Preliminary Estimation of Transfer Function Weights

A Two-Step Regression Approach

STOCKHOLM SCHOOL OF ECONOMICS

The Economic Research Institute

Address: Sveavägen 65, Box 6501, S-113 83 Stockholm, tel 08-736 90 00

Basic Orientation

The Economic Research Institute, at the Stockholm School of Economics (EFI) is devoted to the scientific study of problems in management science and economics. It provides research facilities for scholars belonging to these disciplines, training advanced students in scientific research. The studies to be carried out by the Institute are chosen on the basis of their expected scientific value and their relevance to the research programs outlined by the different sections of the Institute.

Research Divisions:

A	Work Life Issues and Personnel
B	Accounting and Managerial Finance
C	Managerial Economics
D	Marketing, Distribution and Industry Dynamics
ES	Economic Statistics
FI	Finance
F	Public Administration Research Program in Policy Sciences
IEG	International Economics and Geography
I	Applied Information Systems
P	Economic Psychology
S	Economics
CFR	Center for Risk Research

Independent Research Programs:

Man and Organization
The Foundation for Distribution Research

Additional information about research in progress and published reports is described in our project catalogue. The catalogue can be ordered directly from The Economic Research Institute, Box 6501, S-113 83 Stockholm, Sweden.

Per-Olov Edlund

Preliminary Estimation of Transfer Function Weights

A Two-Step Regression Approach

THE ECONOMIC
RESEARCH INSTITUTE

A Dissertation for the
Doctor's Degree in Economic Statistics
Stockholm School of Economics 1989

ISBN 91-7258-291-X
UDK 311.16
330.33.015

Gotab, Stockholm 1989 89797

ACKNOWLEDGEMENTS

First I would like to thank the members of my advisory committee, Professor Anders H. Westlund and Professor Emeritus Erik Ruist at the Stockholm School of Economics and Professor Tarmo Pukkila at the University of Tampere, Finland, for their advice and support during my thesis work. I am also grateful to The Swedish Research Council for Humanities and Social Sciences (HSFR) and the Bank of Sweden Tercentenary Foundation (Riksbankens Jubileumsfond) for their financial support.

Stockholm, August 1989 Per–Olov Edlund

CONTENTS

---

1 Please see next page for previous publication or presentation of the papers.

The four papers have been published or presented as follows:

[A] "On Identification of Transfer Function Models", Research Report 1987:2, Department of Statistics, University of Stockholm. Thesis for the "fil.lic." degree. (Enlarged version of "Identification of the Multi–input Box–Jenkins Transfer Function Model", *Journal of Forecasting, 1984, 3*, p 297–308.)

[B] "On Identification of Transfer Function Models by Biased Regression Methods". Reprint from *Journal of Statistical Computation and Simulation*, vol 31, 1989, pp. 131–148.

[C] "Ridge Estimation of Transfer Function Weights". This is a revised version of Research Paper 6380, Economic Research Institute, Stockholm School of Economics, February 1989. The paper has been conditionally accepted for publication in *Communications in Statistics – Simulation and Computation.*

[D] "On Identification of Transfer Function Models for Business Cycle Forecasting". This is a revised version of a paper presented at "The Ninth International Symposium on Forecasting" in Vancouver, Canada, June 18–21 1989.

1. Introduction

In economic time series modelling, we frequently have to model dynamic relationships where the explanatory variables influence the dependent variable over more than one time period. In traditional econometric models, this usually means inclusion of one or more lagged values of the explanatory variables as well as current values. Such dynamic relationships are found in business cycle forecasting with leading indicators, in marketing models describing the relationship between advertising and sales (carry–over effects) and in many traditional econometric models. There are different ways to model the lag structure. In econometrics, this has traditionally been done by using Almon lags, Koyck schemes or rational lag distributions (e.g., see Johnston 1984, Ch. 9).

In most cases, there is little or nothing in the economic theory that could be used to specify, or in time series terminology, identify the model. In such cases, trial–and–error must be used to obtain a reasonable description of the dynamic relationship. When estimating the identified model, it is often found that the residuals are autocorrelated.

The transfer function model approach proposed by Box and Jenkins (1976) includes methods for identification of the transfer function model as well as coping with the negative effects of autocorrelated residuals. The ordinary single equation regression model is found to be a special case of the more general transfer function model.

Box and Jenkins mainly discuss identification of the single–input transfer function and give very little attention to the more general multi–input model identification. It is therefore interesting to investigate methods for multi–input identification.

The main purpose of this thesis is to describe and evaluate a two–step regression approach proposed by the author (Edlund, 1984). This method is intended for preliminary estimation of the transfer function weights for multi–input transfer function models.

Section 2 gives an introduction to transfer function models and their identification. In Section 3 the two–step regression model is presented and explained. Section 4 summarizes the results from three simulation studies that have been performed to evaluate the regression approach. An empirical comparison between the regression approach and the prewhitening and the cross spectral approaches is presented in Section 5. Finally, some concluding remarks and suggestions for further research are given in Section 6.

2. Transfer function models

The transfer function–noise model proposed by Box and Jenkins (1976) is

$$y_t = c + \frac{\omega(B)}{\delta(B)} x_{t-b} + n_t$$

where y_t is the dependent variable, c is a constant term, x_t is the independent variable, n_t is an error term which represents all the 'missing' x variables plus the pure noise, B is the ordinary lag operator, $\omega(B)$ is a 'moving average' operator, $\delta(B)$ is an 'autoregressive' operator, b is a pure delay parameter which represents the number of complete time intervals before a change in x_t begins to have an effect on y_t. (The y and x variables are assumed to be differenced/transformed to be mean and variance stationary.)

The transfer function $\nu(B)$ is a rational lag structure

$$\nu(B) = \frac{\omega(B)}{\delta(B)} B^b \quad \text{or} \quad \nu(B) = (\nu_0 + \nu_1 B + \nu_2 B^2 + \dots) B^b$$

that may be used to represent any form of linear dynamic relationship between x_t and y_t to any specified degree of accuracy. The weight ν_i is called the impulse response weight at lag i.

The noise term, n_t, may be expressed as an ordinary $ARMA(p,q)$ model of the form

$$n_t = \frac{\theta(B)}{\phi(B)} a_t$$

where $\theta(B)$ is a moving average operator of order q, $\phi(B)$ is an autoregressive operator of order p, and a_t a white noise variable.

The single–input model is easily generalized to an m–input transfer function model,

$$y_t = c + \sum_{j=1}^{m} \frac{\omega_j(B)}{\delta_j(B)} x_{j,t-b_j} + \frac{\theta(B)}{\phi(B)} a_t$$

with m input variables $x_{1t}, x_{2t}, \dots, x_{mt}$. The ordinary multiple regression model,

$$y_t = c + \nu_{10} x_{1t} + \nu_{20} x_{2t} + \dots + \nu_{m0} x_{mt} + a_t$$

is a special case of the more general multi–input transfer function model.

2.1 Identification of transfer function models

To identify the order of the operators $\delta(B)$, $\omega(B)$ and b two stages are performed. First, the impulse response weights ν_i, are preliminarily estimated. Then the pattern of the weights is used to specify the order of the operators (see Box and Jenkins 1976 Fig. 10.6, p 349).

To simplify the work of guessing the suitable order of the operators it seems reasonable to try to get the 'best' possible preliminary estimates of transfer function weights. In this thesis, only the first stage of the identification procedure will be discussed.

2.2 Preliminary estimation of transfer function weights

Box and Jenkins suggested different methods for preliminary estimation of the transfer function weights. They preferred the prewhitening cross–correlation approach for the one–input model. They showed that the weights could also be estimated by Ordinary Least–Squares (*OLS*) but that these estimates have several deficiencies (see Section 3). For the multiple–input case, Box and Jenkins gave a cross–spectral estimator (Box and Jenkins 1976, Ch. 11). This estimator has been investigated and made operational by Pukkila (1982).

Priestley (1971) proposed the covariance contraction method, which is similar to the prewhitening cross–correlation method but where all variables are prewhitened separately. The above methods are described in paper [A] Section 2.2. Haugh and Box (1977) presented a method similar to Priestley's method. Both methods are suitable for the one–input model, but Fask and Robinson (1977) extended Priestley's method to multiple–input models. Tsay (1985) suggested the use of a *VAR* model which would avoid prewhitening and allow for non–stationary variables. This approach would also be useful for testing the hypothesis of unidirectional causality. The *VAR* approach is also suitable for identification of multi–input models.

Another approach not requiring prewhitening of the variables is the Maximum Entropy–Generalized Least–Squares (*ME–GLS*) estimator proposed by Rahiala (1986). First, the residual model part of the transfer function model is estimated from estimates of the cross–spectra of the variables by a maximum entropy method. The impulse response weights are then estimated by an ordinary *GLS* estimator.

Of all these approaches the regression estimator *(OLS)* has the advantage of being rather simple to use, but there are a few estimation problems that have to be taken care of in order to improve the estimates.

3. The two-step regression method

The one–input transfer function–noise model in Section 2 may also be expressed as

$$y_t = c + \nu_0 x_t + \nu_1 x_{t-1} + \nu_2 x_{t-2} + \ldots + n_t$$

This model resembles the ordinary regression model even though there are some differences: the number of parameters to estimate may be infinite and the residuals may be autocorrelated. Also the explanatory variables are functionally related through the autocorrelation function of the x variable. If it would be possible to assume that $\nu_j \approx 0$ for $j > K$, the model could be estimated by ordinary regression methods. Unfortunately, the estimates would have low precision due to multicollinearity caused by the autocorrelation structure of the independent variable. There would also be a loss in efficiency due to the autocorrelation structure of the residuals.

The simple regression model can easily be extended to include more than one input variable, which would make it suitable for preliminary estimation of impulse response weights in multi–input models. In this case, there will be a second source of multicollinearity among the explanatory variables: cross correlation between the input variables.

Pukkila (1980) investigated the regression method without dealing with the multicollinearity and the autocorrelated residuals. His results were quite promising even though the multicollinearity was not very strong. Liu and Hanssens (1982) solved the multicollinearity problem by transforming the y and x variables by a common filter. The filter was constructed to eliminate AR factors with roots close to 1 in the $ARMA$ processes for the x variables. Then generalized least–squares were used to avoid the effects of autocorrelated residuals. Erickson (1981) used the ridge estimator to directly estimate the transfer function weights but did not adjust his estimates for the autocorrelated residuals.

There are several biased regression estimators that can be used to decrease the effects of multicollinearity. Among them are the principal component estimator and the ridge estimators. The principal component estimator works by reducing the rank of the $\mathbf{X'X}$ matrix (or correlation matrix of the x variables) by omitting one or more principal components. There are different ways of choosing the components that should be dropped, see paper [A], Section 3.2.1. The ridge estimator retains all variables; reduction of the multicollinearity is accomplished by adding a small quantity k to the diagonal elements of the correlation matrix before inversion. This

makes the computations numerically more stable, but at the same time the estimates will be biased, see [A] Section 3.2.2.

There are several ways to adjust the estimates for autocorrelated residuals. Apart from Generalized Least–Squares (*GLS*) there are transformation methods as the Cochrane–Orcutt's method and Durbin's method (e.g. Johnston, 1984). These methods assume that the autocorrelation structure can be described by a low order *AR* model (e.g. an *AR*(1) model). The *AR* model is estimated and used to transform the variables. The regression residuals will then be uncorrelated and the efficiency of the estimates increases.

In paper [A] Section 4, a two–step procedure is described that will use biased regression and transformation of variables to reduce the effects of multicollinearity and autocorrelated residuals respectively.

Step 1: ***Identification, estimation and checking of the noise model and transformation of the input and output variables.***

To identify the noise model an estimate of the noise series n_t is needed. The noise series is estimated from

$$\hat{n}_t = y_t - \hat{c} - \sum_{i=1}^{m} \hat{\nu}(B)x_{it} = y_t - \hat{c} - \hat{\nu}_{10}x_{1t} - \hat{\nu}_{11}x_{1,t-1} - \ldots - \hat{\nu}_{mK}x_{m,t-K}$$

where the ν weights are estimated by biased regression. From the estimated n_t, an *ARMA* model can be identified, estimated and checked using the ordinary Box–Jenkins procedures. The estimated model

$$\hat{n}_t = \frac{\hat{\theta}(B)}{\hat{\phi}(B)}\hat{a}_t$$

is then used to transform the original variables

$$\hat{\theta}(B)y'_t = \hat{\phi}(B)y_t \quad \text{all } t \text{ and } \hat{\theta}(B)x'_{jt} = \hat{\phi}(B)x_{jt} \quad j=1,\ldots,m, \text{ all } t$$

Step 2: ***Estimation of the impulse response function from the transformed variables*** y'_t **and** x'_{jt}.

In the second step, the impulse response weights are re–estimated by biased regression on the transformed variables (the transformations do not alter the relationship between x and y)

$$y'_t = c' + \nu_{10}x'_{1t} + \nu_{11}x'_{1,t-1} + \ldots + \nu_{mK}x'_{m,t-K} + a_t$$

Now the residuals a_t will be close to white noise. By using this approach, the

estimate of the residual variance will be unbiased and the estimates of the impulse response weights, ν_{ij} will be less influenced by the multicollinearity. The significance of the weights can be tested by a t–test. In case the residuals are not white noise, the ν weights in the second step can be used to compute the residuals as in the first step. Then the rest of step 1 and 2 can be repeated until the residuals are sufficiently close to white noise.

4. Simulation studies of biased regression estimators

In this Section the three simulation studies in papers [A], [B] and [C] are presented. The first two papers compare principal component and ridge estimators to the ordinary least–squares estimator (*OLS*) for similar experimental designs. The third paper compares five different ridge estimators over 5000 different two–input models, drawn from a large set of possible transfer function models.

4.1 Principal component vs. ridge estimators

The simulation study in paper [A] is an enlarged version of the study by Edlund (1984). The purpose of the study is to investigate the proposed two–step method by comparing one ridge and one principal component estimator to the *OLS* estimator, . The results are obtained both for transformed and untransformed data (Step 1 only).

The study is limited to one model

$$y_t = (2+4B+B^2)x_{1t} + (1-0.6B)^{-1}B^3x_{2t} + n_t$$

where

$$\begin{aligned} n_t &= a_t + 0.75a_{t-1} \\ x_{1,t} &= 0.7x_{1,t-1} + a_{1t} \\ x_{2,t} &= 1.25x_{2,t-1} - 0.75x_{2,t-2} + a_{2t} \end{aligned}$$

The processes a_{1t}, a_{2t} and a_t are normal white noise processes such that a_t is uncorrelated with a_{1t} and a_{2t}, while a_{1t} and a_{2t} are correlated (with covariance matrix Σ).

This two–input model has been used both by Pukkila (1980, 1982) and Damsleth (1979) and can be seen as a model neither too difficult nor too easy to identify. It contains both *AR* and *MA* operators in the transfer functions.

The ridge k value is computed by the Lawless & Wang (1976) method, i.e., $k = p\hat{\sigma}^2/\Sigma\lambda_i\hat{\alpha}_i^2$, where p is the number of explanatory variables, $\hat{\sigma}^2$ is *OLS* estimate of the

standardized σ^2 for the n_t process, λ_i is the i:th eigenvalue of the correlation matrix of the independent variables, and $\hat{\alpha}_i^2$ is the *OLS* estimate of α_i^2 (α_i is the i:th coefficient from regression on principal components). The principal component estimator chooses the first r principal components corresponding to the r largest eigenvalues. The value of r is chosen so that the r components explain at least 99.5 % of the variance of the standardized x variables.

The experimental design covers three different levels of multicollinearity and three levels of signal–to–noise ratio. The degree of multicollinearity is determined by varying the correlation between a_{1t} and a_{2t} (ρ = .58, .87 and .98). The signal–to–noise ratio is determined by varying the residual variance (σ^2 = .01, 1 and 25). 100 observations have been generated for each time series. For each variable, coefficients for lag 0–10 are estimated 50 times for each combination of ρ and σ^2. To simplify the computations, the true rather than estimated model for the residuals is used to transform the y and x variables.

The optimal way of comparing the estimators would be to measure their ability to make the forecaster specify the correct transfer function model in the second stage of the identification procedure. Since the choice of model is rather subjective, it is impractical to include such a measure in a simulation study. Therefore, as a proxy, the *MSE* is used to measure the closeness between the estimated and the true ν weights. It is then assumed that estimates close to the true weights improve the probability of identifying the true model. The following statistics will be used in Tables 4.1 and 4.2. For each estimator

$$MSE = \overline{MSE}(\hat{\nu}) / \overline{MSE}(\hat{\nu})_{OLS} \qquad S = S(\hat{\nu}) / S(\hat{\nu})_{OLS}$$

where,

$$\overline{MSE}(\hat{\nu}) = \sum_{j=1}^{22} MSE(\hat{\nu})_j / 22 \qquad S(\hat{\nu}) = \sum_{j=1}^{22} S(\hat{\nu})_j / 22$$

and

$$MSE(\hat{\nu})_j = \sum_{i=1}^{50} (\hat{\nu}_{ij} - \nu_j)^2 / 50 \qquad S(\hat{\nu})_j = \sum_{i=1}^{50} s_{ij} / 50 \quad j = 1, \ldots, 22$$

where s_{ij} is the estimated standard error of regression coefficient j in replication i.

The computational algorithm is described in [A] Section 5.4. In [A] Section 5.5 the results for a single replication are described.

Table 4.1 shows the total results from the simulations in terms of average *MSE* for each estimator relative to the *MSE* for the *OLS* estimator for untransformed variables (no correction for autocorrelated residuals). There are several interesting results. The ridge estimator is the best estimator for all combinations of ρ and σ^2.

Table 4.1 Estimated ratios of $\overline{MSE}(\hat{\nu})$ for each estimator relative to $\overline{MSE}(\hat{\nu})_{OLS}$ (PC = Principal component, t = transformed)

Estimator	ρ = .58			ρ = .87			ρ = .98		
	σ =.1	σ = 1	σ = 5	σ =.1	σ = 1	σ = 5	σ =.1	σ = 1	σ = 5
OLS	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Ridge	1.00	0.96	0.57	0.99	0.89	0.33	0.99	0.53	0.08
PC	25.21	1.21	0.82	81.04	1.49	0.29	152.48	1.71	0.09
OLS(t)	0.64	0.82	0.82	0.72	0.83	0.83	0.83	0.85	0.85
Ridge(t)	0.64	0.79	0.50	0.72	0.76	0.32	0.82	0.52	0.08
PC(t)	76.54	2.00	0.73	86.50	1.70	0.44	188.79	2.11	0.10
$\overline{MSE}(\hat{\nu})_{OLS}$	.0001	.009	.217	.0003	.021	.053	.001	.133	3.320

Table 4.2 Estimated ratios of average standard error, $\bar{S}(\hat{\nu})$ for each estimator relative to $\bar{S}(\hat{\nu})_{OLS}$ (PC = Principal component, t = transformed)

Estimator	ρ = .58			ρ = .87			ρ = .98		
	σ =.1	σ = 1	σ = 5	σ =.1	σ = 1	σ = 5	σ =.1	σ = 1	σ = 5
OLS	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Ridge	1.00	0.97	0.66	1.00	0.91	0.44	0.99	0.61	0.17
PC	1.29	0.77	0.76	0.96	0.38	0.37	0.90	0.15	0.12
OLS(t)	0.68	0.71	0.71	0.70	0.71	0.71	0.70	0.70	0.70
Ridge(t)	0.68	0.69	0.53	0.70	0.67	0.38	0.70	0.52	0.14
PC(t)	1.81	0.66	0.63	1.47	0.49	0.47	0.60	0.11	0.09
$\bar{S}(\hat{\nu})_{OLS}$	.013	.117	.587	.019	.183	.915	.046	.457	2.286

Transformation of the variables reduces the *MSE* for the ridge and *OLS* estimators. This is not necessarily true for all models. In some cases the transformation may increase the multicollinearity. The relative *MSE* for the ridge estimator decreases as the multicollinearity increases when $\sigma = 1$ or 5 but increases as $\sigma = .1$. The pattern for different σ when holding ρ constant is less clear even though the lowest relative *MSE* is found for $\sigma = 5$. The reduction in *MSE* for the most difficult case when

$\rho = .98$ and $\sigma = 5$ is most impressive for the ridge estimator, a 92 % reduction. In that case, the *OLS* on untransformed variables is totally useless, the average *MSE* is 3.320 compared to the size of the transfer function weights, of which most are less than 1 in absolute value. The principal component estimator gives low *MSE* values only when $\sigma = 5$. When the residual variance is low this estimator gives considerably larger *MSE* values. This may be due to the way the principal components are chosen. The choice of components is not sensitive to the residual variance or the orientation of the coefficient vector relative to the eigenvectors.

As a second measure of performance the average estimated standard error of the coefficients is shown in Table 4.2. The average s.e. of the *OLS* and ridge estimators decreases after transformation and has minimum S values when estimation is difficult (strong multicollinearity and/or large residual variance). The pattern for the principal component estimator is more difficult to describe. In some cases the values are larger after transformation, in other cases lower. The values decrease as the multicollinearity increases and/or signal–to–noise ratio decreases. The reduction in s.e. is quite substantial in the most difficult cases, the s.e. of the ridge estimator is reduced to only 14 % when $\rho = .98$ and $\sigma = 5$. The main benefit from transformation is that we obtain unbiased estimates of the residual variance, σ^2. The average estimates of the residual standard deviation for the *OLS* estimator for estimation on untransformed and transformed data are given in Table 4.3. As can be seen in the table, the transformation gives practically unbiased estimates of the residual standard deviation. (It should be noted that the true rather than estimated models for the residuals were used.)

The results from this simple simulation study are encouraging for the ridge estimator and for the two–step procedure. The results for the principal component estimator indicate that it might be quite dangerous to use it, especially for cases where the multicollinearity is weak and the signal–to–noise ratio is high. As the negative results for the principal component estimator may be dependent on the way the components were chosen, further studies of the two–step method are called for.

In paper [B] the number of estimators has been increased, as well as the number of models. The models now include both a one–input and a three–input model as well as the same two–input model as in [A]

Model 1: $$y_t = (2+4B+B^2)x_{1t} + n_t$$

Model 2: $$y_t = (2+4B+B^2)x_{1t} + (1-0.6B)^{-1}B^3x_{2t} + n_t$$

Model 3: $$y_t = (2+4B+B^2)x_{1t} + (1-0.6B)^{-1}B^3x_{2t} + 2(1-0.8B)^{-1}B^5x_{3t} + n_t$$

Table 4.3 Estimated residual standard deviation for *OLS* on untransformed and transformed data.

True standard deviation	Correlation between a_{1t} and a_{2t}	Data series Untransformed	Transformed
σ = .1	ρ = .58	0.136	0.108
	ρ = .87	0.129	0.103
	ρ = .98	0.123	0.100
σ = 1	ρ = .58	1.214	0.992
	ρ = .87	1.213	0.992
	ρ = .98	1.214	0.991
σ = 5	ρ = .58	6.062	4.957
	ρ = .87	6.063	4.956
	ρ = .98	6.071	4.958

where

$$
\begin{aligned}
n_t &= a_t + 0.75a_{t-1} \\
x_{1,t} &= 0.7x_{1,t-1} + a_{1t} \\
x_{2,t} &= 1.25x_{2,t-1} - 0.75x_{2,t-2} + a_{2t} \\
x_{3,t} &= a_{3t} + 0.8a_{3,t-1}.
\end{aligned}
$$

To further investigate the principal component estimator two new estimators were added to the one in [A]. The first estimator, introduced by Lott (1973), maximizes the adjusted R^2 for the regression of y_t on the principal components. The second estimator chooses components with significant regression coefficients in the principal component regression, see Massy (1965).

Even though the Lawless and Wang (1976) ridge estimator performed well in [A], it is interesting to test other estimators to see if further improvement is possible. In this study, two new estimators are included: Hocking, Speed and Lynn (1976), $k = \hat{\sigma}^2\Sigma\lambda_i^2\hat{\alpha}_i^2/\Sigma\lambda_i^2\hat{\alpha}_i^4$, and Hoerl, Kennard and Baldwin (1975), $k = p\hat{\sigma}^2/\Sigma\hat{\alpha}_i^2$.

The experimental design is more complicated than in [A], as it now also contains such factors as number of input variables (different models), length of time series and number of lags. The ρ values are slightly different from those in [A]. The variances and covariances have been changed to correspond to the covariance matrix used by Pukkila (1982). To reduce the computational burden, only two factors at a time are varied while the other factors are held constant at their default values.

Table 4.4 Estimated ratios of $\overline{MSE}(\hat{\nu})$ for each estimator relative to $\overline{MSE}(\hat{\nu})_{OLS}$ (RR = Ridge regression, PC = Principal component, t = transformed, LW = Lawless and Wang, HSL = Hocking, Speed and Lynn, HKB = Hoerl, Kennard and Baldwin, Σ = The r first PC, t>2 = Significant PC:s, MAX = Maximum adjusted R^2) 100 observations.

Estimator	σ = .1			σ = 1			σ = 5		
	ρ=.61	ρ=.79	ρ=.94	ρ=.61	ρ=.79	ρ=.94	ρ=.61	ρ=.79	ρ=.94
OLS	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
RR LW	1.00	1.00	1.00	0.93	0.89	0.74	0.49	0.37	0.18
RR HSL	1.00	1.00	1.00	1.00	0.99	0.98	0.92	0.88	0.73
RR HKB	1.00	1.00	0.99	0.91	0.84	0.64	0.49	0.38	0.24
PC Σ	45.55	67.93	66.54	1.40	1.56	1.04	0.68	0.51	0.16
PC t>2	0.97	1.34	2.21	1.42	1.37	0.87	0.91	0.57	0.42
PC MAX	0 96	1.24	1.73	1.32	1.26	0.91	0.91	0.58	0.47
OLSt	0.71	0.74	0.80	0.83	0.82	0.81	0.83	0.82	0.81
RRt LW	0.71	0.75	0.80	0.81	0.79	0.69	0.50	0.38	0.19
RRt HSL	0.71	0.74	0.80	0.83	0.82	0.80	0.79	0.76	0.65
RRt HKB	0.71	0.75	0.80	0.82	0.79	0.70	0.53	0.42	0.25
PCt Σ	73.18	91.10	66.25	1.92	2.04	1.16	0.75	0.61	0.29
PCt t>2	0.73	0.80	0.83	1.11	1.22	1.15	0.83	0.65	0.51
PCt MAX	0.75	0.93	0.80	0.95	1.02	0.99	0.81	0.65	0.56
$\overline{MSE}(\hat{\nu})_{OLS}$	.0002	.0004	.0010	.015	.023	.065	.360	.576	1.625

In Table 4.4, the results are given for different combinations of multicollinearity and signal–to–noise ratio corresponding to values in Table 4.1. (As the focus of interest here is multicollinearity, the columns are in different order compared to Table 4.1). We can note that the relative MSE for the ridge estimators tends to decrease as ρ increases, except when $\sigma = .1$. In these cases, the MSE is very small compared to the size of the weights. There is then very little to gain from using ridge regression and the gain decreases as ρ increases. The performances of the LW and HKB estimators are very similar and in most cases these estimators are superior to the HSL estimator. The k value of the HSL estimator is usually the smallest, which makes this estimator very close to the OLS estimator.

The principal component estimators are in some cases superior to the OLS estimators but in some cases much worse. Even though the two new estimators have a more stable performance than the estimator in [A], they never outperform the best

ridge estimators.

A more complete presentation of the results for the other factors is found in [B]. The main conclusions drawn from the simulation study are that the ridge estimators perform well, relative to *OLS* and that even though the new principal component estimators perform well relative to the old one, they may still be worse than the *OLS* estimator in some cases and thus cannot be recommended. Among the ridge estimators the *LW* and *HKB* estimators generally perform better than the *HSL* estimator, which seems to be too close to the *OLS* estimator. The proposed two–step method with the *LW* and *HKB* estimators will in the more difficult cases (strong multicollinearity, low signal–to–noise ratio and short series) reduce the *MSE* substantially compared to *OLS* as well as giving unbiased estimates of the residual variance.

