Slutsky's theorem in the probability limit of (C.20).** Δ

Following corollary C 1, qF^{*} and $qF_{q,s-Mk}$ have the same asymptotic χ^2-distribution with q degrees of freedom.

REFERENCES

1 Amemiya T.and W.A.Fuller:"A Comparative Study of Alternative Estimators in a Distributed Lag Model", Econometrica, vol 35, 1967, pp. 509-529.

2 Anderson, T.W.: An Introduction to Multivariate Statistical Analysis, Wiley, New York, 1959, 374 pp.

3 Barten, A.P. and T.Kloek :"A Generalization of Generalized Least-Squares", Report 6508, Econometric Institute, Netherlands School of Economics, Rotterdam, July 1965, 15 pp. (unbublished)

4 Bellman, R.:Introduction to Matrix Analysis, Wiley,New York 1961.

5 Chung, K.L.:A Course in Probability Theory, Brave and World, Harcourt, New York 1968.

6 Cramèr, H.:Mathematical Methods of Statistics, Princeton University Press, Princeton 1951, 575 pp.

7 Dhrymes, P.H.:Econometrics: Statistical Foundations and Applications. Harper and Row, New York, 1970,592 pp.

8 Goldberger, A.S.: Econometric Theory, Wiley, New York, 1964, 399 pp.

9 Hoeffaing, W. and H. Robbins:" The Central Limit Theorem for dependent random Variables", Duke Mathematical Journal, vol 5, 1948, pp. 773-780

10 Kakwani, N.C.: "The Unbiasedness of Zellner's Seemingly Unrelated Regression Equations Estimators", JASA, vol 62, 1967, pp. 141-142.

11 Kendall, M.G. and A.S. Stuart: The Advanced Theory of Statistics, Vol.I : Distribution Theory, Griffin, London, 2nd ed. 1963, 433 pp.

12 Kmenta, J. and R.F. Gilbert: "Estimation of Seemingly Unrelated Regressions with Autoregressive Disturbances", JASA, vol 65, 1970, pp. 186-197

13 Lawley, D.N.:"Tests of Significance for the Latent Roots of Covariance and Correlation Matrices", Biometrika, vol 43, 1956, pp. 128-136

14 Lewis,T.O. and P.L. Odell: "A Generalization of the Gauss Markov Theorem", JASA, vol 61, 1966, pp. 1063-1066

15 Loève, M.: Probability Theory, Van Nostrand, Princeton, 1955, 685 pp.

16 Malinvaud, E. : Statistical Methods in Econometrics, Rand Mc.Nally and North-Holland, Chicago and Amsterdam, 1966, 631 pp.

17 Neudecker,H.:"The Kronecker Matrix Product and Some of its Applications in Econometrics", Statistica Neerlandica, vol 22, 1968, pp. 69-82.

18 Parks, R.W.: "Efficient Estimation of a System of Regression Equations when Disturbances are both Serially and Contemporaneously correlated", JASA, vol 62, 1967, pp. 500-509.

19 Rao, R.C. : Linear Statistical Inference and its Applications, Wiley, New York, 1965,

20 Rooy, P.J.L. van: 'Pseudorang, Approximanten en Pseudo-Inversen van Matrices met onbekende Rang", Afstudeerverslag, T.H. Eindhoven, 1970, 79 pp. (Dutch-unpublished)

21 Schönfeld, P.: " Generalized Best Linear Unbiased Estimation", Core Discussion Paper 6713, CORE, Louvain, Sept. 1967, 41 pp. (unpublished)

22 Thornber,H.: "An Autoregressive Model: Bayesian versus Sampling Theory Analyses", Technical Report no 6504, Center for Mathematical Studies in Business and Economics, University of Chicago, Chicago, 1965, 26 pp. (unpublished)

23 Vandaele, W.H. : "Zellner's Seemingly Unrelated Regression Equations Estimators: A Survey", EIT-Research Memorandum, Tilburg, 1970 (forthcoming)

24 Wallis, K.F. : "Lagged Dependent Variables and Serially Correlated Errors: A Reappraisal of Three-Pass Least-Squares", Review of Economics and Statistics, 1967, pp. 555-567

25 White, J.S. :" Asymptotic Expansions for the Mean and Variance of the Serial Correlation Coefficient", Biometrika, vol 48, 1961, pp. 85-94

26 Zellner, A. :" An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias", JASA, vol 57, 1962, pp. 348-368

27 Zellner, A. and D.S. Huang: "Further Properties of Efficient Estimators for Seemingly Unrelated Regression Equations", International Economic Review, vol 3, 1962, pp. 300-313

28 Zellner, A. : "Estimators for Seemingly Unrelated Regression Equations: Some Exact Finite Sample Results", JASA, vol 58, 1963, pp. 977-992.

PREVIOUS NUMBERS:

EIT 1	J. Kriens *)	Het verdelen van steekproeven over subpopulaties bij accountantscontroles.
EIT 2	J. P. C. Kleynen *)	Een toepassing van „importance sampling".
EIT 3	S. R. Chowdhury and W. Vandaele *)	A bayesian analysis of heteroscedasticity in regression models.
EIT 4	Prof. drs. J. Kriens *)	De besliskunde en haar toepassingen.
EIT 5	Prof. dr. C. F. Scheffer	Winstkapitalisatie versus dividendkapitalisatie bij het waarderen van aandelen.
EIT 6	S. R. Chowdhury	A bayesian approach in multiple regression analysis with inequality constraints.
EIT 7	P. A. Verheyen	Investeren en onzekerheid.
EIT 8	R. M. J. Heuts en Walter H. Vandaele	Problemen rond niet-lineaire regressie.
EIT 9	S. R. Chowdhury	Bayesian analysis in linear regression with different priors.
EIT 10	A. J. van Reeken	The effect of truncation in statistical computation.
EIT 11	W. H. Vandaele and S. R. Chowdhury	A revised method of scoring.
EIT 12	J. de Blok	Reclame-uitgaven in Nederland.
EIT 13	Walter H. Vandaele	Mødscø, a computer programm for the revised method of scoring.
EIT 14	J. Plasmans	Alternative production models. (Some empirical relevance for postwar Belgian Economy)
EIT 15	D. Neeleman	Multiple regression and serially correlated errors.
EIT 16	H. N. Weddepohl	Vector representation of majority voting.
EIT 17	Walter H. Vandaele	Zellner's seemingly unrelated regression equations estimators: a survey.
EIT 18	J. Plasmans	The general linear seemingly unrelated regression problem. I. Models and inference.

*) not available

EIT 1970