4.2 Ridge estimators

Since the ridge estimators in papers [A] and [B] performed well relative to *OLS*, it is interesting to see if there are other ridge estimators that may perform even better for transfer function models. The choice of an optimal ridge estimator is not easy; there seems to be no generally superior way of determining the value of the shrinkage factor k. Almost every published simulation study seems to favor a different estimator. To some extent, this may be explained by differences in experimental design. It is therefore of interest to compare some of the more well known estimators (that have performed well in other simulation studies) for the particular case of transfer function weight estimation. In this particular case, there are usually a large number of explanatory variables that are auto– and cross correlated and we know that at least some of the estimated weights should be zero. In most simulation studies, the number of explanatory variables has been rather limited and the estimated weights have all been different from zero.

In the study in paper [C], four ordinary ridge estimators and two related estimator are compared. From paper [A] and [B], the Lawless and Wang (1976) and the Hoerl, Kennard and Baldwin (1975) estimators are chosen. The other two ordinary ridge estimators are the *RIDGM* estimator proposed by Dempster, Schatzoff and Wermuth (1977) and the "Correct Orthogonal Variance" (*COV*) or "Normalization Ridge Regression" estimator proposed by Miller and Tracy (1984) and Bulcock, Lee and Luk (1981) respectively. The *RIDGM* estimator is obtained by choosing k so that $\Sigma\, \hat{\alpha}_i^2/[\hat{\sigma}^2/k+\hat{\sigma}^2/\lambda_i] = p$. The *COV* estimator is found by choosing k to satisfy $\Sigma\, \lambda_i/(\lambda_i+k) = p$ $(k \geq 0)$. The fifth estimator is the "Orthogonalizing Ridge"

Table 4.5 Average and median *MSE* values for the estimators.

Estimator	Average MSE	Median MSE
Hoerl et al, *HKB*	.45	.34
Lawless-Wang, *LW*	.45	.31
Dempster et al, *RIDGM*	.43	.30
Bulcock/Miller et al, *COV*	.78	.37
Leskinen, *OR*	.46	.33
Krishnamurthi et al, *Equity*	1.25	.35
Optimal, *Opt*	.36	.25
Minimum, *Min*	.41	.27

(*OR*) estimator proposed by Leskinen (1980). It is not an ordinary ridge estimator since it does not shrink all α_i:s. In the study, only λ_i:s < 1 are shrunk. The shrinkage factor is a *HKB*–type k value. The last estimator is the "Equity" estimator proposed by Krishnamurthi and Rangaswamy (1987). This estimator is not a ridge estimator even though its effects are similar. It uses a square root transformation of the eigenvalues λ_i before inversion to reduce the negative effects of small eigenvalues. To see how close these estimators come to the optimal *MSE*, two optimal ridge estimators are computed. The first minimizes the expected *MSE* given knowledge of the true parameters and the second minimizes the observed *MSE*.

The simulation study in [C] has a different design compared to the studies in [A] and [B]. The design allows us to study the performance of the estimators under a wide range of possible models. The basic model is the same in all replications but the order of the operators as well as the values of the parameters varies for each replication. The basic model is

$$y_t = \frac{\omega_1(B)}{\delta_1(B)} B^{b_1} x_{1t} + \frac{\omega_2(B)}{\delta_2(B)} B^{b_2} x_{2t} + \epsilon_t$$

where

$$x_{1t} = \frac{\theta_1(B)}{\phi_1(B)} a_{1t}, \qquad x_{2t} = \frac{\theta_2(B)}{\phi_2(B)} a_{2t}$$

and

$$\epsilon_t \in \text{iid } N(0,\sigma_\epsilon^2), \quad a_{1t} \in \text{iid } N(0,\sigma_{a_1}^2), \quad a_{2t} \in \text{iid } N(0,\sigma_{a_2}^2)$$

The polynomials $\omega_i(B)$, $\delta_i(B)$, $\theta_i(B)$ and $\phi_i(B)$ are all of order 0, 1 or 2. The pure delay varies between 0 and 3 periods. For practical purposes the ν weights are truncated after lag 10 for both variables which means that a total of 22 regression coefficients are estimated. The simulation procedure is described in more detail in

Table 4.6 Number of times (in %) minimum and maximum *MSE* values for the estimators.

Estimator	Minimum MSE	Maximum MSE
Hoerl et al, *HKB*	11.8	31.1
Lawless–Wang, *LW*	20.9	2.9
Dempster et al, *RIDGM*	28.8	5.6
Bulcock/Miller et al, *COV*	11.3	20.5
Leskinen, *OR*	17.1	11.0
Krishnamurthi et al, *Equity*	10.2	29.0

[C]. A total of 5000 replications of the model are computed with estimates for the six estimators and the two optimal estimators. A number of statistics are computed for each estimator and each replication. Some of these statistics are summarized in Tables 4.5 and 4.6. The replications are also grouped according to the degree of multicollinearity, signal–to–noise ratio and orientation of the coefficient vector relative to the eigenvector corresponding to the smallest eigenvalue. These results are given in [C].

As can be seen in Table 4.5, the distribution of *MSE*:s for the Equity and *COV* estimators is rather skew with the mean *MSE* more than three times as large as the median *MSE* for the Equity estimator and approximately twice as large for the *COV* estimator. For the other estimators, the difference between mean and median values is much smaller. We also note that the *RIDGM* and *LW* estimators come closest to the *Min* and *Opt* estimators, even though the distance to the other estimators is not too large.

Table 4.6 shows the percentage of times each estimator has the minimum and maximum *MSE* among the estimators. Again the *RIDGM* and *LW* estimators show their strength, they rank best in both categories which means they are relatively safe to use, often the minimum *MSE* and very seldom the maximum *MSE*.

When the replications are grouped according to the degree of multicollinearity, signal–to–noise ratio and orientation of the coefficient vector the results from the analysis of all the above replications still hold true, with a few exceptions. When the signal–to–noise ratio is high ($\sigma < .13$) the *RIDGM* and *LW* are not superior, but as the ratio decreases (estimation becomes more difficult) they perform better than the other estimators.

This simulation shows that among the six investigated estimators the *RIDGM* and *LW* estimators overall rank as the top two estimators.

5. Empirical investigation of ridge, prewhitening and cross-spectral estimators

The purpose of paper [D] is to compare the *RIDGM* and *LW* ridge estimators to other estimators on real data. As alternative estimators the prewhitening cross–correlation and cross–spectral analysis methods are chosen. The prewhitening method is, despite its theoretical deficiencies, used by practitioners (see McLeod, 1982) and is the standard method in AUTOBOX, a well–known software package. The cross–spectral analysis method was suggested by Box and Jenkins (1976) as a method appropriate in the multi–input case.

Business cycle forecasting with leading indicators is chosen as a suitable area for building transfer function models. Here we assume that there is a constant lagged relationship between the leading indicators and a reference series. The Swedish index for industrial production (*IIP*) is chosen as the reference series, and as leading indicators the *M*1, *Yield of long term government bonds* and *Share prices at the Stockholm Exchange* are chosen. They constitute the set of financial indicators used by OECD for forecasting the Swedish business cycle. Data are collected from the Main Economic Indicators database at Statistics Sweden for the period January 1960 – December 1979 (240 monthly observations). Results from OECD (1987) and Westlund and Claesson (1989) indicate that the pure delay between the indicators and reference series may vary over time. In the period chosen the relationship between the variables appears to be rather stable. The following notation has been used:

$$
\begin{aligned}
Y_t &= \text{Swedish index for industrial production} \\
X_{1t} &= M1 \text{ (with minor adjustments) deflated by the CPI} \\
X_{2t} &= \text{Yield of long term government bonds} \\
X_{3t} &= \text{Share prices at the Stockholm Exchange}
\end{aligned}
$$

To obtain stationarity the variables are differenced once. It should be noted that both Y_t and X_{1t} are seasonal. For each series a univariate *ARIMA* model is identified, estimated and checked:

$$\nabla_{12} y_t = (1-0.597B)(1-0.372B^{24})a_t, \quad \hat{\sigma}_a = 0.03111$$

$$(1+0.145B-0.304B^2-0.297B^4)\nabla_{12} x_{1t} = (1-0.755B^{12})a_{1t}, \quad \hat{\sigma}_{a_1} = 0.00847$$

$$x_{2t} = 0.00324 + a_{2t}, \quad \hat{\sigma}_{a_2} = 0.02001$$

$$x_{3t} = a_{3t}, \quad \hat{\sigma}_{a_3} = 0.03892$$

where $y = \nabla Y$ and $x_i = \nabla X_i$.

Table 5.1 Estimated impulse response weights, $\hat{\nu}$, between prewhitened *IIP* and filtered *Money supply*, *M*1 and between prewhitened *IIP* and *Yield of long term government bonds* and between prewhitened *IIP* and filtered *Share prices* . Weights larger (absolute value) than 1.64σ are marked with $^{+}$ and larger than 1.96σ with $^{++}$.

***Money supply*, x_{1t}**								
Lags 0–7	-.118	.104	.219	**-.883^{++}**	.209	.558	-.354	**.682^{+}**
Lags 8–15	-.485	.019	.226	.194	.252	.017	.209	**-.626^{+}**
Lag 16–23	-.114	.558	-.213	.163	-.037	.036	-.307	-.283
***Yield of long term government bonds*, x_{2t}** (prewhitening with respect to seasonality)								
Lags 0–7	-.048	-.246	.303	**-.321^{+}**	.047	.144	.179	-.156
Lags 8–15	-.288	.115	.215	.084	-.093	-.183	.136	-.279
Lags 16–23	-.044	.136	.274	-.047	**-.513^{++}**	-.086	**.350^{+}**	.010
***Share prices*, x_{3t}** (prewhitening with respect to seasonality)								
Lags 0–7	-.094	-.083	.109	.046	**-.169^{+}**	.020	.094	.079
Lags 8–15	**.170^{+}**	**-.198^{++}**	-.107	**.250^{++}**	-.136	-.081	**.154^{+}**	.021
Lags 16–23	**-.195^{++}**	-.018	**.196^{++}**	-.010	**.153^{+}**	-.135	-.145	**.175^{+}**

Table 5.2. Estimated impulse response weights, $\hat{\nu}$, between *IIP* and *Money supply*, *M*1, *Yield of long term government bonds* and *Share prices* using cross spectral analysis.

***Money supply*, x_{1t}**								
Lags 0–7	.027	-.053	.403	**-.648^{+}**	.014	.524	-.020	.488
Lags 8–15	-.508	-.230	.102	.559	.521	-.389	-.275	.240
Lags 16–23	-.048	-.033	-.322	.473	.020	-.430	.129	.261
***Yield of long term government bonds*, x_{2t}**								
Lags 0–7	-.136	**-.295^{+}**	.160	.022	.165	-.226	-.048	.146
Lags 8–15	.045	-.062	.022	.036	-.110	-.201	.118	-.048
Lags 16–23	.154	-.198	.114	.136	-.152	-.046	.048	.055
***Share prices*, x_{3t}**								
Lags 0–7	-.038	.073	-.094	.004	.054	-.007	-.053	-.067
Lags 8–15	**.201^{++}**	-.012	**-.141^{+}**	.082	-.035	.118	**-.149^{+}**	.008
Lags 16–23	.033	-.047	.012	-.054	**.147^{+}**	.029	-.113	.055

In [D] three different models are built, with one, two and three input variables. Here only the three–input model will be discussed. The following estimation procedures are used:

For the prewhitening estimator, AUTOBOX is used to prewhiten each input series (one at a time) and the output series, and then to compute the estimated cross correlation function and the impulse response weights.

For the cross–spectral estimator a computer program (by the author) is used to filter the series, compute relevant spectra, then re–color the spectra and estimate the impulse response weights and their standard errors.

The ridge estimators are computed in two steps. First the impulse response weights are estimated by the ridge estimators. Then the residuals are computed and modeled using AUTOBOX. The estimated *ARMA* model for the residuals is then used to transform all variables. Then ridge estimation is used to re–estimate the impulse response weights.

Prewhitening

Both the *Yield* variable and the *Share prices* variable have been seasonally differenced to take care of the seasonality in *IIP*. The estimated weights are given in Table 5.1.

After looking for significant weights with the expected sign the following model is proposed:

$$y_t = \omega_{10}x_{1,t-7} + \omega_{20}x_{2,t-20} + \omega_{30}x_{3,t-11} + N_t$$

with $\nabla_{12}N_t = (1-\theta B)(1-\Theta B^{24})a_t$ as a starting model for the residuals.

Cross spectral

The cross–spectral estimates of the weights are given in Table 5.2. For the three–input case the *M*1 variable has no significant weight with the expected sign. Therefore the following model is proposed:

$$y_t = \omega_{20}x_{2,t-1} + \omega_{30}x_{3,t-8} + N_t$$

with $\nabla_{12}N_t = (1-\theta B)(1-\Theta B^{24})a_t$ as above.

Table 5.3. Estimated impulse response weights, $\hat{\nu}$, between *IIP* and *Money supply*, *M*1, *Yield of long term government bonds* and *Share prices* using ridge regression, *LW* and *RIDGM*. Weights larger (absolute value) than 1.64σ are marked with $^{+}$ and larger than 1.96σ with $^{++}$.

***Money supply*, x_{1t}**								
Lags 0–7	.064	.081	.044	-.064	.040	**.135**$^{++}$	.113	**.201**$^{++}$
Lags 8–15	-.002	.024	.065	.062	.064	-.009	-.001	.008
Lags 16–23	-.029	-.004	-.002	-.017	.044	.004	-.111	-.074
***Yield of long term government bonds*, x_{2t}**								
Lags 0–7	-.027	**-.059**$^{++}$	.028	.005	.035	-.001	-.006	.027
Lags 8–15	-.015	.018	.021	.016	-.022	.003	-.015	-.000
Lags 16–23	.003	.008	.040	.024	**-.049**$^{+}$	**-.051**$^{+}$	.016	-.008
***Share prices*, x_{3t}**								
Lags 0–7	.022	-.004	-.021	-.020	.001	.011	.002	.008
Lags 8–15	.002	.013	.015	.009	-.010	.018	-.009	.010
Lags 16–23	-.013	-.006	.002	-.004	**.029**$^{++}$	**.031**$^{++}$	-.010	-.000

Ridge regression

In Table 5.3 the ridge estimates for the Lawless and Wang estimator are given (The *RIDGM* estimates are similar, see [D] Table 10). Using the two–input model as a starting point and adding the significant weights for the *Share* variable, the proposed model is:

$$y_t = \omega_{10}x_{1,t-7} + (\omega_{20}-\omega_{21}B)x_{2,t-20} + (\omega_{30}-\omega_{31}B)x_{3,t-20} + N_t$$

with $\nabla_{12}N_t = (1-\theta B)(1-\Theta B^{24})a_t$.

Estimated models

Using AUTOBOX the proposed models are estimated. In the model identified using prewhitening the weight for the *Share* variable is not significant. The prewhitening model is reduced to the two–input model. In the cross–spectral model none of the weights are significant. Estimation of the ridge regression model shows that the second weights for the *Yield* and *Share* variables are insignificant. Re–estimation of the model with only one weight for each variable gives the following results:

$$y_t = .826x_{1,t-7} - .132x_{2,t-20} + .074x_{3,t-20} + N_t,$$
$$\nabla_{12}N_t = (1-.772B)(1-.339B^{12}-.605B^{24})a_{1t} \quad \hat{\sigma}_a = .02732$$

The model seems to be an adequate model for the data series. The reduction in residual variance compared to the univariate model for *IIP* is approximately 23 %. This may be an indication that the input variables have a rather predictable pattern that is included in the historical values of *IIP*, or it may be a sign of a poor model. The lead times are long enough to allow for reasonably safe predictions of the *IIP*; the forecast error is approximately 2.7 %.

Conclusions

In OECD (1987) cross–correlation functions between the indicators and the *IIP* are estimated. The time series are detrended and smoothed. Using their maximum correlation lags the following model would be estimated:

$$y'_t = \omega_{10}x'_{1,t-12} + \omega_{20}x'_{2,t-15} + \omega_{30}x'_{3,t-8} + N'_t$$

where y'_t and x'_t indicates detrended and smoothed series. The lags differ considerably from those in the estimated model above. These deviations may, at least in part, be explained by the different transformations used by the OECD (detrending and smoothing) and/or the time period used by OECD (1960–1985).

There is a large variation in the number of significant weights for the estimators (Tables 5.1, 5.2 and 5.3):

Estimator	***M1***	***Yield***	***Share***	**Total**
Prewhitening	3	3	9	15
Cross spectral	1	1	4	6
Ridge regression				
Lawless & Wang	2	3	2	7
RIDGM	3	3	2	8

From this table it appears as if the prewhitening estimator will give too many significant weights, which may be suspected as the estimator does not consider the simultaneous relationships between the input variables. The cross–spectral estimator on the other hand, did not find the significant weights for the *Yield* and *Share* variables. In the study the ridge estimators are somewhere in between and they found all significant models. Among the ridge estimators, the *LW* gives more reasonable estimates of the weights while the *RIDGM* shrinks the estimates too much compared to the other estimators.

From this limited experience, it would be recommendable to use more than one

method for identification. In this way it should be easier to separate the weights that are to be accounted for in the model from those that are spurious.

6. Concluding remarks and further research

The transfer function model can be seen as an important extension of the ordinary single equation regression model. The first step in building a transfer function model is to identify the order of the operators and the pure delay. Good preliminary estimates of the impulse response weights should be of great help in guessing the orders of the model. The prewhitening approach proposed by Box and Jenkins (1976) is primarily suited for the single–input case, whereas in practice multi–input models are of greater interest.

The proposed two–step ridge regression approach is fairly simple to use and seems to compare well to other empirical methods as the prewhitening and cross–spectral methods. The ridge approach is clearly superior to simple *OLS* estimation of the impulse response weights, especially in computationally difficult cases with strong multicollinearity and large residual variance. The ridge estimators also compare favourably to the principal component estimators investigated, even though these estimators in some cases come close to the ridge estimators in performance. Among the ridge estimators, the *RIDGM* estimator proposed by Dempster et al (1977) and the Lawless and Wang (1976) estimator have the best overall performance both in terms of average *MSE* and largest number of times with minimum *MSE* among the investigated estimators.

As the ridge estimators behaved well relative to the prewhitening and cross–spectral methods in the study in paper [D], it would be interesting to perform a simulation study covering these estimators. Such a study should cover a large number of different models and situations, in line with the simulation study in paper [C]. Even though it would probably be difficult to find an overall best method, it would be possible to better specify under which conditions the different methods should be chosen.

As some of the problems with finding reasonable transfer function models in paper [D] could be due to time–varying lag structures further research to study the behavior of the two–step method under different assumptions on time–varying lag structures (pure delay and size of the weights) would be of interest. Such knowledge would be most useful when one tries to build transfer function models for systems that may not be stable over time. Examples of such systems can be found in business cycle forecasting.

7. References

Box, G.E.P. and Jenkins, G.M. (1976). *Time series analysis: Forecasting and control.* 2:nd ed. San Francisco: Holden–Day.

Bulcock, J.W., Lee, W.F. and Luk, W.S. (1981). Ordinary least squares, ridge regression and normalization ridge regression: Some comparisons. Paper presented at the Institute of Theology, University of Uppsala, Sweden, May 1981.

Damsleth, E. (1979). Analysis of multi–input transfer function models when the inputs are correlated. Publ. no 641. Norwegian Computing Center, Oslo, Norway.

Dempster, A.P., Schatzoff, M. and Wermuth, N. (1977). "A simulation study of alternatives to ordinary least squares," *JASA, 72*, 77–91.

Edlund, P.–O. (1984). "Identification of the multi–input Box–Jenkins transfer function model," *Journal of Forecasting, 3*, 297–308.

Erickson, G.M. (1981). "Using ridge regression to estimate directly lagged effects in marketing," *JASA, 76*, 766–773.

Fask, A. and Robinson, P.B. (1977). "Identification of multivariate dynamic econometric models," *American Statistical Association* 1977 Proceedings of the Business and Economic Statistics Section, 284–289.

Haugh, L.D. and Box, G.E.P. (1977). "Identification of dynamic regression (distributed lag) models connecting two time series," *JASA, 72*, 121–130.

Hocking, R.R., Speed, F.M. and Lynn, M.J. (1976). "A class of biased estimators in linear regression," *Technometrics, 18*, 425–437.

Hoerl, A.E., Kennard, R.W. and Baldwin, K.F. (1975). "Ridge regression: some simulations," *Comm. in Statistics, 4*, 105–123.

Johnston, J. (1984). *Econometric methods.* 3:rd ed (International edition). Singapore: McGraw–Hill Book Co.

Krishnamurthi, L. and Rangaswamy, A. (1987). "The equity estimator for marketing research," *Marketing Science, 6*, 336–357.

Lawless, J.F. and Wang, P. (1976). "A simulation study of ridge and other regression estimators," *Comm. in Statistics, A5*, 307–323.

Liu, L.–M. and Hanssens, D. (1982). "Identification of multiple–input transfer function models," *Comm. in Statistics, B11*, 297–314.

Leskinen, E. (1980). On the biased estimation in regression: studies on ridge–type estimation. Jyväskylän Yliopisto, Jyväskylä, Finland.

Lott, W.F. (1973). "The optimal set of principal component restrictions on a least–squares regression," *Comm. in Statistics, 2*, 449–464.

Massy, W.F. (1965). "Principal components regression in exploratory statistical research," *JASA*, *60*, 234–256.

McLeod, G. (1982). *Box–Jenkins in practice, Volume 1*, Lancaster: G. Jenkins & Partners Ltd.

Miller, T.I. and Tracy, R.L. (1984). "Further results concerning ridge regression," *Comm. in Statistics*, *13*, 251–262.

OECD (1987). OECD leading indicators and business cycles in member countries 1960–1985. OECD Dept. of Economics and Statistics, Main Economic Indicators, Sources and Methods, no 39, January 1987.

Priestley, M.B. (1971). "Fitting relationships between time series," *Bull. Inst. Internat. Statist.*, *38*, 1–27, (paper presented at the 38:th session of the I.S.I., Washington, USA).

Pukkila, T. (1980). On transfer function noise model identification, Dept. of Mathematical Sciences, Univ. of Tampere, Finland, (paper presented to the 8:th Nordic conference on Mathematical Statistics held on Ahvenanmaa 26–29.5 1980).

Pukkila, T. (1982). "On the identification of transfer function noise models with several correlated inputs," *Scand. J. Statist.*, *9*, 139–146.

Rahiala, M. (1986). "Identification and preliminary estimation in linear transfer function models," *Scand J Statist*, *13*, 239–255.

Tsay, R.S. (1985). "Model identification in dynamic regression (distributed lag) models," *Journal of Business & Economic Statistics*, *3*, 228–237.

Westlund, A.H. and Claesson, L. (1989). On the estimation of time–varying parameters in transfer function models, (paper presented at the International Conference on "Economic Structural Change; Analysis and Forecasting", Stockholm).

Paper A

On Identification of Transfer Function Models [1]

[1] Research Report 1987:2, Department of Statistics, University of Stockholm. Thesis for the "fil.lic." degree. (Enlarged version of "Identification of the Multi-input Box–Jenkins Transfer Function Model", *Journal of Forecasting, 1984, 3*, p 297–308.)

[On Identification of Transfer Function Models]

ON IDENTIFICATION OF TRANSFER FUNCTION MODELS

by Per–Olov Edlund

SUMMARY

Different ways to identify (preliminarily estimate) the impulse response function of the Box & Jenkins transfer function model are discussed. The discussion is based on the situation when there are several input variables that are correlated with each other. It is found that most of the methods proposed are unsuitable, some are not reliable when there are correlated input variables, and some are expensive or difficult to use. Therefore an extension of a regression approach used by Pukkila (1980) is proposed. The new approach is based on the solution of some problems connected with the application of the regression method in our particular situation, namely the multicollinearity problem and the problem of autocorrelated residuals. It is found that the use of biased regression estimators on variables transformed with respect to the noise model should give better estimates than the usual ordinary regression estimator. To test the new approach a simulation experiment has been designed and performed. The results from the simulations indicate that the proposed method may be of value to the practitioner. It gives estimates with smaller mean squared error and lower estimated standard error.

This is a revised version of a paper presented at the workshop on *"Time Series Analysis in Management"* at the European Institute for Advanced Studies in Management, Brussels, November 26–27, 1981, and at the second *"International Symposium on Forecasting"* held by the International Institute of Forecasters, Istanbul, 6–9 July, 1982. An earlier version of the paper titled *"Identification of the Multi–input Box–Jenkins Transfer Function Model"* has also been published in the Journal of Forecasting, p 297–308, vol 3, 1984. The main additions here are found in Sections 5.4 and 5.5.

KEY WORDS: *Time series Transfer function model Identification procedure Biased regression Monte Carlo*

CONTENTS

1. INTRODUCTION

In economic time series analysis we frequently work with models where it is necessary to include lagged values of the independent variables to make the model realistic. If we e.g. study the relationship between advertising and sales it is reasonable to assume that this month's sales will be effected not only by this month's advertising but also by the advertising expenditures during previous months. In econometrics this possible relationship between past and present advertising and present sales could be represented by means of e.g. Almon–lags, Koyck schemes or rational lag distributions (see e.g. Johnston, 1972, Ch. 10). (These schemes are called "distributed lags").

If we do not have a relevant theory that explicitly tells us the shape of the lag structure it has to be estimated by a trial–and–error method.

Even if an acceptable lag scheme is eventually found there will probably be autocorrelated residuals which will violate the basic assumptions of the regression model.

To avoid these difficulties the transfer function models described by Box and Jenkins (1976) could be used. Their approach gives us a technique for identification of a proper model and allows us to include a model for the residual structure.

Unfortunately their preferred identification procedure was developed for the case when there is only one independent variable (or when the independent variables are mutually uncorrelated). In economic time series analysis this is unlikely to be the case and for that reason it is interesting to try to find another method for identification which will work well when there are several independent variables which are intercorrelated.

The purpose of this study is to investigate an extension of a regression method, proposed by Pukkila (1980), for preliminary estimation of the impulse response function of the transfer function model in the identification phase of building a transfer function model. This extended method will be useful when the input variables are correlated. (Note: In the following, unless otherwise stated, the term identification will be used as a synonym of preliminary estimation of the impulse response function.)

To investigate the possible benefits of the extended method a simulation experiment has been performed. The model used is from Pukkila (1980).

Outline of the paper:

In Chapter 2 the Box–Jenkins transfer function approach is presented and some methods for impulse response function identification are discussed. Chapter 3 deals with the regression method and the solution of the multicollinearity problem by biased regression, and how to adjust for autocorrelated residuals. A suggestion for an extension of Pukkila's regression method, which could be used when the input variables are correlated will be described in Chapter 4. To explore the possible benefits of the new method a simulation experiment has been performed. The outline of the experiment and the results will be reported in Chapter 5. Some concluding remarks will finally be given in Chapter 6.

2. TRANSFER FUNCTION MODELS

The use of univariate models for description and prediction of economic time series has been criticized by econometricians because the *ARMA* models could be used without understanding of the underlying economic system and that *ARMA* models could not be used to predict future values of a series when the system has been "shocked" (e.g. when an extreme value of an exogenous variable has occurred).

To overcome this criticism, Box and Jenkins transfer function models (1976, Ch. 11) may be used instead. These models resemble ordinary regression models but have the advantage of an explicit noise model which allows the residuals to be autocorrelated.

A transfer function model with one input variable, x_t, may be split into two components following Jenkins (1979),

$$y_t = u_t + n_t$$

where y_t is the dependent variable (suitably differenced/transformed to be mean and variance stationary), u_t contains that part of y_t which can be explained exactly in terms of x_t (suitably differenced/transformed to be mean and variance stationary) and n_t is an error term which represents all "missing" x variables plus the pure noise.

The relationship between x_t and u_t can be expressed by a linear dynamic relationship of the kind,

$$u_t - \delta_1 u_{t-1} - \ldots - \delta_r u_{t-r} = \omega_0 x_{t-b} - \omega_1 x_{t-b-1} - \ldots - \omega_s x_{t-b-s}$$

i.e.,

$$u_t = \frac{\omega_0 - \omega_1 B - \ldots - \omega_s B^s}{1 - \delta_1 B - \ldots - \delta_r B^r} x_{t-b} = \frac{\omega(B)}{\delta(B)} x_{t-b} = \nu(B) x_t$$

where $\nu(B) = \frac{\omega(B)}{\delta(B)} B^b$, B is the ordinary lag operator, $\omega(B)$ is a "moving average" operator, $\delta(B)$ is an "autoregressive" operator and b is a pure delay parameter which represents the number of complete time intervals before a change in x_t begins to have an effect on y_t.

The transfer function $\nu(B)$ is a rational lag structure which may represent any linear dynamic relationship between x_t and y_t to any specified degree of accuracy.

This formulation of the transfer function weights has also been used in econometrics, see e.g. Jorgenson (1966). In general differently differenced input and output variables may be used.

n_t may be replaced by an $ARMA(p,q)$ model of the form

$$n_t = c + \frac{\theta(B)}{\phi(B)} a_t$$

where $\phi(B)$ is an autoregressive operator, $\theta(B)$ a moving average operator and a_t a white noise series. If n_t is eliminated between the two expressions above a transfer function–noise model is obtained,

$$y_t = c + \frac{\omega(B)}{\delta(B)} x_{t-b} + \frac{\theta(B)}{\phi(B)} a_t \tag{2.1}$$

When there are more than one input variable, say m variables x_{1t}, x_{2t}, ..., x_{mt}, the expression above is easily generalized to

$$y_t = c + \sum_{j=1}^{m} \frac{\omega_j(B)}{\delta_j(B)} x_{j,t-b_j} + \frac{\theta(B)}{\phi(B)} a_t \tag{2.2}$$

It is also possible to allow the series to be seasonal. It is easy to show that the ordinary regression model is a special case of the more general model (2.2) above.

2.1 Identification of transfer function models

Box and Jenkins (1976, p. 378) suggest the following identification procedure:

(1) Derive rough estimates $\hat{\nu}_j$ of the impulse response weights.

(2) Use the estimates $\hat{\nu}_j$ to make guesses of the orders s and r of the right–hand and left–hand operators, $\omega(B)$ and $\delta(B)$, and of the delay parameter b.

(3) Substitute the estimates $\hat{\nu}_j$ in the equations

$$\begin{aligned}
&\nu_j = 0 && j < b \\
&\nu_j = \delta_1\nu_{j-1} + \delta_2\nu_{j-2} + \ldots + \delta_r\nu_{j-r} + \omega_0 && j = b \\
&\nu_j = \delta_1\nu_{j-1} + \delta_2\nu_{j-2} + \ldots + \delta_r\nu_{j-r} - \omega_{j-b} && j = b+1, b+2, \ldots, b+s \\
&\nu_j = \delta_1\nu_{j-1} + \delta_2\nu_{j-2} + \ldots + \delta_r\nu_{j-r} && j > b+s
\end{aligned} \tag{2.3}$$

with values of r, s and b obtained from (2) to obtain initial estimates of the parameters δ and ω in $\delta(B)$ and $\omega(B)$.

If the true ν_j values were known, b, r and s may be guessed using the following facts. The response weights consist of

(1) b zero values $\nu_0, \nu_1, \ldots, \nu_{b-1}$

(2) a further $s-r+1$ values $\nu_b, \nu_{b+1}, \ldots, \nu_{b+s-r}$ following no fixed pattern (only if $s \geq r$)

(3) values ν_j with $j \geq b+s-r+1$ which follow the pattern of an r:th order difference equation which has r starting values $\nu_{b+s}, \ldots, \nu_{b+s-r+1}$.

In Fig. 2.1 some common forms of $\delta(B)$ and $\omega(B)$ and the corresponding $\nu(B) = \omega(B)/\delta(B)$ are shown. Usually the orders of $\delta(B)$ and $\omega(B)$ are 0, 1 or 2.

When the noise model is included in the transfer function model, a combined transfer function–noise model is obtained. The noise will of course "disturb" the empirical response function in Fig. 2.1. We therefore need a "good" method for estimating $\nu(B)$, a method that will give efficient estimates of $\nu(B)$. The method should also be easy to use and not too expensive in computer time. The aim of this study is to investigate some methods for estimating the impulse response function (transfer function) that have been proposed. Of special interest is the case when there are several correlated input (x) variables in the model.

The "true" identification problem of step 2 and 3 above can be solved in different ways. That part of the identification phase will not be discussed further in this study.

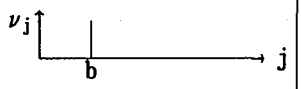
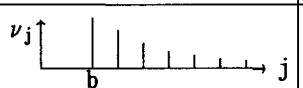
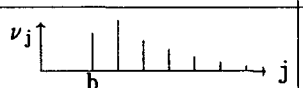
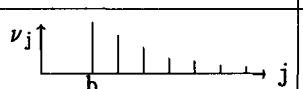
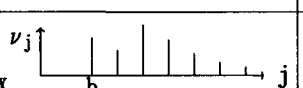
r,s,b	Transfer function	Impulse response, ν_j	ν_j j
0,0,b	$Y_t = \omega_0 B^b X_t$		0 j < b ω_0 j = b 0 j > b
0,1,b	$Y_t = (\omega_0 - \omega_1 B) B^b X_t$		0 j < b ω_0 j = b $-\omega_1$ j = b+1 0 j > b+1
0,2,b	$Y_t = (\omega_0 - \omega_1 B - \omega_2 B^2) B^b X_t$		0 j < b ω_0 j = b $-\omega_1$ j = b+1 $-\omega_2$ j = b+2 0 j > b+2
1,0,b	$(1-\delta_1 B) Y_t = \omega_0 B^b X_t$		0 j < b ω_0 j = b $\delta_1 \nu_{j-1}$ j > b
1,1,b	$(1-\delta_1 B) Y_t = (\omega_0 - \omega_1 B) B^b X_t$		0 j < b ω_0 j = b $\delta_1\omega_0 - \omega_1$ j= b+1 $\delta_1 \nu_{j-1}$ j> b+1
2,0,b	$(1-\delta_1 B - \delta_2 B^2) Y_t = \omega_0 B^b X_t$		0 j < b ω_0 j = b $\delta_1 \nu_{j-1} + \delta_2 \nu_{j-2}$ j > b
2,2,b	$(1-\delta_1 B - \delta_2 B^2) Y_t = (\omega_0 - \omega_1 B - \omega_2 B^2) B^b X_t$		0 j < b ω_0 j = b $\delta_1\omega_0 - \omega_1$ j=b+1 $(\delta_1^2+\omega_2)\omega_0 - \delta_1\omega_1$ $-\omega_2$ j = b+2 $\delta_1 \nu_{j-1} + \delta_2 \nu_{j-2}$ j > b+2

Fig. 2.1. Examples of impulse response functions from transfer functions of order (r,s,b), $(1-\delta_1 B-..-\delta_r B^r) Y_t = (\omega_0 - \omega_1 B-..-\omega_s B^s) B^b X_t$.

2.2 Methods for estimating the impulse response function

In their book, Box and Jenkins discuss three methods for estimating the transfer function weights. Two of these methods are time domain methods, the regression method (see Chapter 3) and the prewhitening–cross–correlation method (Section 2.2.1). The third method is a frequency domain method, the cross spectral analysis method (Section 2.2.3). They found that the regression method had several disadvantages and that the prewhitening–crosscorrelation method was to be preferred. However, they only discuss the case when there is only one input variable. There have also been some other methods proposed or used in practical applications.

Priestley (1971) proposed a method, the covariance contraction method (Section 2.2.2) that is similar to Box and Jenkins' prewhitening–crosscorrelation method. In practical applications the transfer function model could be identified by fitting a model that contains too many (or too few) parameters (Section 2.2.4). By trial–and–error the "right" model will eventually be found.

2.2.1 The prewhitening–crosscorrelation method

In order to study the relationship between x and y, the cross covariance or the cross correlation coefficients at different lags may be computed. The theoretical cross covariance coefficient between x and y at lag k is

$$\gamma_{xy}(k) = E((x_t-\mu_x)(y_{t+k}-\mu_y)) \qquad k = 0, \pm1, \pm2, \ldots \tag{2.4}$$

and between y and x at lag k

$$\gamma_{yx}(k) = E((y_t-\mu_y)(x_{t+k}-\mu_x)) \qquad k = 0, \pm1, \pm2, \ldots \tag{2.5}$$

In general, $\gamma_{xy}(k) \neq \gamma_{yx}(k)$. However, since $\gamma_{xy}(k) = \gamma_{yx}(-k)$ only one function $\gamma_{xy}(k)$ for $k = 0, \pm1, \pm2, \ldots$ is needed. This cross covariance function is not in general symmetric about $k = 0$. The function

$$\rho_{xy}(k) = \frac{\gamma_{xy}(k)}{\sigma_x \cdot \sigma_y} \qquad k = 0, \pm1, \pm2, \ldots \tag{2.6}$$

is called the cross correlation function. In practice these functions are estimated from

$$(2.7) \qquad c_{xy}(k) = \begin{cases} \frac{1}{n}\sum_{t=1}^{n-k}(x_t-\bar{x})(y_{t+k}-\bar{y}) & k = 0, 1, 2, \ldots, n-1 \\ \\ \frac{1}{n}\sum_{t=1}^{n+k}(y_t-\bar{y})(x_{t-k}-\bar{x}) & k = 0, -1, -2, \ldots, -n+1 \end{cases}$$

where $\bar{x}$ and $\bar{y}$ are the means of the x series and y series, respectively. The cross correlation function is then estimated by

$$(2.8) \qquad r_{xy}(k) = \frac{c_{xy}(k)}{s_x s_y} \qquad k = 0, \pm 1, \pm 2, \ldots$$

where $s_x = c_{xx}(0)^{1/2}$ and $s_y = c_{yy}(0)^{1/2}$. If x and y are ergodic processes, $r_{xy}(k)$ dies out fairly rapidly.

To obtain an estimate of ν_k we could compute the regression coefficient of y_t on x_{t-k}. This coefficient may be a poor estimate of ν_k, partly because of the autocorrelation in the x variable. To overcome this problem Box and Jenkins proposed the following method.

If the input follows a white noise process, the regression coefficient would be a fairly good estimate of ν_k. When the input follows some other process it could be transformed to white noise by a linear transformation. If the same transformation is applied to the output series both variables have been prewhitened.

It is assumed that the input process has been suitably differenced to be stationary. Then the differenced series can be represented by an $ARMA(p,q)$ model

$$(2.9) \qquad \phi_x(B)(x_t-\mu_x) = \theta_x(B)\alpha_t$$

or

$$\theta_x^{-1}(B)\phi_x(B)(x_t-\mu_x) = \alpha_t$$

The noise series α_t is then a close approximation to an uncorrelated white noise series. Applying the same transformation to the stationary series y_t gives

$$(2.10) \qquad \beta_t = \theta_x^{-1}(B)\phi_x(B)(y_t-\mu_y).$$

The transfer function–noise model may then be written as

(2.11) $\beta_t = \nu(B)\alpha_t + \epsilon_t$

where $\epsilon_t = \theta_x^{-1}(B)\phi_x(B)n_t$ is the transformed noise series. Since α_t is white noise and n_t is assumed to be independent of the input process, it is possible to obtain the coefficients ν_j from

(2.12) $$\nu_k = \frac{\gamma_{\alpha\beta}(k)}{\sigma_\alpha^2} \qquad k = 0, 1, 2, \ldots$$

where $\gamma_{\alpha\beta}(k)$ is the cross covariance at lag k between α and β. Alternatively (2.12) can be written as

(2.13) $$\nu_k = \frac{\rho_{\alpha\beta}(k)\sigma_\beta}{\sigma_\alpha} \qquad k = 0, 1, 2, \ldots$$

In practice ν_k is estimated by

(2.14) $$\hat{\nu}_k = \frac{s_\beta}{s_\alpha} \cdot r_{\alpha\beta}(k) \qquad k = 0, 1, 2, \ldots$$

When there are more than one input variable, the prewhitening technique can again be applied, if the input processes are not cross correlated, to give the estimates $\hat{\nu}_{kj}$, for $k = 1, 2, \ldots, m$ and $j = 0, 1, 2, \ldots$. If some or all input processes are cross correlated, the prewhitening technique is not directly applicable. For an example, see Damsleth (1979).

In the case of one input variable, the estimate $\hat{\nu}_k$ could be thought of as a regression coefficient of y_t on the variable x_{t-k}. If x_t is autocorrelated the x variables $x_t, x_{t-1}, \ldots, x_{t-k}, \ldots$ will be correlated. This means that we will have multicollinearity between regressors. On the other hand, if the x variable is prewhitened as described above, the regression variables $x_t, x_{t-1}, \ldots, x_{t-k}, \ldots$ will be made orthogonal to each other.

2.2.2 The covariance contraction method

This method suggested by Priestley (1971) is an approach similar to that of the prewhitening method described above. Priestley suggests that both x_t and y_t are

prewhitened by fitting univariate *ARMA* models to each process. This leads to

(2.15) $$\alpha_t = \theta_x^{-1}(B)\phi_x(B)(x_t-\mu_x)$$

and

(2.16) $$\eta_t = \theta_y^{-1}(B)\phi_y(B)(y_t-\mu_y)$$

where α_t and η_t are white noise processes. Now a transfer function model can be fitted to the residuals

(2.17) $$\eta_t + p_1\eta_{t-1} + \ldots + p_n\eta_{t-n} = q_0\alpha_t + q_1\alpha_{t-1} + \ldots + q_m\alpha_{t-m} + \epsilon_t$$

or $$P(B)\eta_t = Q(B)\alpha_t + \epsilon_t$$

The corresponding transfer function model for x_t and y_t is then given by

(2.18) $$P(B)\theta_y^{-1}(B)\phi_y(B)(y_t-\mu_y) = Q(B)\theta_x^{-1}(B)\phi_x(B)(x_t-\mu_x) + \epsilon_t$$

or $$B(B)(y_t-\mu_t) = A(B)(x_t-\mu_x) + N_t'$$

where $$B(B) = P(B)\phi_y(B)\theta_x(B)$$
$$A(B) = Q(B)\phi_x(B)\theta_y(B)$$
$$N_t' = \theta_x(B)\theta_y(B)\epsilon_t$$

The main reason for this approach is that the structures of the operators $A(B)$, $B(B)$ depend on both the autocorrelation and cross correlation structure of x_t and y_t. When individual models are fitted to x_t and y_t the autocorrelation structures are removed and, therefore, it is reasonable to assume that the form of the operators, $Q(B)$ and $P(B)$, will be much simpler than the form of $A(B)$ and $B(B)$.

Then the fact that η_t and α_t are white noise processes is used when the cross covariance function is used to indicate the forms of $P(B)$ and $Q(B)$ as follows. If

(2.19) $$\xi_t = P(B)\eta_t$$

then

$$\xi_t = q_0\alpha_t + q_1\alpha_{t-1} + \ldots + q_m\alpha_{t-m} + \epsilon_t \tag{2.20}$$

The cross covariance function between ξ_t and α_t is given by

$$\rho_{\xi\alpha}(k) = E(\alpha_t\xi_{t+k}) = \begin{cases} q_k & k = 0,\ 1,\ldots,\ m \\ 0 & \text{otherwise} \end{cases} \tag{2.21}$$

This means that the cross covariance function at lag k is simply the coefficient q_k and when $Q(B)$ contains a finite number of terms, $\rho_{\xi\alpha}(k)$ will be zero, except for lags $k =$ 0, 1, 2, ..., m. On the other hand the cross covariance function between η_t and α_t, $\rho_{\eta\alpha}(k)$, will not in general vanish after a finite number of terms since the operator $(P^{-1}(B)Q(B))$, in general, will produce an infinite series in powers of B. Luckily, there is a simple relationship between $\rho_{\eta\alpha}(k)$ and $\rho_{\xi\alpha}(k)$,

$$\rho_{\xi\alpha}(k) = P(B)\rho_{\eta\alpha}(k) \tag{2.22}$$

(where the shift operator B acts on the variable k). Therefore, $P(B)$ may be regarded as the operator which "contracts" the cross covariance function, $\rho_{\eta\alpha}(k)$, into the function $\rho_{\xi\alpha}(k)$.

In practice the estimated cross covariance function, $r_{\eta\alpha}(k)$, is used and a suitable form of $P(B)$ can be found by seeking the filter which causes the function $r_{\eta\alpha}(k)$ to decay quickly to zero.

When the form of $P(B)$ is found, the form of $Q(B)$ may be determined by inspection of the contracted cross covariance function, $r_{\xi\alpha}(k)$. (Only lags for which $r_{\xi\alpha}(k)$ differs significantly from zero are of interest.)

If there is a pure delay between the two series, some of the first q_i are zero. When $P(B)$ and $Q(B)$ have been determined, the structure of $B(B)$ and $A(B)$ can be found by polynomial multiplication.

This method does not seem to be superior to the prewhitening–cross correlation method. In fact, the identification of $P(B)$ seems to be a difficult task in practice (see Liu and Hanssens, 1982).

Haugh and Box (1977) used a similar approach. They estimated a_t and η_t as Priestley, but then they used the Box and Jenkins cross correlation method to obtain an estimate of the impulse response function. The identified model was then combined with the models for x_t and y_t to obtain the transfer function model. Fask

and Robinson (1977) generalized Priestley's approach to multivariate dynamic models.

2.2.3 The cross spectral analysis method

Box and Jenkins (1976, Appendix A11.1) also give an identification method that does not require prewhitening of the input. This method is based on spectral analysis. It could also be extended to multiple (cross correlated) inputs.

They redefine the transfer function $\nu(B)$ so that it could have non–zero impulse response weights ν_k for k a negative integer, so that

$$\nu(B) = \sum_{k=-\infty}^{\infty} \nu_k B^k \tag{2.23}$$

Then if the transfer function–noise model is

$$y_t = \nu(B)x_t + n_t \tag{2.24}$$

the theoretical autocovariance function between x_t and y_t is

$$\gamma_{xy}(k) = \sum_{j=-\infty}^{\infty} \nu_j \gamma_{xx}(k-j) \qquad k = 0, \pm 1, \pm 2, \ldots \tag{2.25}$$

Let

$$\gamma^{xy}(B) = \sum_{k=-\infty}^{\infty} \gamma_{xy}(k) B^k \tag{2.26}$$

denote the cross covariance generating function. Then, multiplying throughout in (2.25) by B^k and summing gives

$$\gamma^{xy}(B) = \nu(B)\gamma^{xx}(B) \tag{2.27}$$

where γ^{xx} is the autocovariance generating function. Substituting $B = e^{-i2\pi f}$ into (2.26) and (2.27) gives

$$\nu\left[e^{-i2\pi f}\right] = \frac{p_{xy}(f)}{p_{xx}(f)} \qquad -\frac{1}{2} \leq f < \frac{1}{2} \tag{2.28}$$

where

$$\nu\left[e^{-i2\pi f}\right] = G(f)e^{i2\pi\phi(f)} = \sum_{k=-\infty}^{\infty} \nu_k e^{-i2\pi fk} \tag{2.29}$$

$p_{xy}(f)$ is the cross spectrum between input and output. (2.29) is called the frequency response function of the system and is the Fourier transform of the impulse response function. Since $\nu(e^{-i2\pi f})$ is complex it can be written as a product involving a gain function $G(f)$ and a phase function $\phi(f)$. If $\nu(e^{-i2\pi f})$ was known, it would be possible to obtain the impulse response function ν_k from

$$\nu_k = \int_{-1/2}^{1/2} \nu\left[e^{-i2\pi f}\right] e^{i2\pi f} df \tag{2.30}$$

In practice, $\nu(e^{-i2\pi f})$ has to be estimated. The integral in (2.30) can be replaced by a finite sum. It is also possible to estimate the noise autocovariance function $\gamma_{nn}(k)$.

For multiple input transfer function models with m input variables it is possible to extend the method above. Let us assume, that after differencing the transfer function–noise model may be written as

$$y_t = \nu_1(B)x_{1,t} + \ldots + \nu_m(B)x_{m,t} + n_t \tag{2.31}$$

Multiplying throughout by $x_{1,t-k}$, $x_{2,t-k}$, ..., $x_{m,t-k}$ in turn, taking expectations and forming the generating functions, the following system of equations is obtained:

$$\begin{aligned}
\gamma^{x_1 y}(B) &= \nu_1(B)\gamma^{x_1 x_1}(B) + \ldots + \nu_m(B)\gamma^{x_1 x_m}(B) \\
&\vdots \qquad\qquad\qquad\qquad \vdots \\
\gamma^{x_m y}(B) &= \nu_1(B)\gamma^{x_m x_1}(B) + \ldots + \nu_m(B)\gamma^{x_m x_m}(B)
\end{aligned} \tag{2.32}$$

Substituting $B = e^{-i2\pi f}$ into (2.32) the spectral equations are obtained

$$p_{x_1 y}(f) = H_1(f)p_{x_1 x_1}(f) + \ldots + H_m(f)p_{x_1 x_m}(f)$$

$$\vdots \qquad\qquad \vdots \tag{2.33}$$

$$p_{x_m y}(f) = H_1(f)p_{x_m x_1}(f) + \ldots + H_m(f)p_{x_m x_m}(f)$$

where $H_j(f) = \nu_j(e^{-i2\pi f})$ can be estimated and substituted into (2.30) to give the ν_{jk}–weights. This method has been described in Pukkila (1979). He has also performed some simulations to investigate the properties of this method. From his results it seems as if this method will work well even when the input processes are cross correlated. One disadvantage with this method is the computational effort needed. For the practitioner it could also be difficult to understand the method since we are partly working in the frequency domain.

2.2.4 The under–/overfitting method

If we have a good theory it may be possible to specify a tentative model. We may then estimate the model and test the coefficients and residuals to see if the model is adequate. If not, it can be modified according to the result of the diagnostic checking.

A slightly different approach would be to overparametrize the model and then delete the parameters that are non–significant. This approach may lead to a non–parsimonious model, there can be common factors in $\delta(B)$ and $\omega(B)$ which may not be detected.

If, on the other hand, the model is underparametrized, parameters are added as they are needed. Also in this case it is possible to stop before the "best" model is reached.

If the number of input variables is large, there may be many models to be estimated and we are not even sure we got the best one.

3. THE REGRESSION METHOD

Box and Jenkins (1976, p 379) also discuss a simple regression method for identification of the impulse response function without prewhitening. They write the model (2.1) without a constant as

$$y_t = \nu_0 x_t + \nu_1 x_{t-1} + \nu_2 x_{t-2} + \ldots + n_t \tag{3.1}$$

where y_t, x_t and n_t are stationary processes with zero means. Then, multiplying throughout in (3.1) by x_{t-k} for $k = 0, 1, 2, \ldots$ gives the following equations

$$x_{t-k}y_t = \nu_0 x_{t-k}x_t + \nu_1 x_{t-k}x_{t-1} + \ldots + x_{t-k}n_t \quad k = 0,1,2,\ldots \tag{3.2}$$

Taking expectations in (3.2), on the assumption that x_{t-k} is uncorrelated with n_t for all k the following set of equations is obtained

$$\gamma_{xy}(k) = \nu_0\gamma_{xx}(k) + \nu_1\gamma_{xx}(k-1) + \ldots \qquad k = 0,1,2,\ldots \tag{3.3}$$

Assuming that $\nu_j \approx 0$ for $k > K$, then it is possible to write the first $K+1$ equations of (3.3) as

$$\boldsymbol{\gamma}_{xy} = \boldsymbol{\Gamma}_{xx}\boldsymbol{\nu} \tag{3.4}$$

where

$$\boldsymbol{\gamma}_{xy} = \begin{bmatrix} \gamma_{xy}(0) \\ \gamma_{xy}(1) \\ \vdots \\ \gamma_{xy}(K) \end{bmatrix} \quad \boldsymbol{\Gamma}_{xx} = \begin{bmatrix} \gamma_{xx}(0) & \gamma_{xx}(1) & \ldots & \gamma_{xx}(K) \\ \gamma_{xx}(1) & \gamma_{xx}(0) & \ldots & \gamma_{xx}(K-1) \\ \vdots & \vdots & & \vdots \\ \gamma_{xx}(K) & \gamma_{xx}(K-1) & \ldots & \gamma_{xx}(0) \end{bmatrix} \quad \boldsymbol{\nu} = \begin{bmatrix} \nu_0 \\ \nu_1 \\ \vdots \\ \nu_K \end{bmatrix}$$

To estimate $\boldsymbol{\nu}$, $\gamma_{xx}(k)$ is replaced by $c_{xx}(k)$ and $\gamma_{xy}(k)$ by $c_{xy}(k)$.

Box and Jenkins point out that these equations,

a) do not in general provide efficient estimates
b) are cumbersome to solve
c) in any case require knowledge of the point K beyond which ν_j is effectively zero.

If there are more than one input variable equations (3.3) and (3.4) could easily

be extended to include several input variables.

In addition to Box and Jenkins' remarks on this method it could be seen that if the input variables are autocorrelated and/or cross correlated then the covariance matrix Γ_{xx} will be multicollinear, i.e. if we view the lagged variables $x_{1,t}, \ldots, x_{1,t-K}, x_{2,t}, \ldots, x_{2,t-K}, \ldots, x_{m,t}, \ldots, x_{m,t-K}$ as different variables these "independent" variables will be correlated. Therefore the estimate $\hat{\nu}$ will have larger variance than if the inputs were white noise and not cross correlated. Even if Γ_{xx} is orthogonal there would still be a problem with autocorrelated residuals from n_t.

To avoid these problems, which may be of great importance when input variables are auto– and cross correlated, the x variables could be transformed with respect to the noise model and then some form of biased regression could be used to reduce the effects of multicollinearity. The problem of multicollinearity and its solution is discussed in the two following Sections. The problem of autocorrelated residuals is discussed in Section 3.3.

From knowledge of the underlying system it may be possible to specify which lags ν_k may be non–zero. This may decrease the order of the Γ_{xx} matrix.

Pukkila (1980) investigated this regression method and found that the estimates were surprisingly good for systems where the input variables were moderately cross correlated. Therefore, it is of interest to study how this method works when the input variables are more seriously cross correlated.

Liu and Hanssens (1982) solved the multicollinearity problem by transforming the y and x variables by a common filter. This filter was constructed to eliminate AR factors with roots close to one in the $ARMA$ processes for the x variables. To avoid the effects of autocorrelated residuals the transfer function weights were estimated with generalized least squares (GLS) rather than ordinary least squares (OLS).

Erickson (1981) used the ridge estimator to estimate the transfer function weights in a "direct lag model" for the famous Lydia Pinkham Data. He used only one input variable and did not correct for autocorrelation in the residuals.

3.1 Multicollinearity

In general, there will be correlation between the variables $x_{1,t}, x_{1,t-1}, \ldots, x_{2,t}, x_{2,t-1}, \ldots, x_{m,t}, x_{m,t-1}, \ldots$. The correlations will be of two different kinds,

a) autocorrelation between $x_{i,t}$ and $x_{i,t+k}$ $\quad i = 1, 2, \ldots, m$

b) cross correlation between $x_{i,t}$ and $x_{i+l,t+k}$ $\quad i = 1, 2, \ldots, m, \; l \neq 0$.

This means that the independent variables are not independent, i.e. there is a

multicollinearity problem. The main consequences of this are (Johnston, 172, p 160):

- The precision of estimation falls. This means that the specific estimates may have very large errors; these errors may be highly correlated, one with another, and finally, the sampling variances of the coefficients will be very large. (It is also possible that the estimated coefficients have the wrong signs.)
- We may be led to drop variables incorrectly because their coefficients are non–significant.
- Estimates of coefficients become very sensitive to particular sets of sample data, the addition of more data may produce large shifts in some of the coefficients.

Several ways to detect and test for multicollinearity in the data matrix have been discussed in literature, see e.g. Webster et al (1974) and Haitovsky (1969). Several "remedies" have also been proposed to decrease the bad effects of multicollinearity, see e.g. Intriligator (1978, Ch. 6), Silvey (1969) and Farrar and Glauber (1967). Some of them are not applicable when economic data are used, and in particular not possible to use when estimation of the impulse response function is of interest.

The author therefore proposes the use of biased regression to reduce the effects of multicollinearity.

The basic idea of biased regression is that if a small bias is introduced in the estimate it is possible to reduce the variance of the estimate considerably. Then the mean squared error, *MSE*, will be lower for the biased estimator than for the *OLS* (*O*rdinary *L*east–*S*quares) estimator. ($MSE = E\{(\hat{\nu}-\nu)^2\} = E\{((\hat{\nu}-E\hat{\nu})+(E\hat{\nu}-\nu))^2\} =$ Variance + Square of bias.)

Various biased estimators have been described in literature. Among them the principal component estimator and the ridge estimator seem to be of particular interest. Other biased regression estimators have been proposed by Marquardt (1970) (Generalized Inverse Estimator), James and Stein (1961), and Webster, Gunst and Mason (1974) (Latent Root Estimator).

3.2 Biased regression methods

In this Section two of the biased estimators that have been proposed and used to decrease the effects of multicollinearity will be briefly described.

3.2.1 The principal component estimator

If the independent variables are correlated then it is possible to describe most of the variation in the dependent variable by a subset of the independent variables. In principal component regression linear combinations of the x variables are constructed, principal components, and then these new variables are used in the regression. The principal components are chosen so that they are pairwise uncorrelated and that the first component will have the maximum possible variance, the second the maximum possible variance among those uncorrelated with the first, and so forth. The first component can be written as,

$$z_{1t} = a_{11}x_{1t} + a_{21}x_{1,t-1} + \ldots + a_{h1}x_{m,t-K} \quad t = 1, \ldots, n \tag{3.5}$$

where $h = m\cdot(K+1)$ (h is the number of independent variables, including lags of the original variables). In matrix form,

$$\mathbf{z}_1 = \mathbf{X}\mathbf{a}_1 \tag{3.6}$$

Usually $\mathbf{a}_1$ is normalized by setting $\mathbf{a}_1'\mathbf{a}_1 = 1$. This means that $\mathbf{z}_1'\mathbf{z}_1 = \lambda_1\mathbf{a}_1'\mathbf{a}_1 = \lambda_1$ where λ_1 is the largest eigenvalue of the $\mathbf{X'X}$ matrix. Continuing in this way gives, in matrix form,

$$\underset{(n\times h)}{\mathbf{Z}} = \underset{(n\times h)}{\mathbf{X}} \cdot \underset{(h\times h)}{\mathbf{A}} \tag{3.7}$$

where $\mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_h)$ is a matrix of eigenvectors and $\mathbf{Z}$ is a $n\times h$ matrix with h principal components. Transformation of the x variables by the $\mathbf{A}$ matrix gives the least squares solution of γ, the transformed coefficients, as

$$\hat{\gamma} = (\mathbf{Z'Z})^{-1}\mathbf{Z'y} \quad \text{or}$$

$$\hat{\gamma} = \Lambda^{-1}\mathbf{Z'y} \tag{3.8}$$

where Λ is a diagonal matrix of size $h\times h$ with $\lambda_1, \lambda_2, \ldots, \lambda_h$ on its diagonal. To obtain $\hat{\nu}$ the transformation is reversed and $\hat{\nu}$ is given as $\hat{\nu} = \mathbf{A}\hat{\gamma}$. If $\mathbf{X'X}$ is orthogonal all λ_i are equal to one. Even though the data matrix Λ is orthogonal the estimates in $\hat{\nu}$ are as imprecise as before. To improve on the *OLS* estimates some of the principal

components have to be deleted (meaning that the effective rank of **X'X** is reduced). This will introduce a bias, but if the data are highly collinear the reduction in the variance of the estimate $\hat{\nu}$ will be larger than the effect of the bias. There are several ways to determine the number of and which of the principal components to delete:

a) The $\hat{\gamma}$ values could be plotted for different numbers of deleted variables. This plot is called a Principal Component Trace by Vinod (1974). From this trace it may be possible to find the point where the $\hat{\gamma}$'s are stabilized and choose the corresponding number of principal components.

b) Marquardt (1970) discusses the principal component estimator and its generalization to non–integer ranks. He calls his estimator the Generalized Inverse estimator. He proposes a criterion for choosing an integer rank that will include "substantially all" of the variation in the x variables. The criterion is that the smallest value of r for which

$$\frac{\sum_{j=h-r}^{h} \lambda_j}{\sum_{j=1}^{h} \lambda_j} < \omega \tag{3.9}$$

is chosen, where $\lambda_1 > \lambda_2 > \ldots > \lambda_h$. Typically ω is selected to be 10^{-5}, or in the interval 10^{-1} to 10^{-7}.

c) Massy (1965) gives two alternative criteria for deleting components,
   (*i*) Delete the components with the smallest eigenvalues.
   (*ii*) Delete the components that are relatively unimportant as predictors of y, i.e. the components with the smallest value of $\hat{\gamma}$ in equation (3.8) are deleted.

   There is no reason why the two criteria should give the same result because y need not be highly correlated with components having large eigenvalues. Greenberg (1975) summarizes this, "including components with small eigenvectors increases variance, while including such components, if correlated with y reduces bias".

The principal component method may be very useful in our case since there are a large number of variables. If the rank of the **X'X** matrix is reduced it will be easier to solve the equation (3.8) for $\hat{\gamma}$.

3.2.2 The ridge estimator

Hoerl and Kennard (1970 a,b) introduced a biased regression method called Ridge Regression. Their estimator may be written as,

$$\hat{\boldsymbol{\nu}}_R = (\mathbf{X}'\mathbf{X}+k\mathbf{I}_h)^{-1}\mathbf{X}'\mathbf{y} \tag{3.10}$$

i.e. a small quantity $k > 0$ is added to the diagonal elements of $\mathbf{X}'\mathbf{X}$ before inversion of the matrix. They showed that the sum of *MSE*:s for individual parameters of the ridge estimator is always lower than the corresponding *MSE* for the *OLS* estimator for some $k < \sigma^2/\gamma^2_{max}$ (where γ_{max} is the largest γ from regression on principal components as described in the previous Section). The summed *MSE* for the *OLS* estimator is,

$$E(L_1^2(\hat{\boldsymbol{\nu}})) = \sigma^2 \sum_{i=1}^{h} \frac{1}{\lambda_i} \tag{3.11}$$

and for the ridge estimator

$$E(L_1^2(\hat{\boldsymbol{\nu}}_R)) = \sigma^2\Sigma \frac{\lambda_i}{(\lambda_i+k)^2} + k^2\Sigma \frac{\lambda_i^2}{(\lambda_i+k)^2} \tag{3.12}$$

where the first term on the right is the variance and the second is the squared bias. As $k \to \infty$ the variance $\to 0$ and the bias $\to \boldsymbol{\nu}'\boldsymbol{\nu}$. The effect of increasing k is to force the $\hat{\nu}_i$ towards zero. k "shrinks" the $\hat{\boldsymbol{\nu}}$ vector.

As can be seen from equation (3.11) the effect of multicollinearity is that the *MSE* is greatly increased. For an orthogonal $\mathbf{X}'\mathbf{X}$ $\lambda_1 = \ldots = \lambda_h = 1$ and $\mathrm{E}(L_1^2(\hat{\boldsymbol{\nu}})) = h\cdot\sigma^2$. When the multicollinearity is strong at least some $0 < \lambda_i << 1$. Then $\mathrm{E}(L_1^2(\hat{\boldsymbol{\nu}})) >> h\cdot\sigma^2$. It may then be assumed that at least some $\hat{\nu}_i$ are too large. This may also be seen from the expected sum of squared coefficients

$$\mathrm{E}(\hat{\boldsymbol{\nu}}'\hat{\boldsymbol{\nu}}) = \boldsymbol{\nu}'\boldsymbol{\nu} + \sigma^2\mathrm{tr}(\mathbf{X}'\mathbf{X})^{-1} = \boldsymbol{\nu}'\boldsymbol{\nu} + \sigma^2 \sum_{i=1}^{h} \frac{1}{\lambda_i} \tag{3.13}$$

which will be larger than $\boldsymbol{\nu}'\boldsymbol{\nu}$ on the average.

This estimator may then be used in our case. However, there is one problem that has to be solved, the value of k has to be determined. Since its optimal value depends

on the unknown parameters σ and γ (or ν) k has to be estimated from our data. Unfortunately there is no guarantee that $L_1^2(\hat{\nu}_R) < L_1^2(\hat{\nu})$ for our estimated k. Therefore many techniques for estimating k have been proposed and a number of simulation studies have been performed to investigate the relative merits of ridge regression v. *OLS* regression and of the methods for choosing k. (See e.g. Dempster, Schatzoff and Wermuth (1977), Wichern and Churchill (1978), Hocking, Speed and Lynn (1976), Gunst and Mason (1977), Lawless and Wang (1976), Hoerl and Kennard (1976), McDonald and Galarneau (1975)). Here only some of the methods are given:

i) Hoerl, Kennard and Baldwin (1975)

$$k = \frac{h \cdot \hat{\sigma}^2}{\Sigma\ \hat{\gamma}_i^2} \qquad \hat{\gamma}_i \text{ is the OLS estimate of } \gamma_i \tag{3.14}$$

ii) Lawless and Wang (1976)

$$k = \frac{h \cdot \hat{\sigma}^2}{\Sigma \lambda_i \hat{\gamma}_i^2} \tag{3.15}$$

iii) Bulcock, Lee and Luk (1981)

Choose k to satisfy

$$\frac{1}{h} \Sigma \frac{\lambda_i}{(\lambda_i + k)^2} = 1 \tag{3.16}$$

iv) Hocking, Speed and Lynn (1976)

$$k = \hat{\sigma}^2 \frac{\Sigma \lambda_i^2 \hat{\gamma}_i^2}{\Sigma \lambda_i^2 \hat{\gamma}_i^4} \tag{3.17}$$

v) Hoerl and Kennard (1976)

An iterative version of *i*) above

$$k_0 = \frac{h \cdot \hat{\sigma}^2}{\Sigma\ \hat{\gamma}_i^2} \tag{3.18}$$

$$k_t = \frac{h \cdot \hat{\sigma}^2}{\Sigma\ \hat{\gamma}^{*2}_{it}} \qquad \hat{\gamma}^*_{it} = \frac{\lambda_i}{(\lambda_i + k_{t-1})}\,\hat{\gamma}_i$$

$$\text{Stop when } \frac{k_{t+1} - k_t}{k_t} \leq \delta = 20(\text{tr}(\mathbf{X}'\mathbf{X})^{-1}/h)^{-1 \cdot 30}$$

vi) Hoerl and Kennard (1970 b)

The Ridge Trace: $\hat{\gamma}_R$ is computed for different values of k and plotted in a diagram with k on the x–axis and $\hat{\gamma}_R(k)$ on the y–axis. The optimal value of k is then determined by inspection. The value of k for which the ridge trace has stabilized is chosen.

In Leskinen (1980) the results from a simulation study are reported. He makes the following conclusions (p 78–9): The ridge estimator is more favorable to the OLS estimator when,

a) the number of explanatory variables increases
b) the degree of multicollinearity increases
c) the signal–to–noise ratio $\gamma'\gamma/\sigma^2$ decreases
d) the direction of the parameter vector changes from the eigenvector corresponding to the smallest eigenvalue of the $\mathbf{X}'\mathbf{X}$ matrix to the eigenvector corresponding to the largest eigenvalue of the $\mathbf{X}'\mathbf{X}$ matrix.

The ridge method described above may be called the Ordinary ridge estimator. The same value k is added to all diagonal elements of $\mathbf{X}'\mathbf{X}$. It is also possible as pointed out in Hoerl and Kennard (1970a p 63) to have a more general form of ridge regression by replacing $k \cdot \mathbf{I}_h$ by $\mathbf{K}$ where $\mathbf{K}$ is a $h \times h$ diagonal matrix with diagonal elements equal to k_i, i.e. a k value is determined for each of the explanatory variables. It is then possible to adjust the bias for each variable. This estimator is called the Generalized ridge estimator. The optimal value of k_i is $k_i = \sigma^2/\gamma_i^2$. As above σ^2 and γ_i are unknown so k_i has to be estimated. Hemmerle (1975) and Goldstein and Smith (1974) have proposed a non–iterative and an iterative method for estimating the k_i respectively.

Since some of the $\hat{\nu}_i$ may be relatively unaffected by the multicollinearity it may be wise to add a value k_i only to those diagonal elements corresponding to variables affected by the collinearity. Then the total bias will be reduced. This estimator, the Directed ridge estimator has been proposed by Guilkey and Murphy (1975).

3.3 Autocorrelated residuals

A basic assumption in regression analysis is that the residuals are uncorrelated. If this is not true the *OLS* estimates will not be efficient (minimum variance) and ordinary tests of significance can not be used. There will also be a bias in the estimation of the variance of the stochastic disturbance term. The *OLS* estimates are still defined, linear, unbiased and consistent. Since we want to test the significance of individual regression coefficients it is important that tests for autocorrelation are performed and if autocorrelation is found the estimates are corrected for this.

In regression there have been several methods proposed for dealing with autocorrelation. These methods, e.g. Durbin's method and Cochrane–Orcutt's method (see e.g. Johnston, 1972), assume that the residuals can be described by an AR model of low order (often an $AR(1)$ model). The technique is to assume and estimate the noise model and then transform the x and y variables according to the noise model. The difficulty is to identify and estimate the noise model. The two methods mentioned above are of two different kinds. Durbin's method is a two–step procedure and Cochrane–Orcutt's method is an iterative method where the iteration is on the estimation of ν and the noise model.

Box and Tiao (1975) used the transformation technique in a case study where they wanted to estimate an intervention function model. In the case of transfer function–noise models the variables can be transformed (if the true model was known) as,

$$\phi(B)y_t = \theta(B)y'_t \qquad \text{all } t$$

(3.19)

$$\phi(B)x_{i,t-j} = \theta(B)x'_{i,t-j} \qquad \begin{matrix} i = 1, \ldots, m \\ j = 0, \ldots, k \end{matrix}$$

where y'_t and $x'_{i,t-j}$ are the transformed variables. Equation (2.2) can then be written as

$$y'_t = c' + \sum_{j=1}^{m} \frac{\omega_j(B)}{\delta_j(B)} x'_{j,t-b_j} + a_t \tag{3.20}$$

This model may then be estimated by one of the biased regression methods of the previous Section.

4. A TWO-STEP PROCEDURE FOR THE IDENTIFICATION OF THE IMPULSE RESPONSE FUNCTION WHEN THE INPUT VARIABLES ARE CORRELATED.

In this Chapter a biased regression method is presented. The method can be used to identify the impulse response function when the input variables are correlated and when the noise model can be described by a (seasonal) *ARMA* model. As has been shown in the previous Chapter it is possible to deal with multicollinearity and autocorrelated residuals by biased regression and transformation respectively. Therefore in the first step the noise model is estimated and the x and y variables are transformed, in the second step the impulse response function is estimated by a biased regression estimator as e.g. the principal component estimator or the ridge estimator.

4.1 Step one: Identification, estimation and checking of the noise model and transformation of the input and output variables

To identify the noise model, Box and Jenkins (1976, p 384) suggest that the noise series n_t is estimated by

$$\hat{n}_t = y_t - \hat{\nu}(B)x_t = y_t - \hat{\nu}_0 x_t - \hat{\nu}_1 x_{t-1} - \hat{\nu}_2 x_{t-2} - \dots \tag{4.1}$$

where the $\hat{\nu}(B)$ are preliminary estimates of the impulse response function. Formula (4.1) may easily be modified to include several input variables. From the noise series $\hat{n}_t$ an *ARMA* model (or a seasonal *ARMA* model) can be identified by the standard procedure of Box and Jenkins. To estimate the impulse response function it is suggested that one of the biased regression methods described in Section 3.2 is used. Those estimators do in general give better estimates of ν and better predictions of future values than the ordinary least–squares estimator. The identified model is then estimated by e.g. Marquardt's (1963) non–linear estimation method. The results are checked and if the model is inadequate it has to be modified and re–estimated till it is found acceptable. The so obtained model

$$\hat{n}_t = \frac{\hat{\theta}(B)}{\hat{\phi}(B)}\hat{a}_t \tag{4.2}$$

is then used to transform the original variables $y_t, x_{1t}, \dots, x_{mt}$. This gives

$$\hat{\theta}(B)y_t' = \hat{\phi}(B)y_t \quad \text{all } t \tag{4.3}$$
$$\hat{\theta}(B)x_{jt}' = \hat{\phi}(B)x_{jt} \quad j = 1, \ldots, m \quad \text{all } t$$

4.2 Step two: Estimation of the impulse response function from the transformed variables y_t' and x_{jt}'

In the second step

$$y_t' = c' + \nu_{10}x_{1,t}' + \nu_{11}x_{1,t-1}' + \ldots + \nu_{mk}x_{m,t-k}' + a_t \tag{4.4}$$

is estimated by a biased regression method. In (4.4) the residuals a_t will be (almost) white noise and by using biased regression the bad effects of multicollinearity are decreased. The estimates $\hat{\nu}_{ij}$ will then hopefully be good estimates of ν_{ij} and the significance of the coefficients can be tested by the standard t–test. From the estimated values $\hat{\nu}_{ij}$ the transfer function model can be identified as described in Section 2.1.

If it is believed that the estimated noise model is inadequate, i.e the estimated residuals, $\hat{a}_t$, in (4.4) are not white noise, step one can be repeated with the estimates $\hat{\nu}_{ij}$ from (4.4) substituted into (4.1). The noise model is then re–identified and re–estimated, the variables transformed and (4.4) re–estimated. This procedure can be repeated until the estimated residuals, $\hat{a}_t$, in (4.4) are white noise.

5. EVALUATION BY SIMULATION

Since it is not possible to evaluate the proposed identification method of Chapter 4 by analytical methods the author has chosen to perform a simulation study.

5.1 The model

Of course it is an impossible task to make a complete evaluation of the proposed procedure by the use of simulation techniques. The results may be dependent on the model chosen and its parameter values. Therefore the aim is not at a complete evaluation, but merely at showing the effects for one model. The model chosen is from Pukkila (1979) and has also been used in Damsleth (1979) and Pukkila (1980). Pukkila considers two similar models, here the two input processes model has been chosen

$$y_t = (2+4B+B^2)x_{1,t} + (1-.6B)^{-1}B^3x_{2,t} + n_t \tag{5.1}$$

where
$$n_t = a_t + .75a_{t-1}$$
$$x_{1,t} = .7x_{1,t-1} + a_{1t}$$
$$x_{2,t} = 1.25x_{2,t-1} - .75x_{2,t-2} + a_{2t}$$

The processes a_{1t}, a_{2t} and a_t are normal white noise processes such that a_t is uncorrelated with a_{1t} and a_{2t}, and a_{1t} and a_{2t} are correlated with the covariance matrix Σ.

If we expand the model we obtain

$$y_t = 2x_{1,t} + 4x_{1,t-1} + x_{1,t-2} + x_{2,t-3} + .6x_{2,t-4} + .36x_{2,t-5} + \ldots + .0279936x_{2,t-10} + \ldots + n_t \tag{5.2}$$

We are interested in trying to estimate the coefficients in eq. (5.2) by the use of the regression methods presented in Chapter 3.

5.2 Criteria for comparing estimators

The most commonly used criterion is the *MSE* criterion which may be computed for each coefficient or for the whole vector of coefficients.

We have

$$(5.3) \qquad MSE(\hat{\nu}) = E(\hat{\nu}-\nu)'(\hat{\nu}-\nu) = \sum_{i=1}^{h} E(\hat{\nu}_i-\nu_i)^2$$

where $\hat{\nu}$ is the estimated vector of coefficients. The mean squared error may be thought of as the sum of two components, bias^2 + variance. In this study main focus has been on the total $MSE(\hat{\nu})$ as in (5.3). The MSE has also been computed for each coefficient.

Another possible criterion is the generalized mean squared error

$$(5.4) \qquad GMSE(\hat{\nu}) = E(\hat{\nu}-\nu)'\mathbf{G}(\hat{\nu}-\nu)$$

where G is a symmetric positive semi–definite matrix of order $h \times h$. If $\mathbf{G} = \mathbf{I}_h$ we obtain the $MSE(\hat{\nu})$ in (5.3). Usually $\mathbf{G} = \mathbf{X}'\mathbf{X}$ so that the $GMSE(\hat{\nu})$ is a measure of the predictive ability of the $\hat{\nu}_i$:s. Since the primary interest is not in prediction this criterion will not be used.

When the impulse response weights are estimated by regression methods not only the true values of the ν_i:s are of interest, but also their standard errors (for testing their significance). Therefore estimators which give low standard errors for the estimated coefficients are of interest. As a measure of the ability to identify the significant coefficients the standard error of the $\hat{\nu}_i$:s have been computed.

5.3 Selected estimators

In Chapter 3 several biased regression estimators were discussed. For this study two of the more well known estimators have been selected; the principal component estimator and the ridge estimator.

For the ridge estimator the k–value has been computed by the method of Lawless and Wang (1976), i.e.

$$(5.5) \qquad k = \frac{h \cdot \hat{\sigma}^2}{\Sigma \lambda_i \hat{\gamma}_i^2}$$

where $\hat{\gamma}_i$ is the OLS estimate of γ_i, λ_i is the i:th eigenvalue and $\hat{\sigma}^2$ is the OLS estimate of the standardized σ^2.

The principal component estimator uses the first r principal components corresponding to the r largest eigenvalues. The value of r has been determined so that the r components contain at least 99.5 % of the variance of the 22 standardized x variables. Another way of choosing the principal components is to choose the components which contributes significantly to the explanation of y. There may be some of the components with small eigenvalues that have significant $\hat{\gamma}$:s. This would have been more time consuming in terms of CPU time.

As a standard of comparison the *OLS* estimates have also been computed .

For these three estimators, estimates are computed both for the original data and for the transformed data.

5.4 The experimental design

As was noted in Section 3.2.2 the ridge estimator is more favorable to the *OLS* estimator when

a) the number of explanatory variables increases,
b) the degree of multicollinearity increases,
c) the signal–to–noise ratio $\gamma'\gamma/\sigma^2$ decreases,
d) the direction of the parameter vector γ changes from the eigenvector corresponding to the smallest eigenvalue of the $\mathbf{X'X}$ matrix to the eigenvector corresponding to the largest eigenvalue of the $\mathbf{X'X}$.

In this experiment the effects of changes in b) and c) have been studied by varying the degree of multicollinearity by using different values of the covariance between a_{1t} and a_{2t} and by varying the signal–to–noise ratio by using different values of σ^2.

The number of explanatory variables is 22, since lags up to 10 for each variable have been used. (As can be seen from eq (5.2) the value of $\nu_{2,10} \approx 0$, i.e. the cut–off is after lag 10.). More x variables could have been included by using more explanatory variables or by using a larger lag for each variable. Since the ridge estimator is known to be more superior to the *OLS* estimator when the number of variables increases, there is no need to include more variables at this preliminary stage.

It would be more interesting to experiment with different orientations of the parameter vector γ but that is left for a later study. (The number of possible combinations of levels on the multicollinearity, signal–to–noise ratio and parameter vector orientation would soon become very large.)

The degree of multicollinearity may be varied in two different ways. First, the intra correlation (autocorrelation) may be changed by changing the parameters of the

two input processes. The coefficient of $x_{1,t-1}$ could e.g. be changed to, say, .9. Then the autocorrelation for the x_{1t} series would be increased, i.e. the correlation between $x_{1,t}$ and $x_{1,t-k}$ would increase etc..

The second way is to change the inter correlation (cross correlation) between the two variables. This is done by changing the off–diagonal element of the covariance matrix Σ, i.e. the correlation between a_{1t} and a_{2t} is changed.

Since the prewhitening technique seems to be less efficient when the cross correlations are strong, it is interesting to see if the proposed method may be an improvement in such situations, and therefore only the latter way to change the multicollinearity in the X'X matrix has been used.

In the experiment the following Σ matrices were used

$$\Sigma = \begin{bmatrix} 3 & 2 \\ 2 & 4 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} 3 & 3 \\ 3 & 4 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} 3 & 3.4 \\ 3.4 & 4 \end{bmatrix}$$

which gives the following approximate correlations between a_{1t} and a_{2t}; .58, .87 and .98 respectively. (The variance for a_1 and the covariances are not the same as in Pukkila (1979). He used only one Σ matrix with the correlation coefficient .612.)

The second factor considered in this experiment is the signal–to–noise ratio. The ratio has been changed by using different values of σ^2. $\sigma^2 = .01$, 1 and 25 were used.

Fifty replications were made for each combination of Σ and σ^2.

The practical work has been carried out as follows:

(The program was written in FORTRAN–77 and run on a PRIME 750 computer. All real variables have been declared in double precision (REAL*8). NAG–routines have been used to generate random normal deviates and to compute correlations, eigenvalues and eigenvectors.)

a) 150 random normal deviates were generated for each of the two x–variables and for 50 series of a_t.

b) The x_1–series was computed from the normal deviates. Only the last $100+k$ values have been used where k is the lag.

c) For each covariance matrix
   1) the x_2–series was computed
   2) the correlation matrix was computed
   3) the eigenvalues and eigenvectors were computed
   4) the number of principal components to be used was determined so that the components correspond to 99.5 % of the variation in the x–variables.

d) For each value of the residual variance and for each replication
   1) y_t was computed

2) the coefficient vectors and the corresponding variances were estimated. In matrix form the following relations were used

OLS:

$$\hat{\gamma}_{OLS} = \Lambda^{-1}\mathbf{Z}'\mathbf{y}, \quad \hat{\nu}_{OLS} = \mathbf{T}\mathbf{A}\hat{\gamma}_{OLS} \qquad \text{where } \mathbf{T} \text{ is a diagonal matrix with the scale factors } s_y/s_{x_j} \text{ on its diagonal}$$

(5.6)

$$V(\hat{\nu}_{OLS}) = \hat{\sigma}^2_{OLS}\mathbf{T}\mathbf{A}\Lambda^{-1}\mathbf{A}'\mathbf{T}' \text{ where } \hat{\sigma}^2_{OLS} = \frac{1-\hat{\gamma}'_{OLS}\mathbf{Z}'\mathbf{y}}{100-22}$$

Ridge:

$$\hat{\gamma}_{RR} = \mathbf{K}\hat{\gamma}_{OLS}, \quad \hat{\nu}_{RR} = \mathbf{T}\mathbf{A}\hat{\gamma}_{RR} \qquad \text{where } \mathbf{K} \text{ is a diagonal matrix with } \lambda_i/(\lambda_i+k) \text{ on its diagonal}$$

(5.7)

$$V(\hat{\nu}_{RR}) = \hat{\sigma}^2_{RR}\mathbf{T}\mathbf{A}\mathbf{K}\Lambda^{-1}\mathbf{K}'\mathbf{A}'\mathbf{T}' \text{ where } \hat{\sigma}^2_{RR} = \frac{1-\hat{\gamma}'_{RR}\mathbf{Z}'\mathbf{y}-k\hat{\gamma}'_{RR}\hat{\gamma}_{RR}}{100-22}$$

Principal component:

$$\hat{\gamma}_{PC} = \Lambda_r^{-1}\mathbf{A}'_r\mathbf{X}'\mathbf{y}, \quad \hat{\nu}_{PC} = \mathbf{T}\mathbf{A}_r\hat{\gamma}_{PC} \qquad \text{where } \Lambda_r^{-1} \text{ and } \mathbf{A}_r \text{ contain the first } r \text{ (largest) eigenvalues and eigenvectors respectively}$$

(5.8)

$$V(\hat{\nu}_{PC}) = \hat{\sigma}^2_{PC}\mathbf{T}\mathbf{A}_r\Lambda_r^{-1}\mathbf{A}'_r\mathbf{T}' \text{ where } \hat{\sigma}^2_{PC} = \frac{1-\hat{\gamma}'_{PC}\mathbf{Z}'_r\mathbf{y}}{100-22}$$

e) The x variables and the y variable were transformed with the true, rather than the estimated, noise model

$$y'_t + .75y'_{t-1} = y_t \text{ or}$$

(5.9)

$$y'_t = y_t - .75y'_{t-1}$$

and

$$x'_{it} + .75x'_{it-1} = x_{it} \text{ or}$$

(5.10)

$$x'_{it} = x_{it} - .75x'_{it-1} \quad i = 1, 2$$

This was done in order to simplify the computations. Of course, in practice the noise model has to be estimated as described in Chapter 4.

f) For each estimator and for each coefficient, j, the following statistics were computed

$$(5.11) \qquad MSE(\hat{\nu})_j = \sum_{i=1}^{50} (\hat{\nu}_{ij} - \nu_j)^2/50 \qquad j = 1, \ldots, 22$$

$$(5.12) \qquad BIAS(\hat{\nu})_j = \sum_{i=1}^{50} (\hat{\nu}_{ij} - \nu_j)/50 \qquad j = 1, \ldots, 22$$

$$(5.13) \qquad S(\hat{\nu})_j = \sum_{i=1}^{50} s_{ij}/50 \qquad j = 1, \ldots, 22$$

where s_{ij} is the estimated standard error of regression coefficient j in the i:th replication.

Then the following statistics were computed

$$(5.14) \qquad \overline{MSE}(\hat{\nu}) = \sum_{j=1}^{22} MSE(\hat{\nu})_j/22$$

$$(5.15) \qquad \overline{BIAS}(\hat{\nu}) = \sum_{j=1}^{22} BIAS(\hat{\nu})_j/22$$

$$(5.16) \qquad \overline{S}(\hat{\nu}) = \sum_{j=1}^{22} S(\hat{\nu})_j/22$$

$\overline{MSE}(\hat{\nu})$ measures the average MSE over all coefficients and all replications, $\overline{BIAS}(\hat{\nu})$ measures the average bias over all coefficients and all replications, and $\overline{S}(\hat{\nu})$ measures the average standard error of the estimate for all coefficients and over all replications.

In order to compare the suggested procedure and its estimators the ratios

$$(5.17) \qquad MSE_m = \frac{\overline{MSE}(\hat{\nu})_m}{\overline{MSE}(\hat{\nu})_{OLS}} \quad \text{and} \quad S_m = \frac{\overline{S}(\hat{\nu})_m}{\overline{S}(\hat{\nu})_{OLS}}$$

where $m = RR$, PC, $OLS(t)$, $RR(t)$ and $PC(t)$ were computed (t = transformed).

By using the same time series for variable x_1 for all simulations and the same time series for variable x_2 for all replications with a given combination of values of Σ and σ^2 the variance of the estimators have been reduced. This means that the comparisons between different combinations of Σ and σ^2 will be less influenced by the fact that only 50 replications were made. By this technique we will also avoid the problem of having stochastic x variables.

5.5 Results of simulations

Before presenting the summarized results from the simulations the results for one replication where $\rho_{a_1a_2} = .87$ and $\sigma^2 = 1$ are presented.

For this replication the y variable was computed from the two x series (x_{1t} and x_{2t}) and from a residual series. The two x series remained the same for all replications with a given combination of $\rho_{a_1a_2}$ and σ^2. Different sets of residuals where used for each replication.

All series were transformed by the true residual model (see Section 5.4) to reduce the autocorrelation in the residual series and to make the estimated residual variance from OLS regression an estimate of the true variance of the residual series. The true models for the transformed variables, x'_{1t} and x'_{2t} are,

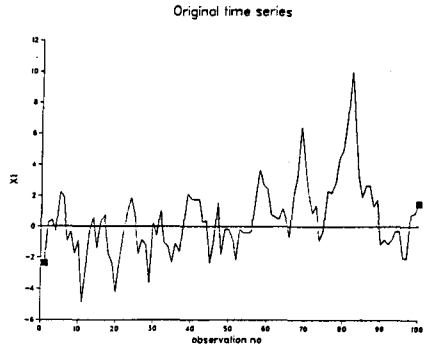
$$x'_{1t} = -0.05x'_{1t-1} + 0.525x'_{1t-2} + a'_{1t}$$

$$x'_{2t} = 0.5x'_{2t-1} + 0.1875x'_{2t-2} - 0.5625x'_{2t-3} + a'_{2t}$$

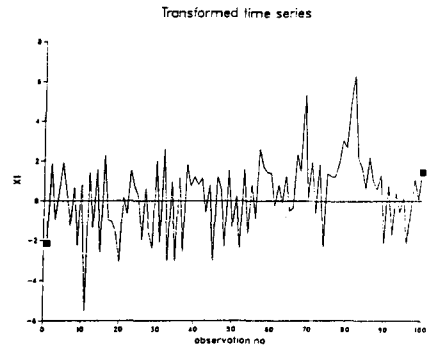
i.e. the orders of the two AR processes have increased from $AR(1)$ to $AR(2)$ and from $AR(2)$ to $AR(3)$ respectively.

The simulated series are shown in Fig. 5.1. Apart from the x_1 series the original and transformed series look very much the same. Please note that the scales are different, the transformed series have lower variances. The series are given in Appendix 1.

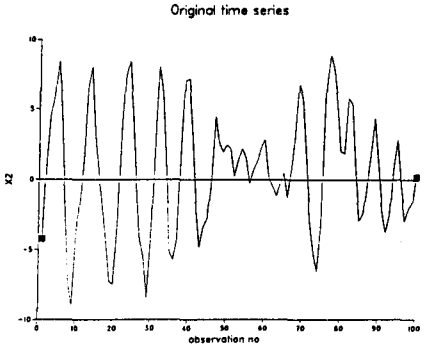
In Table 5.1 the autocorrelation and cross correlation functions are given for x_{1t},



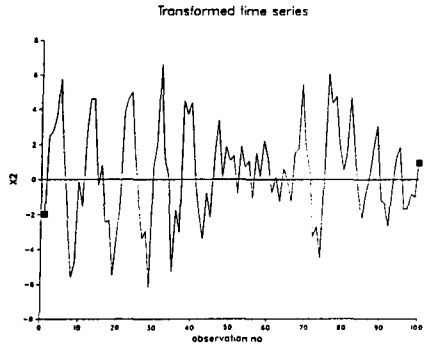
x_1 series, original



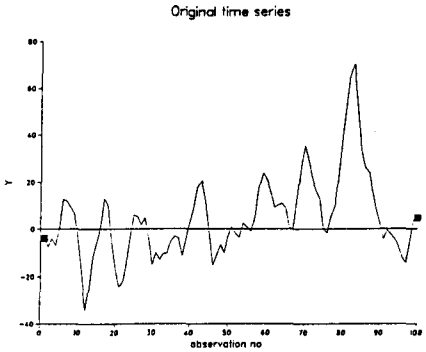
x_1 series, transformed



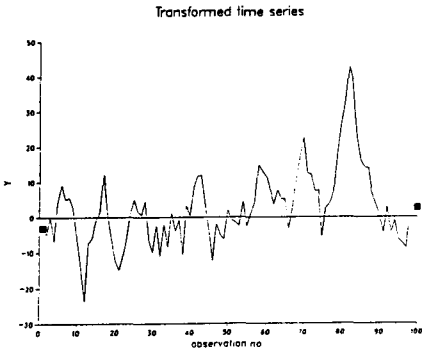
x_2 series, original



x_2 series, transformed



y series, original



y series, transformed

Fig 5.1. Simulated series, x_1, x_2 and y, original and transformed.

Table 5.1. Simulated (r) and theoretical (ρ) autocorrelation and cross correlation functions for original and transformed variables.

	Original variables				Transformed variables			
	x_{1t}		x_{2t}		x'_{1t}		x'_{2t}	
lag k	$r(k)$	$\rho(k)$	$r(k)$	$\rho(k)$	$r(k)$	$\rho(k)$	$r(k)$	$\rho(k)$
0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	0.74	0.70	0.71	0.71	–0.03	–0.11	0.52	0.51
2	0.53	0.49	0.11	0.14	0.55	0.53	0.13	0.16
3	0.36	0.34	–0.44	–0.36	–0.04	–0.08	–0.45	–0.39
4	0.23	0.24	–0.68	–0.55	0.26	0.28	–0.56	–0.45
5	0.14	0.17	–0.56	–0.42	–0.05	–0.06	–0.52	–0.39
6	0.14	0.12	–0.20	–0.11	0.12	0.15	–0.12	–0.06
7	0.18	0.08	0.20	0.17	0.08	–0.04	0.16	0.15
8	0.22	0.06	0.44	0.30	0.07	0.08	0.41	0.28
9	0.25	0.04	0.46	0.25	0.20	–0.02	0.36	0.20
10	0.22	0.03	0.29	0.08	0.01	0.04	0.25	0.07

a) Simulated (r) and theoretical (ρ) autocorrelation functions.

	Original variables		Transformed variables	
lag k	$r_{x_1x_2}(k)$	$\rho_{x_1x_2}(k)$	$r_{x'_1x'_2}(k)$	$\rho_{x'_1x'_2}(k)$
–10	0.19	0.02	0.13	0.02
–9	0.14	0.02	0.08	–0.01
–8	0.05	0.03	0.04	0.05
–7	–0.04	0.05	–0.06	–0.01
–6	–0.08	0.07	–0.02	0.09
–5	–0.05	0.10	–0.09	–0.01
–4	0.05	0.14	0.09	0.16
–3	0.18	0.20	0.04	0.00
–2	0.29	0.28	0.29	0.31
–1	0.40	0.41	0.09	0.04
0	0.52	0.58	0.55	0.59
1	0.33	0.42	0.09	0.13
2	0.01	0.09	0.09	0.15
3	–0.27	–0.20	–0.28	–0.23
4	–0.36	–0.32	–0.19	–0.16
5	–0.27	–0.25	–0.24	–0.21
6	–0.01	–0.07	0.02	0.00
7	0.27	0.10	0.16	0.05
8	0.41	0.18	0.29	0.14
9	0.34	0.15	0.23	0.08
10	0.12	0.05	0.06	0.04

b) Simulated (r) and theoretical (ρ) cross correlation functions.

x_{2t}, x'_{1t} and x'_{2t}. They correspond fairly well to their theoretical values. (The autocorrelation and cross correlation functions for other values of $\rho_{a_1 a_2}$ are given in Appendix 2 (simulated) and Appendix 4 (theoretical).) When estimating the impulse response weights these correlation functions determine the correlation matrix (the $X'X$ matrix).

This can be seen in Fig. 5.2 and 5.3 where the correlation matrices for the original and transformed variables are shown. Fig. 5.2 shows the correlation between the original 22 variables ($x_{1t}, \ldots, x_{1t-10}, x_{2t}, \ldots, x_{2t-10}$) and Fig. 5.3 the corresponding

	x_1											x_2										
t	t-1	t-2	t-3	t-4	t-5	t-6	t-7	t-8	t-9	t-10	t	t-1	t-2	t-3	t-4	t-5	t-6	t-7	t-8	t-9	t-10	
1.00	.74	.53	.36	.23	.14	.14	.18	.22	.25	.22	.52	.40	.29	.18	.05	-.05	-.08	-.04	.05	.14	.19	$x_{1,t}$
	1.00	.74	.53	.36	.23	.15	.15	.20	.23	.26	.33	.53	.41	.29	.17	.04	-.06	-.07	-.03	.06	.14	$x_{1,t-1}$
		1.00	.73	.52	.35	.23	.16	.17	.21	.25	.01	.34	.53	.41	.27	.16	.03	-.05	-.06	-.02	.06	$x_{1,t-2}$
			1.00	.74	.53	.35	.23	.15	.16	.21	-.27	.00	.34	.54	.41	.27	.16	.03	-.05	-.06	-.02	$x_{1,t-3}$
				1.00	.74	.52	.34	.20	.11	.13	-.36	-.28	-.01	.34	.54	.41	.29	.16	.02	-.07	-.07	$x_{1,t-4}$
					1.00	.73	.51	.31	.17	.09	-.27	-.37	-.29	-.02	.34	.54	.43	.29	.14	.00	-.08	$x_{1,t-5}$
						1.00	.73	.50	.30	.17	-.01	-.26	-.37	-.30	-.02	.33	.54	.43	.29	.14	.00	$x_{1,t-6}$
							1.00	.73	.51	.32	.27	.01	-.25	-.38	-.31	-.04	.32	.54	.43	.29	.14	$x_{1,t-7}$
								1.00	.75	.53	.41	.29	.03	-.25	-.40	-.33	-.06	.32	.54	.43	.28	$x_{1,t-8}$
									1.00	.76	.34	.43	.30	.01	-.29	-.42	-.34	-.04	.32	.53	.40	$x_{1,t-9}$
										1.00	.12	.36	.44	.29	-.02	-.31	-.43	-.33	-.03	.32	.51	$x_{1,t-10}$
											1.00	.71	.11	-.44	-.68	-.56	-.20	.20	.44	.46	.29	$x_{2,t}$
												1.00	.71	.10	-.45	-.69	-.57	-.19	.21	.45	.45	$x_{2,t-1}$
													1.00	.70	.08	-.46	-.69	-.56	-.18	.21	.44	$x_{2,t-2}$
														1.00	.70	.09	-.45	-.69	-.57	-.19	.21	$x_{2,t-3}$
															1.00	.70	.10	-.45	-.71	-.59	-.20	$x_{2,t-4}$
																1.00	.71	.10	-.46	-.71	-.58	$x_{2,t-5}$
																	1.00	.70	.09	-.46	-.71	$x_{2,t-6}$
																		1.00	.71	.10	-.46	$x_{2,t-7}$
																			1.00	.70	.09	$x_{2,t-8}$
																				1.00	.70	$x_{2,t-9}$
																					1.00	$x_{2,t-10}$

Fig 5.2. The correlation matrix for the original variables when $\rho_{a_1 a_2} \approx .87$

	x_1											x_2										
t	t-1	t-2	t-3	t-4	t-5	t-6	t-7	t-8	t-9	t-10	t	t-1	t-2	t-3	t-4	t-5	t-6	t-7	t-8	t-9	t-10	
1.00	-.03	.55	-.04	.26	-.05	.12	.08	.07	.20	.01	.55	.09	.29	.04	.09	-.09	-.02	-.06	.04	.08	.13	$x_{1,t}$
	1.00	-.02	.55	-.04	.27	-.05	.12	.08	.09	.20	.09	.54	.09	.29	.04	.09	-.10	-.02	-.05	.05	.09	$x_{1,t-1}$
		1.00	-.02	.54	-.04	.27	-.04	.13	.10	.10	.09	.10	.55	.10	.28	.03	.08	-.09	-.01	-.05	.05	$x_{1,t-2}$
			1.00	-.02	.55	-.05	.27	-.04	.14	.11	-.28	.09	.11	.55	.09	.27	.02	.07	-.09	-.01	-.05	$x_{1,t-3}$
				1.00	-.02	.55	-.06	.27	-.06	.13	-.19	-.28	.08	.11	.56	.10	.29	.02	.07	-.10	-.01	$x_{1,t-4}$
					1.00	-.02	.54	-.06	.24	-.06	-.24	-.19	-.29	.07	.10	.55	.11	.29	.01	.05	-.11	$x_{1,t-5}$
						1.00	-.03	.54	-.08	.23	.02	-.25	-.19	-.30	.08	.11	.56	.11	.29	.01	.05	$x_{1,t-6}$
							1.00	-.02	.54	-.06	.16	.04	-.23	-.19	-.31	.06	.09	.55	.11	.29	.00	$x_{1,t-7}$
								1.00	.00	.55	.29	.17	.05	-.23	-.20	-.32	.05	.09	.56	.12	.29	$x_{1,t-8}$
									1.00	.03	.23	.32	.19	.05	-.26	-.23	-.33	.05	.10	.55	.11	$x_{1,t-9}$
										1.00	.06	.24	.33	.20	.04	-.27	-.24	-.33	.06	.12	.55	$x_{1,t-10}$
											1.00	.52	.13	-.45	-.56	-.52	-.12	.16	.41	.36	.25	$x_{2,t}$
												1.00	.53	.13	-.45	-.56	-.53	-.12	.17	.42	.36	$x_{2,t-1}$
													1.00	.52	.11	-.47	-.57	-.52	-.11	.17	.41	$x_{2,t-2}$
														1.00	.51	.10	-.47	-.57	-.52	-.11	.17	$x_{2,t-3}$
															1.00	.52	.12	-.46	-.57	-.54	-.12	$x_{2,t-4}$
																1.00	.53	.12	-.47	-.59	-.54	$x_{2,t-5}$
																	1.00	.53	.11	-.47	-.58	$x_{2,t-6}$
																		1.00	.53	.12	-.47	$x_{2,t-7}$
																			1.00	.53	.11	$x_{2,t-8}$
																				1.00	.52	$x_{2,t-9}$
																					1.00	$x_{2,t-10}$

Fig 5.3. The correlation matrix for the transformed variables when $\rho_{a_1 a_2} \approx .87$.

Table 5.2. Some measures of multicollinearity for the two correlation matrices of Fig 5.2 and 5.3.

Original variables				Transformed variables			
i	λ_i	%	Σ%	i	λ_i	%	Σ%
1	5.964	27.1	27.1	1	4.758	21.6	21.6
2	4.935	22.4	49.5	2	3.893	17.7	39.3
3	4.796	21.8	71.3	3	2.942	13.4	52.7
4	1.997	9.1	80.4	4	2.589	11.8	64.5
5	1.109	5.0	85.5	5	1.808	8.2	72.7
6	0.875	4.0	89.4	6	1.343	6.1	78.8
7	0.462	2.1	91.5	7	0.930	4.2	83.0
8	0.364	1.7	93.2	8	0.660	3.0	86.0
9	0.301	1.4	94.6	9	0.648	2.9	89.0
10	0.233	1.1	95.6	10	0.521	2.4	91.3
11	0.186	0.8	96.5	11	0.492	2.2	93.6
12	0.173	0.8	97.2	12	0.468	2.1	95.7
13	0.164	0.7	98.0	13	0.271	1.2	96.9
14	0.143	0.7	98.6	14	0.183	0.8	97.8
15	0.129	0.6	99.2	15	0.121	0.5	98.3
16	0.091	0.4	99.6	16	0.096	0.4	98.7
17	0.034	0.2	99.8	17	0.077	0.4	99.1
18	0.012	0.1	99.9	18	0.074	0.3	99.4
19	0.009	0.0	99.9	19	0.045	0.2	99.6
20	0.009	0.0	99.9	20	0.042	0.2	99.8
21	0.007	0.0	100.0	21	0.021	0.1	99.9
22	0.007	0.0	100.0	22	0.018	0.1	100.0
	22.000	100.0			22.000	100.0	

Sum of variance inflation factors:

$\Sigma\ 1/\lambda_i = 672.40$ $\Sigma\ 1/\lambda_i = 217.86$

The spectral condition number:

$\lambda_1/\lambda_{22} = 848.52$ $\lambda_1/\lambda_{22} = 267.82$

Determinant of the correlation matrix:

$|\mathbf{X'X}| = \lambda_1 \cdot \ldots \cdot \lambda_{22} = 4.90423 \cdot 10^{-17}$ $|\mathbf{X'X}| = \lambda_1 \cdot \ldots \cdot \lambda_{22} = 3.73064 \cdot 10^{-11}$

Haitovskys test variable:

$\chi^2_H(231) = 0.000$ $\chi^2_H(231) = 0.000$

Table 5.3. Estimates of impulse response weights for all estimators and the corresponding theoretical weights.

	Untransformed variables			Transformed variables			Theoretical
	OLS	Ridge	*PC*	*OLS*	Ridge	*PC*	weights
$\hat{\nu}_{10}$	2.15	2.16	2.35	2.07	2.07	2.27	2.00
$\hat{\nu}_{11}$	3.90	3.88	3.76	3.97	3.93	3.81	4.00
$\hat{\nu}_{12}$	1.03	1.01	0.79	1.03	1.01	0.78	1.00
$\hat{\nu}_{13}$	–0.09	–0.05	0.20	–0.12	–0.08	0.21	0.00
$\hat{\nu}_{14}$	–0.12	–0.12	–0.07	–0.09	–0.06	0.00	0.00
$\hat{\nu}_{15}$	0.11	0.10	–0.02	0.15	0.13	–0.16	0.00
$\hat{\nu}_{16}$	0.24	0.22	0.09	0.17	0.16	0.23	0.00
$\hat{\nu}_{17}$	–0.03	–0.02	0.20	0.00	0.02	0.19	0.00
$\hat{\nu}_{18}$	–0.22	–0.20	–0.01	–0.08	–0.08	–0.12	0.00
$\hat{\nu}_{19}$	0.03	0.01	–0.25	–0.08	–0.09	–0.16	0.00
$\hat{\nu}_{110}$	0.08	0.08	0.12	0.05	0.05	0.06	0.00
$\hat{\nu}_{20}$	–0.07	–0.08	–0.24	–0.07	–0.07	–0.23	0.00
$\hat{\nu}_{21}$	0.01	0.02	0.17	–0.01	0.02	0.19	0.00
$\hat{\nu}_{23}$	–0.07	–0.05	0.08	–0.06	–0.06	0.04	0.00
$\hat{\nu}_{23}$	1.02	0.97	0.61	1.04	1.01	0.67	1.00
$\hat{\nu}_{24}$	0.57	0.59	0.75	0.52	0.54	0.66	0.60
$\hat{\nu}_{25}$	0.32	0.32	0.39	0.30	0.32	0.53	0.36
$\hat{\nu}_{26}$	0.05	0.05	0.07	0.12	0.11	–0.10	0.22
$\hat{\nu}_{27}$	0.25	0.23	0.04	0.18	0.16	0.11	0.13
$\hat{\nu}_{28}$	0.15	0.14	0.07	0.04	0.04	0.14	0.08
$\hat{\nu}_{29}$	–0.22	–0.20	0.06	–0.05	–0.04	–0.05	0.05
$\hat{\nu}_{210}$	0.14	0.13	–0.06	0.03	0.03	–0.01	0.03
MSE	0.015	0.013	0.032	0.006	0.005	0.032	
Bias	–0.010	–0.012	–0.016	–0.016	–0.015	–0.018	

Table 5.4. Estimates of standard errors for the estimated impulse response weights for all estimators.

$s_{\hat{\nu}_{ij}}$	Untransformed variables			Transformed variables		
	OLS	Ridge	*PC*	*OLS*	Ridge	*PC*
1,0	0.16	0.15	0.09	0.13	0.12	0.09
1	0.20	0.18	0.11	0.13	0.12	0.11
2	0.20	0.18	0.10	0.16	0.15	0.09
3	0.20	0.19	0.11	0.16	0.15	0.10
4	0.20	0.19	0.11	0.16	0.15	0.11
5	0.20	0.19	0.11	0.16	0.15	0.10
6	0.20	0.18	0.11	0.16	0.15	0.11
7	0.20	0.18	0.10	0.16	0.15	0.08
8	0.20	0.18	0.10	0.13	0.13	0.11
9	0.16	0.15	0.10	0.11	0.11	0.09
10	0.10	0.10	0.09	0.08	0.08	0.08
2,0	0.14	0.13	0.06	0.11	0.11	0.08
1	0.21	0.19	0.04	0.12	0.11	0.10
2	0.22	0.20	0.04	0.13	0.12	0.08
3	0.23	0.20	0.04	0.14	0.13	0.06
4	0.23	0.20	0.04	0.14	0.13	0.09
5	0.23	0.20	0.05	0.14	0.13	0.09
6	0.22	0.20	0.04	0.14	0.13	0.09
7	0.22	0.19	0.03	0.14	0.13	0.06
8	0.22	0.19	0.03	0.11	0.10	0.08
9	0.19	0.17	0.03	0.11	0.10	0.10
10	0.12	0.11	0.05	0.09	0.09	0.08

matrix for the transformed variables.

As can be seen in Fig. 5.2 the "upper" triangle contains the autocorrelations of variable x_{1t} and the "lower" triangle the autocorrelations of x_{2t}. The "square" shows the cross correlations between the two variables.

The effects of transforming x_{1t} and x_{2t} may be seen in Fig. 5.3 where some of the correlations drop significantly. The transformation may be viewed as a sort of differencing. The effect in this case is a reduction of the multicollinearity. For the two correlation matrices the eigenvalues and some other interesting measures of multicollinearity are shown in Table 5.2. (All correlation matrices and eigenvalues for the simulated series are given in Appendix 3. Their theoretical values are given in Appendix 5). From the results in Table 5.2 the following may be noted:

a) The sum of variance inflation factors are much larger than for an orthogonal matrix of the same order, about 29 and 8 times resp.

b) The spectral condition number is 1 if the correlation matrix is orthogonal. This

measure also shows that the X'X matrices are illconditioned, especially the first one.

c) Haitovskys test variable is effectively zero for both matrices. This means that there are ≈ 100 % multicollinearity in the matrices.

d) A relatively low number of principal components may represent most of the variation in the x variables. In order not to introduce too much bias in the principal components estimator the number of retained components were chosen so that they correspond to at least 99.5 % of the variation in the x variables.

The estimated regression coefficients for all estimators together with their theoretical values are shown in Table 5.3.

We can see that the ridge estimators are better than the corresponding *OLS* estimator in terms of *MSE* and that the principal components estimators are rather poor compared to the other estimators even though the bias is of the same order. Both the *OLS* and ridge estimators are much better on the transformed variables than the original variables. *OLS* on untransformed variables overestimated the true residual standard deviation (1.28 vs 1.00). On transformed variables the residual standard deviation was 1.01 which is very close to the true value. The principal component estimator used 16 components for the untransformed variables and 19 components for the transformed variables.

The estimated standard errors are shown in Table 5.4. We can again observe that the ridge estimator is at least as good as the OLS estimator. We can also notice that the principal components estimator now has much lower values that the other estimators.

Now the results for all simulations will be presented. The results have been summarized in tables to ease comparisons.

In Table 5.5 the average *MSE* values are presented. It may be noted that:

a) The *MSE* for the *OLS* estimator increases roughly proportional to the residual variance for a given value of ρ.

b) The ridge estimator (*RR*) used is always better than (or equal to) the corresponding *OLS* estimator. The principal component estimator (*PC*) is better than the corresponding *OLS* estimator when the signal–to–noise ratio is low ($\sigma_a = 5$). The *RR* estimator is with one exception always better than the *PC* estimator. The largest reductions in *MSE* from using *RR* are obtained when the average *MSE* is large (signal–to–noise ratio is low) and when the correlation between a_{1t} and a_{2t} is high (strong multicollinearity).

Table 5.5. Estimated ratios of $\widehat{MSE}(\hat{\nu})_m$ to $\widehat{MSE}(\hat{\nu})_{OLS}$

MSE_m	ρ = .58			ρ = .87			ρ = .98		
Estimator	σ =.1	σ = 1	σ = 5	σ =.1	σ = 1	σ = 5	σ =.1	σ = 1	σ = 5
OLS	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
RR	1.00	0.96	0.57	0.99	0.89	0.33	0.99	0.53	0.08
PC	25.21	1.21	0.82	81.04	1.49	0.29	152.48	1.71	0.09
OLS(t)	0.64	0.82	0.82	0.72	0.83	0.83	0.83	0.85	0.85
RR(t)	0.64	0.79	0.50	0.72	0.76	0.32	0.82	0.52	0.08
PC(t)	76.54	2.00	0.73	86.50	1.70	0.44	188.79	2.11	0.10
$MSE(\hat{\nu})_{OLS}$	0.0001	0.009	0.217	0.0003	0.021	0.530	0.001	0.133	3.320

c) The estimates on transformed variables are always better (or equal to) estimates on original variables for *OLS* and *RR*.

d) The *PC* estimator performs very poorly except when the signal–to–noise is low. This may be an effect of the criteria used for selecting principal components. In Table 5.6 the number of principal components used for different correlations between a_{1t} and a_{2t} are shown. In Table 5.7 the average number of deleted significant principal components are shown for different combinations of correlation and residual standard deviation.

From Table 5.7 it is quite clear that the effect of deleting components is most serious when the residual standard deviation is low. We may therefore expect that the *PC* estimator would do much better if the components that are significant are selected rather than the r components with largest eigenvalues (which is based only on the correlation matrix and not on the residual standard deviation).

Table 5.8 shows the average bias for the estimators. As may be expected, on the average the absolute value of the bias increases as the multicollinearity increases and the signal–to–noise ratio decreases. Still, the largest average bias is not more than 0.0184 which is small compared to the values of most of the coefficients. For individual coefficients the bias may be much larger than the averages given above.

Table 5.6. The number of principal components, r, used for different correlations between a_{1t} and a_{2t}.

Correlation	Data series Untransformed	Transformed
$\rho = .58$	19	21
$\rho = .87$	16	19
$\rho = .98$	13	13

Table 5.7. Average number of deleted significant principal components for different combinations of correlations between a_{1t} and a_{2t} and residual standard deviations.

Correlation between a_{1t} and a_{2t}	True standard deviation	Data series Untransformed no	%	Transformed no	%
$\rho = .58$	$\sigma = .1$	3.00	100	1.00	100
(3 and 1 *pc*	$\sigma = 1$	0.10	3	0.76	76
deleted)	$\sigma = 5$	0.02	1	0.08	8
$\rho = .87$	$\sigma = .1$	4.34	72	3.00	100
(6 and 3 *pc*	$\sigma = 1$	0.82	14	1.12	37
deleted)	$\sigma = 5$	0.20	3	0.20	7
$\rho = .98$	$\sigma = .1$	5.86	65	7.88	88
(9 *pc* dele-	$\sigma = 1$	2.34	26	2.56	28
ted)	$\sigma = 5$	0.86	10	0.76	8

Table 5.8. Estimated average bias for the estimators ($x10^{-4}$).

BIAS	ρ = .58			ρ = .87			ρ = .98		
Estimator	σ =.1	σ = 1	σ = 5	σ =.1	σ = 1	σ = 5	σ =.1	σ = 1	σ = 5
OLS	–3	8	55	4	17	75	12	38	154
RR	–3	4	–16	4	13	7	11	25	–4
PC	–5	6	53	–24	–9	57	–79	–78	–75
OLS(t)	–7	6	66	1	17	87	10	42	184
RR(t)	–7	–2	–108	1	12	–29	10	38	26
PC(t)	–31	–18	41	–19	–3	68	82	82	81

From Table 5.9 below it may be noted that:

a) The average standard error for the OLS estimator increases roughly proportional to the residual standard deviation for a given value of ρ.

b) When we look at the average standard error for the coefficients we find the same relationship between *OLS* and *RR* as before, but that the *PC* estimator nearly always is the best estimator in its class (original vs transformed data). It is interesting to notice that the reduction in $S(\hat{\nu})$ is largest when the multicollinearity is strong and when the signal–to–noise ratio is small. As may be seen from the bottom row of the Table it is almost impossible to obtain any significant coefficient from the *OLS* estimator when $\rho = .98$ and $\sigma = 5$. The average s.e. is then about 2.3 as compared with the coefficients of which some are less than 1 and the largest 4. If we then use the *RR* estimator on the transformed variables the average s.e. decreases to about .3. The *RR* is of course a much better estimator in this situation since we are interested in identifying significant coefficients.

c) The estimated s.e. from the transformed data are in general lower than the corresponding estimate from the original data when the multicollinearity is large and/or the signal–to–noise ratio is low. This is to be expected since when we have autocorrelated residuals the estimate of the variance of a_t will be biased (too large). The transformation procedure will then give a practically unbiased estimate of σ_a (and unbiased estimates of the s.e. of $\hat{\nu}_i$). This also gives us the possibility to use the standard t–test to test the significance of the estimated coefficients, see Table 5.10.

Table 5.9. Estimated ratios of $S(\hat{\nu})_m$ to $S(\hat{\nu})_{OLS}$

S_m Estimator	ρ = .58			ρ = .87			ρ = .98		
	σ =.1	σ = 1	σ = 5	σ =.1	σ = 1	σ = 5	σ =.1	σ = 1	σ = 5
OLS	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
RR	1.00	0.97	0.66	1.00	0.91	0.44	0.99	0.61	0.17
PC	1.29	0.77	0.76	0.96	0.38	0.37	0.90	0.15	0.12
OLS(t)	0.68	0.71	0.71	0.70	0.71	0.71	0.70	0.70	0.70
RR(t)	0.68	0.69	0.53	0.70	0.67	0.38	0.70	0.52	0.14
PC(t)	1.81	0.66	0.63	1.47	0.49	0.47	0.60	0.11	0.09
$\overline{S(\hat{\nu})}_{OLS}$	0.013	0.117	0.587	0.019	0.183	0.915	0.046	0.457	2.286

Table 5.10. The estimated residual standard deviation for *OLS* on untransformed and transformed data.

True standard deviation	Correlation between a_{1t} and a_{2t}	Data series Untransformed	Transformed
σ = .1	ρ = .58	0.136	0.108
	ρ = .87	0.129	0.103
	ρ = .98	0.123	0.100
σ = 1	ρ = .58	1.214	0.992
	ρ = .87	1.213	0.992
	ρ = .98	1.214	0.991
σ = 5	ρ = .58	6.062	4.957
	ρ = .87	6.063	4.956
	ρ = .98	6.071	4.958

Table 5.11 The average estimated value of k in ridge regression on untransformed and transformed data.

Correlation between a_{1t} and a_{2t}	True standard deviation	Data series Untransformed	Transformed
$\rho = .58$	$\sigma = .1$	0.00001	0.00003
	$\sigma = 1$	0.00117	0.00214
	$\sigma = 5$	0.02813	0.05055
$\rho = .87$	$\sigma = .1$	0.00001	0.00002
	$\sigma = 1$	0.00101	0.00186
	$\sigma = 5$	0.02449	0.04438
$\rho = .98$	$\sigma = .1$	0.00001	0.00002
	$\sigma = 1$	0.00093	0.00172
	$\sigma = 5$	0.02259	0.04108

In Table 5.11 the values of the shrinkage parameter k for the ridge estimator are given. The values are proportional to the residual variance of the simulated series as can be seen from eq. (5.5). Except when $\sigma = 5$ the value of k is very small.

If any recommendations are to be made from this study it would be to advice the practitioner to start by studying the eigenvalues of the correlation matrix. If these show clear signs of multicollinearity or if it is believed that the signal–to–noise ratio is low, then the ridge estimator should be used after the transformation of the original variables. If the signal–to–noise ratio is high it may still be wise to use the ridge estimator on the transformed variables, even though the gain in terms of *MSE* may not be so dramatic.

6. CONCLUDING REMARKS

In this paper different ways to estimate the impulse response function weights in the Box–Jenkins transfer function model have been discussed (Chapter 2). The discussion was based on the case when there are several input variables that are correlated with each other. It was found that most of the methods proposed are unsuitable, some are not reliable when there are correlated input variables, and some are expensive or difficult to use. Therefore an extension of a simple regression approach used by Pukkila (1980) was proposed. The new approach is based on the solution of some problems in connection with the application of the regression method in our particular situation, namely the multicollinearity problem and the problem of autocorrelated residuals. It was found that the use of biased regression estimators on variables transformed with respect to the noise model should give better estimates than the ordinary regression estimator (Chapter 4).

To test the new approach a simulation experiment was designed and performed. The results from the simulations indicate that the proposed method may be of value to the practitioner (Chapter 5).

It seems as if the greatest benefit of the proposed method is the possible reduction in the s.e. of $\hat{v}_i$. The use of biased regression decreases the risk of obtaining too large estimates of the coefficients. This reduces the risk of over–parameterization of the identified transfer function model. The method should be easy to apply since almost all computer systems have programs for biased regression.

It was noted that the two criteria used gave somewhat different results. This indicates that the results may be dependent on the criteria chosen. One other criterion that may be of interest is the proportion of times that the different estimators were better than the *OLS* estimator. In Bulcock et al (1981) some other criteria are discussed.

Since the results of the simulation study are very promising further studies will be made in the following areas:

a) Lawless and Wangs method of determining k was used. Several other methods have been proposed and some of them may be particularly suitable for this type of data. Further research is needed in this area. It is also interesting to investigate other criteria for choosing r, the number of components in the *PC* estimator. This may have changed the conclusions regarding the benefits of the *PC* estimator, especially in terms of the *MSE*.

b) A comparative simulation study including some of the other methods discussed in Chapter 2. Then different models (different orientations of the parameter vector) would be used. Other relevant aspects to study are the effects of varying the number of lags included (including too few lags), the length of the time series, effects of seasonality, other distributions for the independent variables and the residual model.

c) Some real world applications to study the practical use of the proposed methods.

When drawing conclusions from this study it should be remembered that the conclusions are based on simulations of one model (under different conditions). Even though the results are in line with what may be expected one should be very careful to say anything about the gain in *MSE* from using the proposed method. It is likely that the proposed method gives smaller or equal *MSE* than ordinary least–squares but that the gain in *MSE* may vary between models.

When we evaluate the results of the simulation study we should keep in mind that in real world applications we face problems like how many, and which, x variables to be included (the missing variables problem) and measurement errors. It is also possible that the transfer function model is not providing good fit. It may therefore be better to start with a multivariate model to make sure of the direction of causality, before choosing the transfer function model.

Finally it is noted that even if we may identify a suitable transfer function it may still be difficult to obtain meaningful estimates if the input processes are correlated. On the other hand, with the proposed method we may delete variables that are not to be included, and then reduce the effects of multicollinearity on the remaining variables. This may be the greatest advantage with the proposed method compared to *OLS*.

7. REFERENCES

Box, G.E.P., Jenkins, G.M., 1976, *Time Series Analysis: Forecasting and control*, 2:nd ed., Holden–Day, San Francisco.

Box, G.E.P., Tiao, G.C., 1975, Intervention Analysis with Applications to Economic and Environmental Problems, *JASA*, *70*, p 70–9.

Bulcock, J.W., Lee, W.F., Luk, W.S., 1981, Ordinary Least Squares, Ridge Regression and Normalization Ridge Regression: Some Comparisons, Paper presented at the Institute of Theology, University of Uppsala, May 1981.

Damsleth, E., 1979, Analysis of Multi–input Transfer Function Models when the Inputs are Correlated, Norwegian Computing Center, Oslo, publ. no 641.

Dempster, A.P., Schatzoff, M., Wermuth, N., 1977, A Simulation Study of Alternatives to Ordinary Least Squares, *JASA*, *72*, p 77–91.

Edlund, P.–O., 1984, Identification of the Multi–input Box–Jenkins Transfer Function Model, *Journal of Forecasting*, *3*, p 297–308.

Erickson, G.M., 1981, Using Ridge Regression to Estimate Directly Lagged Effects in Marketing, *JASA*, *76*, p 766–773.

Farrar, D.E., Glauber, R.R., 1967, Multicollinearity in Regression Analysis: The problem revisited, *Rev. of Economics and Statistics*, *49*, p 92–107.

Fask, A., Robinson, P.B., 1977, Identification of Multivariate Dynamic Econometric Models, *American Statistical Association, 1977* Proceedings of the Business and Economic Statistics Section, p 284–289.

Goldstein, M, Smith, A.F.M., 1974, Ridge–type Estimators for Regression Analysis, *JRSS ser.B*, *36*, p 284–91.

Greenberg, E., 1975, Minimum Variance Properties of Principal Component Regression, *JASA*, *70*, p 194–7.

Guilkey, D.K., Murphy, J.L., 1975, Directed Ridge Regression Techniques in Cases of Multicollinearity, *JASA*, *70*, p 769–75.

Gunst, R.F., Mason, R.L., 1977, Biased Estimation in Regression: An Evaluation Using Mean Squared Error, *JASA*, *72*, p 616–28.

Haitovsky, Y., 1969, Multicollinearity in Regression Analysis: Comment, *Rev. of Economics and Statistics*, *51*, p 486–9.

Haugh, L.D., Box, G.E.P., 1977, Identification of Dynamic Regression (Distributed Lag) Models Connecting Two Time Series, *JASA*, *72*, p 121–30.

Hemmerle, W.J., 1975, An Explicit Solution for Generalized Ridge Regression, *Technometrics*, *17*, p 309–14.

Hocking, R.R., Speed, F.M., Lynn, M.J., 1976, A Class of Biased Estimators in Linear Regression, *Technometrics, 18,* p 425–37.

Hoerl, A.E., Kennard, R.W., 1970a, Ridge Regression: Biased Estimation for Nonorthogonal Problems, *Technometrics, 12,* p 55–67.

Hoerl, A.E., Kennard, R.W., 1970b, Ridge Regression: Applications to Nonorthogonal Problems, *Technometrics, 12,* p 69–82.

Hoerl, A.E., Kennard, R.W., 1976, Ridge Regression: Iterative Estimation of the Biasing Parameter, *Comm. in Statistics, A5,* p 77–88.

Hoerl, A.E., Kennard, R.W., Baldwin, K.F., 1975, Ridge Regression: Some Simulations, *Comm. in Statistics, 4,* p 105–23.

Intrilligator, M.D., 1978, *Econometric Models, Techniques and Applications,* Prentice–Hall Inc, Englewood Cliffs, N.J.

James, W., Stein, G., 1961, Estimation with Quadratic Loss, Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, vol 1, ed. by Jerzy Neyman, Berkeley and Los Angeles, University of California Press, p 361–79.

Jenkins, G.M., 1979, *Practical Experiences with Modelling and Forecasting Time Series,* G Jenkins & Partners Ltd, St Helier, Jersey.

Johnston, J., 1972, *Econometric Methods,* 2:nd ed., McGraw–Hill Kogakusha Ltd, Tokyo.

Jorgenson, D.W., 1966, Rational Distributed Lag Functions, *Econometrica, 32,* p 135–48.

Lawless, J.F., Wang, P., 1976, A Simulation Study of Ridge and other Regression Estimators, *Comm. in Statistics, A5,* p 307–23.

Leskinen, E., 1980, On the Biased Estimation in Regression: Studies on Ridge–type Estimation, Jyväskylän Yliopisto, Jyväskylä, Finland.

Liu, L.–M., Hanssens, D., 1982, Identification of Multiple–input Transfer Function Models, *Comm. in Statistics, B11,* p 297–314.

Marquardt, D.W., 1963, An Algorithm for Least Squares Estimation of Nonlinear Parameters, *Jour. Soc. Ind. Appl. Math.,* 11, p 431–41.

Marquardt, D.W., 1970, Generalized Inverses, Ridge Regression, Biased Linear and Nonlinear Estimation, *Technometrics, 12,* p 591–612.

Massy, W.F., 1965, Principal Components Regression in Exploratory Statistical Research, *JASA, 60,* p 234–56.

McDonald, G.C., Galarneau, D.I., 1975, A Monte Carlo Evaluation of Some Ridge–type Estimators, *JASA, 70,* p 407–16.

Priestley, M.B., 1971, Fitting Relationships between Time Series, *Bull. Inst. Internat. Statist., 38,* p 1–27, (paper presented at the 38:th session of the I.S.I., Washington, USA).

Pukkila, T., 1979, On the Identification of Transfer Function Noise Models with Several Correlated Input Processes Using Frequency Domain Tools, Acta Universitatis Tamperensis, ser A, vol 110, Tampere, Finland.

Pukkila, T., 1980, On Transfer Function Noise Model Identification, Dept. of Mathematical Sciences, Univ. of Tampere, Finland, (paper presented to the 8:th Nordic conference on Mathematical Statistics held on Ahvenanmaa 26–29.5 1980).

Silvey, S.D., 1969, Multicollinearity and Imprecise Estimation, *JRSS ser. B, 31*, p 539–52.

Vinod, H.D., 1974, Ridge Estimation of a Trans–Log Production Function, 1974 Business and Economic Statistics Section Proceedings, ASA, p 596–601.

Webster, J.T., Gunst, R.F., Mason, R.L., 1974, Latent Root Regression Analysis, *Technometrics, 16*, p 513–22.

Wichern, D.W., Churchill, G.A., 1978, A Comparison of Ridge Estimators, *Technometrics, 20*, p 301–11.

APPENDIX 1. Original and transformed time series used in the numerical example in Section 5.5.

t	y_t	x_{1t}	x_{2t}	t	y_t	x_{1t}	x_{2t}
1	−3.994	−2.368	−4.219	51	0.798	−0.743	2.193
2	−7.524	0.262	0.985	52	−1.996	−2.200	0.214
3	−4.072	0.447	4.604	53	−3.624	−0.181	1.353
4	−7.008	−0.299	5.643	54	2.462	−0.417	2.212
5	−0.364	2.245	8.451	55	0.773	−0.413	1.633
6	12.750	1.911	5.357	56	−0.984	−0.300	−0.280
7	11.913	−0.916	−1.559	57	4.788	1.925	0.724
8	9.321	−0.274	−7.359	58	17.881	3.665	1.277
9	6.828	−1.773	−8.954	59	23.849	2.713	2.315
10	−3.764	−0.902	−3.685	60	21.033	2.452	2.886
11	−18.346	−4.919	−1.668	61	16.014	0.770	0.133
12	−34.255	−2.717	1.403	62	9.166	0.607	−0.411
13	−25.222	−0.289	6.533	63	10.419	0.476	−1.150
14	−11.614	0.561	8.055	64	10.985	1.169	−0.345
15	−5.467	−1.429	3.093	65	8.966	0.424	0.461
16	0.553	0.320	0.511	66	0.392	−0.707	−1.282
17	13.010	0.750	−1.839	67	−0.422	2.119	0.555
18	9.921	−1.775	−4.145	68	13.236	3.242	2.925
19	−5.582	−2.350	−7.267	69	25.944	6.449	6.782
20	−17.292	−4.259	−7.459	70	35.488	4.145	5.801
21	−24.318	−2.130	−3.402	71	29.385	2.054	1.735
22	−22.114	−0.538	3.181	72	21.335	0.842	−2.927
23	−14.535	1.041	7.533	73	16.387	1.373	−5.145
24	−3.344	1.879	8.459	74	13.177	−0.930	−6.530
25	6.111	0.763	4.912	75	−0.026	−0.324	−3.058
26	5.309	−1.805	−0.314	76	−1.812	2.301	6.332
27	1.722	−0.860	−4.276	77	5.659	2.179	8.914
28	4.844	−1.198	−5.478	78	9.561	2.765	8.015
29	−3.628	−3.650	−8.443	79	23.406	4.423	5.584
30	−15.159	0.214	−3.959	80	40.201	4.948	2.028
31	−9.931	−0.618	2.435	81	52.154	6.903	1.863
32	−13.006	1.005	8.062	82	66.850	10.008	5.791
33	−10.441	−1.082	6.179	83	70.441	6.850	5.350
34	−10.060	−1.313	1.045	84	51.629	3.270	0.421
35	−5.178	−2.335	−5.179	85	33.036	1.881	−2.922
36	−3.138	−1.106	−5.717	86	26.055	2.663	−2.430
37	−3.821	−1.692	−4.336	87	24.341	2.628	−0.450
38	−11.325	−0.109	2.243	88	16.865	1.293	1.871
39	−4.796	2.109	7.080	89	8.510	1.730	4.379
40	2.793	1.768	7.169	90	2.947	−1.176	1.091
41	8.682	1.718	3.377	91	−4.248	−0.852	−2.293
42	18.001	1.748	−1.823	92	−0.445	−1.150	−3.731
43	20.673	0.275	−4.825	93	−2.227	−0.887	−2.531
44	13.190	0.379	−3.329	94	−4.161	−0.282	0.867
45	−0.610	−2.394	−2.755	95	−7.061	−0.274	2.827
46	−15.254	−1.034	−0.180	96	−12.270	−2.037	−0.279
47	−11.237	1.535	4.456	97	−14.402	−2.047	−2.954
48	−6.547	−1.843	2.633	98	−5.690	0.766	−2.109
49	−10.192	−0.181	1.975	99	3.689	0.855	−1.665
50	−2.728	−0.163	2.500	100	4.793	1.471	0.172

t	y'_t	x'_{1t}	x'_{2t}	t	y'_t	x'_{1t}	x'_{2t}
1	−3.234	−2.161	−1.994	51	−0.740	0.251	1.383
2	−5.099	1.883	2.480	52	−1.441	−2.388	−0.823
3	−0.248	−0.965	2.743	53	−2.543	1.611	1.971
4	−6.822	0.425	3.585	54	4.369	−1.625	0.734
5	4.752	1.926	5.763	55	−2.504	0.806	1.083
6	9.186	0.467	1.035	56	0.894	−0.904	−1.092
7	5.024	−1.267	−2.336	57	4.117	2.603	1.542
8	5.553	0.676	−5.607	58	14.793	1.713	0.120
9	2.663	−2.281	−4.749	59	12.755	1.428	2.225
10	−5.761	0.809	−0.123	60	11.467	1.381	1.217
11	−14.025	−5.525	−1.575	61	7.414	−0.266	−0.780
12	−23.736	1.427	2.585	62	3.605	0.807	0.175
13	−7.420	−1.359	4.594	63	7.715	−0.129	−1.281
14	−6.049	1.580	4.609	64	5.199	1.266	0.616
15	−0.930	−2.614	−0.364	65	5.066	−0.525	−0.001
16	1.251	2.281	0.783	66	−3.407	−0.313	−1.281
17	12.071	−0.961	−2.426	67	2.134	2.354	1.515
18	0.868	−1.055	−2.325	68	11.636	1.477	1.788
19	−6.233	−1.559	−5.523	69	17.217	5.342	5.441
20	−12.617	−3.090	−3.317	70	22.575	0.139	1.720
21	−14.855	0.187	−0.914	71	12.453	1.950	0.444
22	−10.973	−0.678	3.866	72	11.995	−0.620	−3.260
23	−6.305	1.550	4.634	73	7.391	1.838	−2.701
24	1.385	0.717	4.983	74	7.634	−2.309	−4.504
25	5.072	0.226	1.175	75	−5.751	1.407	0.320
26	1.505	−1.974	−1.195	76	2.502	1.246	6.092
27	0.594	0.620	−3.380	77	3.783	1.245	4.345
28	4.398	−1.663	−2.943	78	6.724	1.831	4.756
29	−6.927	−2.403	−6.235	79	18.363	3.050	2.017
30	−9.963	2.016	0.717	80	26.429	2.660	0.515
31	−2.458	−2.130	1.898	81	32.332	4.907	1.477
32	−11.162	2.602	6.639	82	42.601	6.327	4.683
33	−2.070	−3.033	1.200	83	38.491	2.105	1.838
34	−8.508	0.962	0.145	84	22.761	1.691	−0.958
35	1.203	−3.057	−5.288	85	15.965	0.613	−2.204
36	−4.041	1.186	−1.751	86	14.081	2.204	−0.777
37	−0.790	−2.581	−3.022	87	13.780	0.975	0.133
38	−10.733	1.827	4.510	88	6.530	0.562	1.771
39	3.254	0.739	3.698	89	3.612	1.308	3.051
40	0.353	1.214	4.395	90	0.238	−2.157	−1.198
41	8.417	0.808	0.080	91	−4.426	0.766	−1.395
42	11.688	1.142	−1.883	92	2.875	−1.725	−2.685
43	11.907	−0.581	−3.413	93	−4.383	0.407	−0.518
44	4.260	0.815	−0.769	94	−0.874	−0.587	1.255
45	−3.805	−3.005	−2.178	95	−6.406	0.166	1.886
46	−12.400	1.220	1.453	96	−7.465	−2.161	−1.693
47	−1.937	0.620	3.366	97	−8.803	−0.426	−1.684
48	−5.094	−2.309	0.109	98	0.913	1.085	−0.846
49	−6.371	1.550	1.893	99	3.005	0.041	−1.030
50	2.050	−1.326	1.080	100	2.539	1.440	0.945

APPENDIX 2. Autocorrelation and cross correlation functions for simulated processes.

	$\rho_{a_1 a_2} \approx .58$				$\rho_{a_1 a_2} \approx .98$			
lag k	x_1 $r(k)$	x_2 $r(k)$	x_1' $r(k)$	x_2' $r(k)$	x_1 $r(k)$	x_2 $r(k)$	x_1' $r(k)$	x_2' $r(k)$
0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	0.74	0.68	–0.03	0.43	0.74	0.73	–0.03	0.57
2	0.53	0.06	0.55	0.10	0.53	0.17	0.55	0.18
3	0.36	–0.45	–0.04	–0.47	0.36	–0.39	–0.04	–0.40
4	0.23	–0.63	0.26	–0.45	0.23	–0.68	0.26	–0.58
5	0.14	–0.45	–0.05	–0.41	0.14	–0.61	–0.05	–0.56
6	0.14	–0.10	0.12	0.00	0.14	–0.27	0.12	–0.21
7	0.18	0.19	0.08	0.12	0.18	0.16	0.08	0.15
8	0.22	0.31	0.07	0.30	0.22	0.47	0.07	0.42
9	0.25	0.27	0.20	0.15	0.25	0.53	0.20	0.47
10	0.22	0.16	0.01	0.16	0.22	0.37	0.01	0.31

	$r_{x_1 x_2}$		$r_{x_1' x_2'}$		$r_{x_1 x_2}$		$r_{x_1' x_2'}$	
lag k	$r(k)$	$r(-k)$	$r(k)$	$r(-k)$	$r(k)$	$r(-k)$	$r(k)$	$r(-k)$
0	0.27	0.27	0.32	0.32	0.63	0.63	0.64	0.64
1	0.16	0.17	0.01	–0.01	0.41	0.51	0.14	0.15
2	–0.05	0.10	0.03	0.11	0.05	0.38	0.12	0.38
3	–0.21	0.08	–0.21	0.01	–0.28	0.22	–0.29	0.05
4	–0.25	0.03	–0.10	0.05	–0.40	0.06	–0.23	0.10
5	–0.17	–0.01	–0.18	–0.05	–0.30	–0.06	–0.26	–0.11
6	0.03	–0.04	0.07	0.01	–0.04	–0.09	–0.01	–0.03
7	0.25	–0.04	0.12	–0.06	0.25	–0.04	0.17	–0.05
8	0.34	0.00	0.28	0.03	0.40	0.07	0.26	0.04
9	0.23	0.05	0.12	0.00	0.36	0.18	0.28	0.13
10	0.02	0.09	0.04	0.09	0.17	0.23	0.07	0.13

APPENDIX 3. Correlation matrices and eigenvalues for simulated processes.

1.00	0.74	0.53	0.36	0.23	0.14	0.14	0.18	0.22	0.25	0.22	0.27	0.17	0.10	0.08	0.03	-0.01	-0.04	-0.04	0.00	0.05	0.09
0.74	1.00	0.74	0.53	0.36	0.23	0.15	0.15	0.20	0.23	0.26	0.16	0.28	0.18	0.10	0.08	0.03	-0.02	-0.04	-0.03	0.01	0.05
0.53	0.74	1.00	0.73	0.52	0.35	0.23	0.16	0.17	0.21	0.25	-0.05	0.16	0.29	0.18	0.10	0.07	0.02	-0.02	-0.04	-0.03	0.00
0.36	0.53	0.73	1.00	0.74	0.53	0.35	0.23	0.15	0.16	0.21	-0.21	-0.05	0.16	0.29	0.19	0.10	0.07	0.02	-0.02	-0.03	-0.03
0.23	0.36	0.52	0.74	1.00	0.74	0.52	0.34	0.20	0.11	0.13	-0.25	-0.22	-0.08	0.15	0.29	0.19	0.12	0.08	0.01	-0.03	-0.04
0.14	0.23	0.35	0.53	0.74	1.00	0.73	0.51	0.31	0.17	0.09	-0.17	-0.26	-0.25	-0.09	0.15	0.29	0.20	0.12	0.07	0.00	-0.04
0.14	0.15	0.23	0.35	0.52	0.73	1.00	0.73	0.50	0.30	0.17	0.03	-0.17	-0.26	-0.25	-0.10	0.14	0.29	0.20	0.12	0.06	-0.01
0.18	0.15	0.16	0.23	0.34	0.51	0.73	1.00	0.73	0.51	0.32	0.25	0.04	-0.16	-0.27	-0.26	-0.11	0.13	0.29	0.20	0.11	0.05
0.22	0.20	0.17	0.15	0.20	0.31	0.50	0.73	1.00	0.75	0.53	0.34	0.25	0.05	-0.17	-0.28	-0.27	-0.12	0.13	0.29	0.19	0.09
0.25	0.23	0.21	0.16	0.11	0.17	0.30	0.51	0.75	1.00	0.76	0.23	0.34	0.26	0.04	-0.18	-0.29	-0.28	-0.11	0.13	0.26	0.15
0.22	0.26	0.25	0.21	0.13	0.09	0.17	0.32	0.53	0.76	1.00	0.02	0.24	0.35	0.25	0.02	-0.20	-0.30	-0.27	-0.10	0.12	0.24
0.27	0.16	-0.05	-0.21	-0.25	-0.17	0.03	0.25	0.34	0.23	0.02	1.00	0.68	0.06	-0.45	-0.63	-0.45	-0.10	0.19	0.31	0.27	0.16
0.17	0.28	0.16	-0.05	-0.22	-0.26	-0.17	0.04	0.25	0.34	0.24	0.68	1.00	0.68	0.05	-0.47	-0.63	-0.45	-0.09	0.19	0.30	0.25
0.10	0.18	0.29	0.16	-0.08	-0.25	-0.26	-0.16	0.05	0.26	0.35	0.06	0.68	1.00	0.67	0.03	-0.48	-0.64	-0.44	-0.10	0.17	0.28
0.08	0.10	0.18	0.29	0.15	-0.09	-0.25	-0.27	-0.17	0.04	0.25	-0.45	0.05	0.67	1.00	0.67	0.03	-0.48	-0.64	-0.45	-0.11	0.16
0.01	0.08	0.10	0.19	0.29	0.15	-0.10	-0.26	-0.28	-0.18	0.02	-0.63	-0.47	0.03	0.67	1.00	0.67	0.04	-0.48	-0.65	-0.47	-0.12
-0.01	0.03	0.07	0.10	0.19	0.29	0.14	-0.11	-0.27	-0.29	-0.20	-0.45	-0.63	-0.48	0.03	0.67	1.00	0.67	0.04	-0.48	-0.65	-0.47
-0.04	-0.02	0.02	0.07	0.12	0.20	0.29	0.13	-0.12	-0.28	-0.30	-0.10	-0.45	-0.64	-0.48	0.04	0.67	1.00	0.67	0.04	-0.47	-0.64
-0.04	-0.04	-0.02	0.02	0.08	0.12	0.20	0.29	0.13	-0.11	-0.27	0.19	-0.09	-0.44	-0.64	-0.48	0.04	0.67	1.00	0.67	0.05	-0.47
0.00	-0.03	-0.04	-0.02	0.01	0.07	0.12	0.20	0.29	0.13	-0.10	0.31	0.19	-0.10	-0.45	-0.65	-0.48	0.04	0.67	1.00	0.67	0.04
0.05	0.01	-0.03	-0.03	-0.03	0.00	0.06	0.11	0.19	0.26	0.12	0.27	0.30	0.17	-0.11	-0.47	-0.65	-0.47	0.05	0.67	1.00	0.67
0.09	0.05	0.00	-0.03	-0.04	-0.04	-0.01	0.05	0.09	0.15	0.24	0.16	0.25	0.28	0.16	-0.12	-0.47	-0.64	-0.47	0.04	0.67	1.00

Correlation matrix for the original variables when $\rho_{a_1 a_2} \approx .58$.

1.00	-0.03	0.55	-0.04	0.26	-0.05	0.12	0.08	0.07	0.20	0.01	0.32	-0.01	0.11	0.01	0.05	-0.05	0.01	-0.06	0.03	0.00	0.09
-0.03	1.00	-0.02	0.55	-0.04	0.27	-0.05	0.12	0.08	0.09	0.20	0.01	0.32	0.00	0.12	0.01	0.06	-0.05	0.01	-0.06	0.04	0.00
0.55	-0.02	1.00	-0.02	0.54	-0.04	0.27	-0.04	0.13	0.10	0.10	0.03	0.02	0.32	0.00	0.11	0.01	0.05	-0.05	0.01	-0.06	0.01
-0.04	0.55	-0.02	1.00	-0.02	0.55	-0.05	0.27	-0.04	0.14	0.11	-0.21	0.03	0.03	0.33	0.00	0.11	0.00	0.04	-0.05	0.01	-0.06
0.26	-0.04	0.54	-0.02	1.00	-0.02	0.55	-0.06	0.27	-0.06	0.13	-0.10	-0.22	0.02	0.02	0.34	0.00	0.12	0.01	0.04	-0.05	0.02
-0.05	0.27	-0.04	0.55	-0.02	1.00	-0.02	0.54	-0.06	0.24	-0.06	-0.18	-0.10	-0.24	0.01	0.01	0.33	0.01	0.13	0.00	0.03	-0.06
0.12	-0.05	0.27	-0.05	0.55	-0.02	1.00	-0.03	0.54	-0.08	0.23	0.07	-0.18	-0.11	-0.24	0.01	0.01	0.33	0.01	0.13	0.00	0.03
0.08	0.12	-0.04	0.27	-0.06	0.54	-0.03	1.00	-0.02	0.54	-0.06	0.12	0.07	-0.18	-0.11	-0.25	0.00	0.00	0.33	0.02	0.12	-0.01
0.07	0.08	0.13	-0.04	0.27	-0.06	0.54	-0.02	1.00	0.00	0.55	0.28	0.12	0.08	-0.17	-0.11	-0.26	0.00	0.00	0.34	0.02	0.12
0.20	0.09	0.10	0.14	-0.06	0.24	-0.08	0.54	0.00	1.00	0.03	0.12	0.28	0.13	0.08	-0.20	-0.12	-0.27	0.00	0.01	0.31	0.00
0.01	0.20	0.10	0.11	0.13	-0.06	0.23	-0.06	0.55	0.03	1.00	0.04	0.13	0.29	0.14	0.08	-0.20	-0.14	-0.27	0.00	0.01	0.31
0.32	0.01	0.03	-0.21	-0.10	-0.18	0.07	0.12	0.28	0.12	0.04	1.00	0.43	0.10	-0.47	-0.45	-0.41	0.00	0.12	0.30	0.15	0.16
-0.01	0.32	0.02	0.03	-0.22	-0.10	-0.18	0.07	0.12	0.28	0.13	0.43	1.00	0.43	0.10	-0.47	-0.46	-0.41	0.01	0.12	0.30	0.15
0.11	0.00	0.32	0.03	0.02	-0.24	-0.11	-0.18	0.08	0.13	0.29	0.10	0.43	1.00	0.43	0.09	-0.49	-0.46	-0.41	0.01	0.11	0.29
0.01	0.12	0.00	0.33	0.02	0.01	-0.24	-0.11	-0.17	0.08	0.14	-0.47	0.10	0.43	1.00	0.41	0.07	-0.49	-0.46	-0.41	0.00	0.10
0.05	0.01	0.11	0.00	0.34	0.01	0.01	-0.25	-0.11	-0.20	0.08	-0.45	-0.47	0.09	0.41	1.00	0.41	0.09	-0.49	-0.47	-0.42	-0.01
-0.05	0.06	0.01	0.11	0.00	0.33	0.01	0.00	-0.26	-0.12	-0.20	-0.41	-0.46	-0.49	0.07	0.41	1.00	0.42	0.10	-0.49	-0.48	-0.43
0.01	-0.05	0.05	0.00	0.12	0.01	0.33	0.00	0.00	-0.27	-0.14	0.00	-0.41	-0.46	-0.49	0.09	0.42	1.00	0.42	0.09	-0.48	-0.47
-0.06	0.01	-0.05	0.04	0.01	0.13	0.01	0.33	0.00	0.00	-0.27	0.12	0.01	-0.41	-0.46	-0.49	0.10	0.42	1.00	0.42	0.10	-0.48
0.03	-0.06	0.01	-0.05	0.04	0.00	0.13	0.02	0.34	0.01	0.00	0.30	0.12	0.01	-0.41	-0.47	-0.49	0.09	0.42	1.00	0.42	0.09
0.00	0.04	-0.06	0.01	-0.05	0.03	0.00	0.12	0.02	0.31	0.01	0.15	0.30	0.11	0.00	-0.42	-0.48	-0.48	0.10	0.42	1.00	0.41
0.09	0.00	0.03	-0.06	0.02	-0.06	0.03	-0.01	0.12	0.00	0.31	0.16	0.15	0.29	0.10	-0.01	-0.43	-0.47	-0.48	0.09	0.41	1.00

Correlation matrix for the transformed variables when $\rho_{a_1 a_2} \approx .58$.

Some measures of multicollinearity for the two simulated correlation matrices with $\rho_{a_1 a_2} \approx .58$.

	Original variables				Transformed variables		
i	λ_i	%	Σ%	i	λ_i	%	Σ%
1	4.983	22.6	22.6	1	3.783	17.2	17.2
2	4.506	20.5	43.1	2	3.220	14.6	31.8
3	4.037	18.3	61.5	3	2.825	12.8	44.7
4	1.973	9.0	70.5	4	2.365	10.7	55.4
5	1.673	7.6	78.1	5	1.652	7.5	62.9
6	1.106	5.0	83.1	6	1.336	6.1	69.0
7	0.855	3.9	87.0	7	1.118	5.1	74.1
8	0.505	2.3	89.3	8	0.920	4.2	78.3
9	0.394	1.8	91.1	9	0.721	3.3	81.5
10	0.391	1.8	92.8	10	0.597	2.7	84.3
11	0.319	1.5	94.3	11	0.529	2.4	86.7
12	0.277	1.3	95.5	12	0.508	2.3	89.0
13	0.229	1.0	96.6	13	0.491	2.2	91.2
14	0.176	0.8	97.4	14	0.472	2.1	93.3
15	0.166	0.8	98.1	15	0.306	1.4	94.7
16	0.157	0.7	98.9	16	0.293	1.3	96.1
17	0.101	0.5	99.3	17	0.249	1.1	97.2
18	0.039	0.2	99.5	18	0.223	1.0	98.2
19	0.031	0.1	99.6	19	0.143	0.7	98.9
20	0.031	0.1	99.8	20	0.129	0.6	99.5
21	0.025	0.1	99.9	21	0.063	0.3	99.7
22	0.024	0.1	100.0	22	0.056	0.3	100.0
	22.000	100.0			22.000	100.0	

Sum of variance inflation factors:

$\Sigma\ 1/\lambda_i = 221.02$ $\Sigma\ 1/\lambda_i = 79.17$

The spectral condition number:

$\lambda_1/\lambda_{22} = 204.53$ $\lambda_1/\lambda_{22} = 67.22$

Determinant of the correlation matrix:

$|\mathbf{X'X}| = \lambda_1 \cdot \ldots \cdot \lambda_{22} = 4.73826 \cdot 10^{-12}$ $|\mathbf{X'X}| = \lambda_1 \cdot \ldots \cdot \lambda_{22} = 1.63570 \cdot 10^{-6}$

Haitovskys test variable:

$\chi^2_H(231) = 0.000$ $\chi^2_H(231) = 0.000$

1.00	0.74	0.53	0.36	0.23	0.14	0.14	0.18	0.22	0.25	0.22	0.63	0.51	0.38	0.22	0.06	-0.06	-0.09	-0.04	0.07	0.18	0.23
0.74	1.00	0.74	0.53	0.36	0.23	0.15	0.15	0.20	0.23	0.26	0.41	0.63	0.52	0.38	0.21	0.04	-0.07	-0.09	-0.02	0.08	0.18
0.53	0.74	1.00	0.73	0.52	0.35	0.23	0.16	0.17	0.21	0.25	0.05	0.42	0.63	0.51	0.35	0.19	0.04	-0.07	-0.07	-0.01	0.08
0.36	0.53	0.73	1.00	0.74	0.53	0.35	0.23	0.15	0.16	0.21	-0.28	0.04	0.41	0.63	0.50	0.35	0.19	0.04	-0.07	-0.07	-0.01
0.23	0.36	0.52	0.74	1.00	0.74	0.52	0.34	0.20	0.11	0.13	-0.40	-0.28	0.03	0.42	0.63	0.51	0.36	0.19	0.02	-0.08	-0.08
0.14	0.23	0.35	0.53	0.74	1.00	0.73	0.51	0.31	0.17	0.09	-0.30	-0.40	-0.29	0.03	0.42	0.64	0.52	0.36	0.17	0.00	-0.09
0.14	0.15	0.23	0.35	0.52	0.73	1.00	0.73	0.50	0.30	0.17	-0.04	-0.29	-0.40	-0.29	0.03	0.41	0.64	0.52	0.36	0.17	0.00
0.18	0.15	0.16	0.23	0.34	0.51	0.73	1.00	0.73	0.51	0.32	0.25	-0.01	-0.27	-0.40	-0.31	0.01	0.39	0.63	0.52	0.37	0.17
0.22	0.20	0.17	0.15	0.20	0.31	0.50	0.73	1.00	0.75	0.53	0.40	0.28	0.01	-0.28	-0.43	-0.33	-0.01	0.39	0.64	0.53	0.36
0.25	0.23	0.21	0.16	0.11	0.17	0.30	0.51	0.75	1.00	0.76	0.36	0.43	0.30	-0.01	-0.32	-0.45	-0.34	0.00	0.40	0.63	0.51
0.22	0.26	0.25	0.21	0.13	0.09	0.17	0.32	0.53	0.76	1.00	0.17	0.39	0.44	0.28	-0.05	-0.35	-0.46	-0.32	0.02	0.41	0.62
0.63	0.41	0.05	-0.28	-0.40	-0.30	-0.04	0.25	0.40	0.36	0.17	1.00	0.73	0.17	-0.39	-0.68	-0.61	-0.27	0.16	0.47	0.53	0.37
0.51	0.63	0.42	0.04	-0.28	-0.40	-0.29	-0.01	0.28	0.43	0.39	0.73	1.00	0.74	0.16	-0.40	-0.69	-0.62	-0.26	0.18	0.48	0.53
0.38	0.52	0.63	0.41	0.03	-0.29	-0.40	-0.27	0.01	0.30	0.44	0.17	0.74	1.00	0.72	0.14	-0.41	-0.69	-0.61	-0.24	0.19	0.48
0.22	0.38	0.51	0.63	0.42	0.03	-0.29	-0.40	-0.28	-0.01	0.28	-0.39	0.16	0.72	1.00	0.72	0.15	-0.41	-0.69	-0.61	-0.25	0.19
0.06	0.21	0.35	0.50	0.63	0.42	0.03	-0.31	-0.43	-0.32	-0.05	-0.68	-0.40	0.14	0.72	1.00	0.73	0.16	-0.41	-0.70	-0.62	-0.25
-0.06	0.04	0.19	0.35	0.51	0.64	0.41	0.01	-0.33	-0.45	-0.35	-0.61	-0.69	-0.41	0.15	0.73	1.00	0.73	0.15	-0.42	-0.71	-0.62
-0.09	-0.07	0.04	0.19	0.36	0.52	0.64	0.39	-0.01	-0.34	-0.46	-0.27	-0.62	-0.69	-0.41	0.16	0.73	1.00	0.73	0.14	-0.42	-0.71
-0.04	-0.09	-0.07	0.04	0.19	0.36	0.52	0.63	0.39	0.00	-0.32	0.16	-0.26	-0.61	-0.69	-0.41	0.15	0.73	1.00	0.73	0.14	-0.42
0.07	-0.02	-0.07	-0.07	0.02	0.17	0.36	0.52	0.64	0.40	0.02	0.47	0.18	-0.24	-0.61	-0.70	-0.42	0.14	0.73	1.00	0.73	0.14
0.18	0.08	-0.01	-0.07	-0.08	0.00	0.17	0.37	0.53	0.63	0.41	0.53	0.48	0.19	-0.25	-0.62	-0.71	-0.42	0.14	0.73	1.00	0.73
0.23	0.18	0.08	-0.01	-0.08	-0.09	0.00	0.17	0.36	0.51	0.62	0.37	0.53	0.48	0.19	-0.25	-0.62	-0.71	-0.42	0.14	0.73	1.00

Correlation matrix for the original variables when $\rho_{a_1 a_2} \approx .98$.

1.00	-0.03	0.55	-0.04	0.26	-0.05	0.12	0.08	0.07	0.20	0.01	0.64	0.15	0.38	0.05	0.10	-0.11	-0.03	-0.05	0.04	0.13	0.13
-0.03	1.00	-0.02	0.55	-0.04	0.27	-0.05	0.12	0.08	0.09	0.20	0.14	0.63	0.15	0.38	0.05	0.10	-0.11	-0.03	-0.05	0.05	0.13
0.55	-0.02	1.00	-0.02	0.54	-0.04	0.27	-0.04	0.13	0.10	0.10	0.12	0.15	0.64	0.15	0.36	0.03	0.09	-0.11	-0.03	-0.04	0.05
-0.04	0.55	-0.02	1.00	-0.02	0.55	-0.05	0.27	-0.04	0.14	0.11	-0.29	0.12	0.16	0.64	0.14	0.35	0.03	0.09	-0.11	-0.02	-0.03
0.26	-0.04	0.54	-0.02	1.00	-0.02	0.55	-0.06	0.27	-0.06	0.13	-0.23	-0.30	0.11	0.15	0.64	0.15	0.36	0.02	0.08	-0.12	-0.02
-0.05	0.27	-0.04	0.55	-0.02	1.00	-0.02	0.54	-0.06	0.24	-0.06	-0.26	-0.22	-0.30	0.11	0.15	0.63	0.16	0.36	0.02	0.06	-0.13
0.12	-0.05	0.27	-0.05	0.55	-0.02	1.00	-0.03	0.54	-0.08	0.23	-0.01	-0.26	-0.23	-0.30	0.12	0.16	0.64	0.16	0.35	0.01	0.06
0.08	0.12	-0.04	0.27	-0.06	0.54	-0.03	1.00	-0.02	0.54	-0.06	0.17	0.01	-0.24	-0.23	-0.31	0.09	0.14	0.63	0.17	0.36	0.01
0.07	0.08	0.13	-0.04	0.27	-0.06	0.54	-0.02	1.00	0.00	0.55	0.26	0.18	0.02	-0.24	-0.23	-0.32	0.08	0.14	0.64	0.17	0.36
0.20	0.09	0.10	0.14	-0.06	0.24	-0.08	0.54	0.00	1.00	0.03	0.28	0.30	0.21	0.02	-0.28	-0.27	-0.34	0.09	0.15	0.64	0.17
0.01	0.20	0.10	0.11	0.13	-0.06	0.23	-0.06	0.55	0.03	1.00	0.07	0.29	0.32	0.21	0.01	-0.29	-0.28	-0.33	0.09	0.17	0.64
0.64	0.14	0.12	-0.29	-0.23	-0.26	-0.01	0.17	0.26	0.28	0.07	1.00	0.57	0.18	-0.40	-0.58	-0.56	-0.21	0.15	0.42	0.47	0.31
0.15	0.63	0.15	0.12	-0.30	-0.22	-0.26	0.01	0.18	0.30	0.29	0.57	1.00	0.58	0.18	-0.41	-0.59	-0.58	-0.21	0.16	0.44	0.47
0.38	0.15	0.64	0.16	0.11	-0.30	-0.23	-0.24	0.02	0.21	0.32	0.18	0.58	1.00	0.58	0.15	-0.43	-0.60	-0.57	-0.19	0.18	0.44
0.05	0.38	0.15	0.64	0.15	0.11	-0.30	-0.23	-0.24	0.02	0.21	-0.40	0.18	0.58	1.00	0.57	0.14	-0.43	-0.60	-0.56	-0.19	0.18
0.10	0.05	0.36	0.14	0.64	0.15	0.12	-0.31	-0.23	-0.28	0.01	-0.58	-0.41	0.15	0.57	1.00	0.58	0.16	-0.42	-0.61	-0.58	-0.19
-0.11	0.10	0.03	0.35	0.15	0.63	0.16	0.09	-0.32	-0.27	-0.29	-0.56	-0.59	-0.43	0.14	0.58	1.00	0.59	0.16	-0.43	-0.62	-0.58
-0.03	-0.11	0.09	0.03	0.36	0.16	0.64	0.14	0.08	-0.34	-0.28	-0.21	-0.58	-0.60	-0.43	0.16	0.59	1.00	0.58	0.15	-0.44	-0.62
-0.05	-0.03	-0.11	0.09	0.02	0.36	0.16	0.63	0.14	0.09	-0.33	0.15	-0.21	-0.57	-0.60	-0.42	0.16	0.58	1.00	0.58	0.15	-0.44
0.04	-0.05	-0.03	-0.11	0.08	0.02	0.35	0.17	0.64	0.15	0.09	0.42	0.16	-0.19	-0.56	-0.61	-0.43	0.15	0.58	1.00	0.59	0.15
0.13	0.05	-0.04	-0.02	-0.12	0.06	0.01	0.36	0.17	0.64	0.17	0.47	0.44	0.18	-0.19	-0.58	-0.62	-0.44	0.15	0.59	1.00	0.58
0.13	0.13	0.05	-0.03	-0.02	-0.13	0.06	0.01	0.36	0.17	0.64	0.31	0.47	0.44	0.18	-0.19	-0.58	-0.62	-0.44	0.15	0.58	1.00

Correlation matrix for the transformed variables when $\rho_{a_1 a_2} \approx .98$.

Some measures of multicollinearity for the two simulated correlation matrices with $\rho_{a_1 a_2} \approx .98$.

	Original variables				Transformed variables		
i	λ_i	%	Σ%	i	λ_i	%	Σ%
1	6.333	28.8	28.8	1	5.153	23.4	23.4
2	5.201	23.6	52.4	2	4.180	19.0	42.4
3	5.078	23.1	75.5	3	2.954	13.4	55.9
4	1.986	9.0	84.5	4	2.770	12.6	68.4
5	0.915	4.2	88.7	5	1.910	8.7	77.1
6	0.717	3.3	92.0	6	1.338	6.1	83.2
7	0.421	1.9	93.9	7	0.837	3.8	87.0
8	0.337	1.5	95.4	8	0.627	2.9	89.9
9	0.261	1.2	96.6	9	0.566	2.6	92.4
10	0.198	0.9	97.5	10	0.492	2.2	94.7
11	0.175	0.8	98.3	11	0.463	2.1	96.8
12	0.169	0.8	99.1	12	0.408	1.9	98.6
13	0.148	0.7	99.7	13	0.214	1.0	99.6
14	0.022	0.1	99.8	14	0.024	0.1	99.7
15	0.017	0.1	99.9	15	0.016	0.1	99.8
16	0.012	0.1	100.0	16	0.013	0.1	99.8
17	0.005	0.0	100.0	17	0.010	0.0	99.9
18	0.002	0.0	100.0	18	0.009	0.0	99.9
19	0.001	0.0	100.0	19	0.006	0.0	100.0
20	0.001	0.0	100.0	20	0.006	0.0	100.0
21	0.001	0.0	100.0	21	0.003	0.0	100.0
22	0.001	0.0	100.0	22	0.002	0.0	100.0
	22.000	100.0			22.000	100.0	

Sum of variance inflation factors:

$\Sigma\ 1/\lambda_i = 5069.50$ $\qquad$ $\Sigma\ 1/\lambda_i = 1516.92$

The spectral condition number:

$\lambda_1/\lambda_{22} = 7306.35$ $\qquad$ $\lambda_1/\lambda_{22} = 2120.83$

Determinant of the correlation matrix:

$|\mathbf{X'X}| = \lambda_1 \cdot \ldots \cdot \lambda_{22} = 2.34242 \cdot 10^{-25}$ $\qquad$ $|\mathbf{X'X}| = \lambda_1 \cdot \ldots \cdot \lambda_{22} = 2.73348 \cdot 10^{-19}$

Haitovskys test variable:

$\chi^2_H(231) = 0.000$ $\qquad$ $\chi^2_H(231) = 0.000$

APPENDIX 4. Autocorrelation and cross correlation functions for theoretical processes.

	$\rho_{a_1 a_2} \approx .58$				$\rho_{a_1 a_2} \approx .98$			
	x_1	x_2	x_1'	x_2'	x_1	x_2	x_1'	x_2'
lag k	$\rho(k)$	$\rho(k)$	$\rho(k)$	$\rho(k)$	$\rho(k)$	$\rho(k)$	$\rho(k)$	$\rho(k)$
0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	0.70	0.71	–0.11	0.51	0.70	0.71	–0.11	0.51
2	0.49	0.14	0.53	0.16	0.49	0.14	0.53	0.16
3	0.34	–0.36	–0.08	–0.39	0.34	–0.36	–0.08	–0.39
4	0.24	–0.55	0.28	–0.45	0.24	–0.55	0.28	–0.45
5	0.17	–0.42	–0.06	–0.39	0.17	–0.42	–0.06	–0.39
6	0.12	–0.11	0.15	–0.06	0.12	–0.11	0.15	–0.06
7	0.08	0.17	–0.04	0.15	0.08	0.17	–0.04	0.15
8	0.06	0.30	0.08	0.28	0.06	0.30	0.08	0.28
9	0.04	0.25	–0.02	0.20	0.04	0.25	–0.02	0.20
10	0.03	0.08	0.04	0.07	0.03	0.08	0.04	0.07

	$\rho_{x_1 x_2}$		$\rho_{x_1' x_2'}$		$\rho_{x_1 x_2}$		$\rho_{x_1' x_2'}$	
lag k	$\rho(k)$	$\rho(-k)$	$\rho(k)$	$\rho(-k)$	$\rho(k)$	$\rho(-k)$	$\rho(k)$	$\rho(-k)$
0	0.39	0.39	0.39	0.39	0.66	0.66	0.67	0.67
1	0.28	0.27	0.09	0.03	0.48	0.46	0.15	0.04
2	0.06	0.19	0.10	0.21	0.10	0.32	0.17	0.35
3	–0.14	0.13	–0.15	0.00	–0.23	0.26	–0.26	0.01
4	–0.21	0.09	–0.11	0.11	–0.36	0.16	–0.18	0.18
5	–0.17	0.07	–0.14	–0.00	–0.28	0.11	–0.24	–0.01
6	–0.05	0.05	–0.00	0.06	–0.08	0.08	–0.01	0.10
7	0.07	0.03	0.03	–0.00	0.11	0.05	0.05	–0.01
8	0.12	0.02	0.09	0.03	0.20	0.04	0.16	0.05
9	0.10	0.02	0.05	–0.00	0.17	0.03	0.09	–0.01
10	0.03	0.01	0.03	0.02	0.06	0.02	0.05	0.03

APPENDIX 5. Correlation matrices and eigenvalues for theoretical processes.

1.00	0.70	0.49	0.34	0.24	0.17	0.12	0.08	0.06	0.04	0.03	0.39	0.27	0.19	0.13	0.09	0.07	0.05	0.03	0.02	0.02	0.01
0.70	1.00	0.70	0.49	0.34	0.24	0.17	0.12	0.08	0.06	0.04	0.28	0.39	0.27	0.19	0.13	0.09	0.07	0.05	0.03	0.02	0.02
0.49	0.70	1.00	0.70	0.49	0.34	0.24	0.17	0.12	0.08	0.06	0.06	0.28	0.39	0.27	0.19	0.13	0.09	0.07	0.05	0.03	0.02
0.34	0.49	0.70	1.00	0.70	0.49	0.34	0.24	0.17	0.12	0.08	-0.14	0.06	0.28	0.39	0.27	0.19	0.13	0.09	0.07	0.05	0.03
0.24	0.34	0.49	0.70	1.00	0.70	0.49	0.34	0.24	0.17	0.12	-0.21	-0.14	0.06	0.28	0.39	0.27	0.19	0.13	0.09	0.07	0.05
0.17	0.24	0.34	0.49	0.70	1.00	0.70	0.49	0.34	0.24	0.17	-0.17	-0.21	-0.14	0.06	0.28	0.39	0.27	0.19	0.13	0.09	0.07
0.12	0.17	0.24	0.34	0.49	0.70	1.00	0.70	0.49	0.34	0.24	-0.05	-0.17	-0.21	-0.14	0.06	0.28	0.39	0.27	0.19	0.13	0.09
0.08	0.12	0.17	0.24	0.34	0.49	0.70	1.00	0.70	0.49	0.34	0.07	-0.05	-0.17	-0.21	-0.14	0.06	0.28	0.39	0.27	0.19	0.13
0.06	0.08	0.12	0.17	0.24	0.34	0.49	0.70	1.00	0.70	0.49	0.12	0.07	-0.05	-0.17	-0.21	-0.14	0.06	0.28	0.39	0.27	0.19
0.04	0.06	0.08	0.12	0.17	0.24	0.34	0.49	0.70	1.00	0.70	0.10	0.12	0.07	-0.05	-0.17	-0.21	-0.14	0.06	0.28	0.39	0.27
0.03	0.04	0.06	0.08	0.12	0.17	0.24	0.34	0.49	0.70	1.00	0.03	0.10	0.12	0.07	-0.05	-0.17	-0.21	-0.14	0.06	0.28	0.39
0.39	0.28	0.06	-0.14	-0.21	-0.17	-0.05	0.07	0.12	0.10	0.03	1.00	0.71	0.14	-0.36	-0.55	-0.42	-0.11	0.17	0.30	0.25	0.08
0.27	0.39	0.28	0.06	-0.14	-0.21	-0.17	-0.05	0.07	0.12	0.10	0.71	1.00	0.71	0.14	-0.36	-0.55	-0.42	-0.11	0.17	0.30	0.25
0.19	0.27	0.39	0.28	0.06	-0.14	-0.21	-0.17	-0.05	0.07	0.12	0.14	0.71	1.00	0.71	0.14	-0.36	-0.55	-0.42	-0.11	0.17	0.30
0.13	0.19	0.27	0.39	0.28	0.06	-0.14	-0.21	-0.17	-0.05	0.07	-0.36	0.14	0.71	1.00	0.71	0.14	-0.36	-0.55	-0.42	-0.11	0.17
0.09	0.13	0.19	0.27	0.39	0.28	0.06	-0.14	-0.21	-0.17	-0.05	-0.55	-0.36	0.14	0.71	1.00	0.71	0.14	-0.36	-0.55	-0.42	-0.11
0.07	0.09	0.13	0.19	0.27	0.39	0.28	0.06	-0.14	-0.21	-0.17	-0.42	-0.55	-0.36	0.14	0.71	1.00	0.71	0.14	-0.36	-0.55	-0.42
0.05	0.07	0.09	0.13	0.19	0.27	0.39	0.28	0.06	-0.14	-0.21	-0.11	-0.42	-0.55	-0.36	0.14	0.71	1.00	0.71	0.14	-0.36	-0.55
0.03	0.05	0.07	0.09	0.13	0.19	0.27	0.39	0.28	0.06	-0.14	0.17	-0.11	-0.42	-0.55	-0.36	0.14	0.71	1.00	0.71	0.14	-0.36
0.02	0.03	0.05	0.07	0.09	0.13	0.19	0.27	0.39	0.28	0.06	0.30	0.17	-0.11	-0.42	-0.55	-0.36	0.14	0.71	1.00	0.71	0.14
0.02	0.02	0.03	0.05	0.07	0.09	0.13	0.19	0.27	0.39	0.28	0.25	0.30	0.17	-0.11	-0.42	-0.55	-0.36	0.14	0.71	1.00	0.71
0.01	0.02	0.02	0.03	0.05	0.07	0.09	0.13	0.19	0.27	0.39	0.08	0.25	0.30	0.17	-0.11	-0.42	-0.55	-0.36	0.14	0.71	1.00

Correlation matrix for the original variables when $\rho_{a_1 a_2} \approx .58$.

1.00	-0.11	0.53	-0.08	0.28	-0.06	0.15	-0.04	0.08	-0.02	0.04	0.39	0.03	0.21	0.00	0.11	0.00	0.06	0.00	0.03	0.00	0.02
-0.11	1.00	-0.11	0.53	-0.08	0.28	-0.06	0.15	-0.04	0.08	-0.02	0.09	0.39	0.03	0.21	0.00	0.11	0.00	0.06	0.00	0.03	0.00
0.53	-0.11	1.00	-0.11	0.53	-0.08	0.28	-0.06	0.15	-0.04	0.08	0.10	0.09	0.39	0.03	0.21	0.00	0.11	0.00	0.06	0.00	0.03
-0.08	0.53	-0.11	1.00	-0.11	0.53	-0.08	0.28	-0.06	0.15	-0.04	-0.15	0.10	0.09	0.39	0.03	0.21	0.00	0.11	0.00	0.06	0.00
0.28	-0.08	0.53	-0.11	1.00	-0.11	0.53	-0.08	0.28	-0.06	0.15	-0.11	-0.15	0.10	0.09	0.39	0.03	0.21	0.00	0.11	0.00	0.06
-0.06	0.28	-0.08	0.53	-0.11	1.00	-0.11	0.53	-0.08	0.28	-0.06	-0.14	-0.11	-0.15	0.10	0.09	0.39	0.03	0.21	0.00	0.11	0.00
0.15	-0.06	0.28	-0.08	0.53	-0.11	1.00	-0.11	0.53	-0.08	0.28	0.00	-0.14	-0.11	-0.15	0.10	0.09	0.39	0.03	0.21	0.00	0.11
-0.04	0.15	-0.06	0.28	-0.08	0.53	-0.11	1.00	-0.11	0.53	-0.08	0.03	0.00	-0.14	-0.11	-0.15	0.10	0.09	0.39	0.03	0.21	0.00
0.08	-0.04	0.15	-0.06	0.28	-0.08	0.53	-0.11	1.00	-0.11	0.53	0.09	0.03	0.00	-0.14	-0.11	-0.15	0.10	0.09	0.39	0.03	0.21
-0.02	0.08	-0.04	0.15	-0.06	0.28	-0.08	0.53	-0.11	1.00	-0.11	0.05	0.09	0.03	0.00	-0.14	-0.11	-0.15	0.10	0.09	0.39	0.03
0.04	-0.02	0.08	-0.04	0.15	-0.06	0.28	-0.08	0.53	-0.11	1.00	0.03	0.05	0.09	0.03	0.00	-0.14	-0.11	-0.15	0.10	0.09	0.39
0.39	0.09	0.10	-0.15	-0.11	-0.14	0.00	0.03	0.09	0.05	0.03	1.00	0.51	0.16	-0.39	-0.45	-0.39	-0.06	0.15	0.28	0.20	0.07
0.03	0.39	0.09	0.10	-0.15	-0.11	-0.14	0.00	0.03	0.09	0.05	0.51	1.00	0.51	0.16	-0.39	-0.45	-0.39	-0.06	0.15	0.28	0.20
0.21	0.03	0.39	0.09	0.10	-0.15	-0.11	-0.14	0.00	0.03	0.09	0.16	0.51	1.00	0.51	0.16	-0.39	-0.45	-0.39	-0.06	0.15	0.28
0.00	0.21	0.03	0.39	0.09	0.10	-0.15	-0.11	-0.14	0.00	0.03	-0.39	0.16	0.51	1.00	0.51	0.16	-0.39	-0.45	-0.39	-0.06	0.15
0.11	0.00	0.21	0.03	0.39	0.09	0.10	-0.15	-0.11	-0.14	0.00	-0.45	-0.39	0.16	0.51	1.00	0.51	0.16	-0.39	-0.45	-0.39	-0.06
0.00	0.11	0.00	0.21	0.03	0.39	0.09	0.10	-0.15	-0.11	-0.14	-0.39	-0.45	-0.39	0.16	0.51	1.00	0.51	0.16	-0.39	-0.45	-0.39
0.06	0.00	0.11	0.00	0.21	0.03	0.39	0.09	0.10	-0.15	-0.11	-0.06	-0.39	-0.45	-0.39	0.16	0.51	1.00	0.51	0.16	-0.39	-0.45
0.00	0.06	0.00	0.11	0.00	0.21	0.03	0.39	0.09	0.10	-0.15	0.15	-0.06	-0.39	-0.45	-0.39	0.16	0.51	1.00	0.51	0.16	-0.39
0.03	0.00	0.06	0.00	0.11	0.00	0.21	0.03	0.39	0.09	0.10	0.28	0.15	-0.06	-0.39	-0.45	-0.39	0.16	0.51	1.00	0.51	0.16
0.00	0.03	0.00	0.06	0.00	0.11	0.00	0.21	0.03	0.39	0.09	0.20	0.28	0.15	-0.06	-0.39	-0.45	-0.39	0.16	0.51	1.00	0.51
0.02	0.00	0.03	0.00	0.06	0.00	0.11	0.00	0.21	0.03	0.39	0.07	0.20	0.28	0.15	-0.06	-0.39	-0.45	-0.39	0.16	0.51	1.00

Correlation matrix for the transformed variables when $\rho_{a_1 a_2} \approx .58$.

Some measures of multicollinearity for the two theoretical correlation matrices with $\rho_{a_1 a_2} \approx .58$.

	Original variables				Transformed variables		
i	λ_i	%	Σ%	i	λ_i	%	Σ%
1	4.786	21.8	21.8	1	3.684	16.7	16.7
2	4.425	20.1	41.9	2	3.208	14.6	31.3
3	3.914	17.8	59.7	3	3.111	14.1	45.5
4	2.517	11.4	71.1	4	2.329	10.6	56.1
5	1.352	6.1	77.2	5	1.773	8.1	64.1
6	0.925	4.2	81.4	6	1.474	6.7	70.8
7	0.754	3.4	84.9	7	1.000	4.5	75.4
8	0.627	2.9	87.7	8	0.802	3.6	79.0
9	0.489	2.2	89.9	9	0.725	3.3	82.3
10	0.461	2.1	92.0	10	0.558	2.5	84.8
11	0.331	1.5	93.5	11	0.482	2.2	87.0
12	0.280	1.3	94.8	12	0.445	2.0	89.1
13	0.248	1.1	95.9	13	0.429	2.0	91.0
14	0.224	1.0	97.0	14	0.415	1.9	92.9
15	0.201	0.9	97.9	15	0.384	1.7	94.6
16	0.189	0.9	98.7	16	0.331	1.5	96.1
17	0.123	0.6	99.3	17	0.219	1.0	97.1
18	0.061	0.3	99.6	18	0.205	0.9	98.1
19	0.036	0.2	99.7	19	0.131	0.6	98.7
20	0.024	0.1	99.8	20	0.120	0.5	99.2
21	0.019	0.1	99.9	21	0.089	0.4	99.6
22	0.016	0.1	100.0	14	0.086	0.4	100.0
	22.000	100.0			22.000	100.0	

Sum of variance inflation factors:

$\Sigma\ 1/\lambda_i = 245.69$ $\qquad$ $\Sigma\ 1/\lambda_i = 71.00$

The spectral condition number:

$\lambda_1/\lambda_{22} = 300.76$ $\qquad$ $\lambda_1/\lambda_{22} = 42.75$

Determinant of the correlation matrix:

$|\mathbf{X'X}| = \lambda_1 \cdot \ldots \cdot \lambda_{22} = 1.04877 \cdot 10^{-11}$ $\qquad$ $|\mathbf{X'X}| = \lambda_1 \cdot \ldots \cdot \lambda_{22} = 1.89621 \cdot 10^{-6}$

Haitovskys test variable:

$\chi^2_H(231) = 0.000$ $\qquad$ $\chi^2_H(231) = 0.000$

1.00	0.70	0.49	0.34	0.24	0.17	0.12	0.08	0.06	0.04	0.03	0.58	0.41	0.28	0.20	0.14	0.10	0.07	0.05	0.03	0.02	0.02
0.70	1.00	0.70	0.49	0.34	0.24	0.17	0.12	0.08	0.06	0.04	0.42	0.58	0.41	0.28	0.20	0.14	0.10	0.07	0.05	0.03	0.02
0.49	0.70	1.00	0.70	0.49	0.34	0.24	0.17	0.12	0.08	0.06	0.09	0.42	0.58	0.41	0.28	0.20	0.14	0.10	0.07	0.05	0.03
0.34	0.49	0.70	1.00	0.70	0.49	0.34	0.24	0.17	0.12	0.08	-0.20	0.09	0.42	0.58	0.41	0.28	0.20	0.14	0.10	0.07	0.05
0.24	0.34	0.49	0.70	1.00	0.70	0.49	0.34	0.24	0.17	0.12	-0.32	-0.20	0.09	0.42	0.58	0.41	0.28	0.20	0.14	0.10	0.07
0.17	0.24	0.34	0.49	0.70	1.00	0.70	0.49	0.34	0.24	0.17	-0.25	-0.32	-0.20	0.09	0.42	0.58	0.41	0.28	0.20	0.14	0.10
0.12	0.17	0.24	0.34	0.49	0.70	1.00	0.70	0.49	0.34	0.24	-0.07	-0.25	-0.32	-0.20	0.09	0.42	0.58	0.41	0.28	0.20	0.14
0.08	0.12	0.17	0.24	0.34	0.49	0.70	1.00	0.70	0.49	0.34	0.10	-0.07	-0.25	-0.32	-0.20	0.09	0.42	0.58	0.41	0.28	0.20
0.06	0.08	0.12	0.17	0.24	0.34	0.49	0.70	1.00	0.70	0.49	0.18	0.10	-0.07	-0.25	-0.32	-0.20	0.09	0.42	0.58	0.41	0.28
0.04	0.06	0.08	0.12	0.17	0.24	0.34	0.49	0.70	1.00	0.70	0.15	0.18	0.10	-0.07	-0.25	-0.32	-0.20	0.09	0.42	0.58	0.41
0.03	0.04	0.06	0.08	0.12	0.17	0.24	0.34	0.49	0.70	1.00	0.05	0.15	0.18	0.10	-0.07	-0.25	-0.32	-0.20	0.09	0.42	0.58
0.58	0.42	0.09	-0.20	-0.32	-0.25	-0.07	0.10	0.18	0.15	0.05	1.00	0.71	0.14	-0.36	-0.55	-0.42	-0.11	0.17	0.30	0.25	0.08
0.41	0.58	0.42	0.09	-0.20	-0.32	-0.25	-0.07	0.10	0.18	0.15	0.71	1.00	0.71	0.14	-0.36	-0.55	-0.42	-0.11	0.17	0.30	0.25
0.28	0.41	0.58	0.42	0.09	-0.20	-0.32	-0.25	-0.07	0.10	0.18	0.14	0.71	1.00	0.71	0.14	-0.36	-0.55	-0.42	-0.11	0.17	0.30
0.20	0.28	0.41	0.58	0.42	0.09	-0.20	-0.32	-0.25	-0.07	0.10	-0.36	0.14	0.71	1.00	0.71	0.14	-0.36	-0.55	-0.42	-0.11	0.17
0.14	0.20	0.28	0.41	0.58	0.42	0.09	-0.20	-0.32	-0.25	-0.07	-0.55	-0.36	0.14	0.71	1.00	0.71	0.14	-0.36	-0.55	-0.42	-0.11
0.10	0.14	0.20	0.28	0.41	0.58	0.42	0.09	-0.20	-0.32	-0.25	-0.42	-0.55	-0.36	0.14	0.71	1.00	0.71	0.14	-0.36	-0.55	-0.42
0.07	0.10	0.14	0.20	0.28	0.41	0.58	0.42	0.09	-0.20	-0.32	-0.11	-0.42	-0.55	-0.36	0.14	0.71	1.00	0.71	0.14	-0.36	-0.55
0.05	0.07	0.10	0.14	0.20	0.28	0.41	0.58	0.42	0.09	-0.20	0.17	-0.11	-0.42	-0.55	-0.36	0.14	0.71	1.00	0.71	0.14	-0.36
0.03	0.05	0.07	0.10	0.14	0.20	0.28	0.41	0.58	0.42	0.09	0.30	0.17	-0.11	-0.42	-0.55	-0.36	0.14	0.71	1.00	0.71	0.14
0.02	0.03	0.05	0.07	0.10	0.14	0.20	0.28	0.41	0.58	0.42	0.25	0.30	0.17	-0.11	-0.42	-0.55	-0.36	0.14	0.71	1.00	0.71
0.02	0.02	0.03	0.05	0.07	0.10	0.14	0.20	0.28	0.41	0.58	0.08	0.25	0.30	0.17	-0.11	-0.42	-0.55	-0.36	0.14	0.71	1.00

Correlation matrix for the original variables when $\rho_{a_1 a_2} \approx .87$.

1.00	-0.11	0.53	-0.08	0.28	-0.06	0.15	-0.04	0.08	-0.02	0.04	0.59	0.04	0.31	0.00	0.16	-0.01	0.09	-0.01	0.05	-0.01	0.02
-0.11	1.00	-0.11	0.53	-0.08	0.28	-0.06	0.15	-0.04	0.08	-0.02	0.13	0.59	0.04	0.31	0.00	0.16	-0.01	0.09	-0.01	0.05	-0.01
0.53	-0.11	1.00	-0.11	0.53	-0.08	0.28	-0.06	0.15	-0.04	0.08	0.15	0.13	0.59	0.04	0.31	0.00	0.16	-0.01	0.09	-0.01	0.05
-0.08	0.53	-0.11	1.00	-0.11	0.53	-0.08	0.28	-0.06	0.15	-0.04	-0.23	0.15	0.13	0.59	0.04	0.31	0.00	0.16	-0.01	0.09	-0.01
0.28	-0.08	0.53	-0.11	1.00	-0.11	0.53	-0.08	0.28	-0.06	0.15	-0.16	-0.23	0.15	0.13	0.59	0.04	0.31	0.00	0.16	-0.01	0.09
-0.06	0.28	-0.08	0.53	-0.11	1.00	-0.11	0.53	-0.08	0.28	-0.06	-0.21	-0.16	-0.23	0.15	0.13	0.59	0.04	0.31	0.00	0.16	-0.01
0.15	-0.06	0.28	-0.08	0.53	-0.11	1.00	-0.11	0.53	-0.08	0.28	0.00	-0.21	-0.16	-0.23	0.15	0.13	0.59	0.04	0.31	0.00	0.16
-0.04	0.15	-0.06	0.28	-0.08	0.53	-0.11	1.00	-0.11	0.53	-0.08	0.05	0.00	-0.21	-0.16	-0.23	0.15	0.13	0.59	0.04	0.31	0.00
0.08	-0.04	0.15	-0.06	0.28	-0.08	0.53	-0.11	1.00	-0.11	0.53	0.14	0.05	0.00	-0.21	-0.16	-0.23	0.15	0.13	0.59	0.04	0.31
-0.02	0.08	-0.04	0.15	-0.06	0.28	-0.08	0.53	-0.11	1.00	-0.11	0.08	0.14	0.05	0.00	-0.21	-0.16	-0.23	0.15	0.13	0.59	0.04
0.04	-0.02	0.08	-0.04	0.15	-0.06	0.28	-0.08	0.53	-0.11	1.00	0.04	0.08	0.14	0.05	0.00	-0.21	-0.16	-0.23	0.15	0.13	0.59
0.59	0.13	0.15	-0.23	-0.16	-0.21	0.00	0.05	0.14	0.08	0.04	1.00	0.51	0.16	-0.39	-0.45	-0.39	-0.06	0.15	0.28	0.20	0.07
0.04	0.59	0.13	0.15	-0.23	-0.16	-0.21	0.00	0.05	0.14	0.08	0.51	1.00	0.51	0.16	-0.39	-0.45	-0.39	-0.06	0.15	0.28	0.20
0.31	0.04	0.59	0.13	0.15	-0.23	-0.16	-0.21	0.00	0.05	0.14	0.16	0.51	1.00	0.51	0.16	-0.39	-0.45	-0.39	-0.06	0.15	0.28
0.00	0.31	0.04	0.59	0.13	0.15	-0.23	-0.16	-0.21	0.00	0.05	-0.39	0.16	0.51	1.00	0.51	0.16	-0.39	-0.45	-0.39	-0.06	0.15
0.16	0.00	0.31	0.04	0.59	0.13	0.15	-0.23	-0.16	-0.21	0.00	-0.45	-0.39	0.16	0.51	1.00	0.51	0.16	-0.39	-0.45	-0.39	-0.06
-0.01	0.16	0.00	0.31	0.04	0.59	0.13	0.15	-0.23	-0.16	-0.21	-0.39	-0.45	-0.39	0.16	0.51	1.00	0.51	0.16	-0.39	-0.45	-0.39
0.09	-0.01	0.16	0.00	0.31	0.04	0.59	0.13	0.15	-0.23	-0.16	-0.06	-0.39	-0.45	-0.39	0.16	0.51	1.00	0.51	0.16	-0.39	-0.45
-0.01	0.09	-0.01	0.16	0.00	0.31	0.04	0.59	0.13	0.15	-0.23	0.15	-0.06	-0.39	-0.45	-0.39	0.16	0.51	1.00	0.51	0.16	-0.39
0.05	-0.01	0.09	-0.01	0.16	0.00	0.31	0.04	0.59	0.13	0.15	0.28	0.15	-0.06	-0.39	-0.45	-0.39	0.16	0.51	1.00	0.51	0.16
-0.01	0.05	-0.01	0.09	-0.01	0.16	0.00	0.31	0.04	0.59	0.13	0.20	0.28	0.15	-0.06	-0.39	-0.45	-0.39	0.16	0.51	1.00	0.51
0.02	-0.01	0.05	-0.01	0.09	-0.01	0.16	0.00	0.31	0.04	0.59	0.07	0.20	0.28	0.15	-0.06	-0.39	-0.45	-0.39	0.16	0.51	1.00

Correlation matrix for the transformed variables when $\rho_{a_1 a_2} \approx .87$.

Some measures of multicollinearity for the two theoretical correlation matrices with $\rho_{a_1 a_2} \approx .87$.

	Original variables				Transformed variables		
i	λ_i	%	Σ%	i	λ_i	%	Σ%
1	5.187	23.6	23.6	1	3.905	17.7	17.7
2	4.855	22.1	45.6	2	3.439	15.6	33.4
3	4.355	19.8	65.4	3	3.354	15.2	48.6
4	2.705	12.3	77.7	4	2.602	11.8	60.5
5	1.227	5.6	83.3	5	2.008	9.1	69.6
6	0.844	3.8	87.1	6	1.604	7.3	76.9
7	0.575	2.6	89.8	7	1.086	4.9	81.8
8	0.401	1.8	91.6	8	0.816	3.7	85.5
9	0.324	1.5	93.1	9	0.671	3.0	88.6
10	0.273	1.2	94.3	10	0.543	2.5	91.0
11	0.239	1.1	95.4	11	0.503	2.3	93.3
12	0.219	1.0	96.4	12	0.475	2.2	95.5
13	0.213	1.0	97.3	13	0.299	1.4	96.8
14	0.200	0.9	98.3	14	0.147	0.7	97.5
15	0.190	0.9	99.1	15	0.136	0.6	98.1
16	0.096	0.4	99.6	16	0.127	0.6	98.7
17	0.043	0.2	99.8	17	0.075	0.3	99.0
18	0.021	0.1	99.8	18	0.070	0.3	99.4
19	0.013	0.1	99.9	19	0.043	0.2	99.6
20	0.009	0.0	99.9	20	0.040	0.2	99.7
21	0.007	0.0	100.0	21	0.029	0.1	99.9
22	0.006	0.0	100.0	22	0.029	0.1	100.0
	22.000	100.0			22.000	100.0	

Sum of variance inflation factors:

$\Sigma\ 1/\lambda_i = 642.09$ $\qquad$ $\Sigma\ 1/\lambda_i = 181.66$

The spectral condition number:

$\lambda_1/\lambda_{22} = 915.50$ $\qquad$ $\lambda_1/\lambda_{22} = 135.66$

Determinant of the correlation matrix:

$|\mathbf{X'X}| = \lambda_1 \cdot \ldots \cdot \lambda_{22} = 9.42407 \cdot 10^{-16}$ $\qquad$ $|\mathbf{X'X}| = \lambda_1 \cdot \ldots \cdot \lambda_{22} = 1.70391 \cdot 10^{-10}$

Haitovskys test variable:

$\chi^2_H(231) = 0.000$ $\qquad$ $\chi^2_H(231) = 0.000$

1.00	0.70	0.49	0.34	0.24	0.17	0.12	0.08	0.06	0.04	0.03	0.66	0.46	0.32	0.23	0.16	0.11	0.08	0.05	0.04	0.03	0.02
0.70	1.00	0.70	0.49	0.34	0.24	0.17	0.12	0.08	0.06	0.04	0.48	0.66	0.46	0.32	0.23	0.16	0.11	0.08	0.05	0.04	0.03
0.49	0.70	1.00	0.70	0.49	0.34	0.24	0.17	0.12	0.08	0.06	0.10	0.48	0.66	0.46	0.32	0.23	0.16	0.11	0.08	0.05	0.04
0.34	0.49	0.70	1.00	0.70	0.49	0.34	0.24	0.17	0.12	0.08	-0.23	0.10	0.48	0.66	0.46	0.32	0.23	0.16	0.11	0.08	0.05
0.24	0.34	0.49	0.70	1.00	0.70	0.49	0.34	0.24	0.17	0.12	-0.36	-0.23	0.10	0.48	0.66	0.46	0.32	0.23	0.16	0.11	0.08
0.17	0.24	0.34	0.49	0.70	1.00	0.70	0.49	0.34	0.24	0.17	-0.28	-0.36	-0.23	0.10	0.48	0.66	0.46	0.32	0.23	0.16	0.11
0.12	0.17	0.24	0.34	0.49	0.70	1.00	0.70	0.49	0.34	0.24	-0.08	-0.28	-0.36	-0.23	0.10	0.48	0.66	0.46	0.32	0.23	0.16
0.08	0.12	0.17	0.24	0.34	0.49	0.70	1.00	0.70	0.49	0.34	0.11	-0.08	-0.28	-0.36	-0.23	0.10	0.48	0.66	0.46	0.32	0.23
0.06	0.08	0.12	0.17	0.24	0.34	0.49	0.70	1.00	0.70	0.49	0.20	0.11	-0.08	-0.28	-0.36	-0.23	0.10	0.48	0.66	0.46	0.32
0.04	0.06	0.08	0.12	0.17	0.24	0.34	0.49	0.70	1.00	0.70	0.17	0.20	0.11	-0.08	-0.28	-0.36	-0.23	0.10	0.48	0.66	0.46
0.03	0.04	0.06	0.08	0.12	0.17	0.24	0.34	0.49	0.70	1.00	0.06	0.17	0.20	0.11	-0.08	-0.28	-0.36	-0.23	0.10	0.48	0.66
0.66	0.48	0.10	-0.23	-0.36	-0.28	-0.08	0.11	0.20	0.17	0.06	1.00	0.71	0.14	-0.36	-0.55	-0.42	-0.11	0.17	0.30	0.25	0.08
0.46	0.66	0.48	0.10	-0.23	-0.36	-0.28	-0.08	0.11	0.20	0.17	0.71	1.00	0.71	0.14	-0.36	-0.55	-0.42	-0.11	0.17	0.30	0.25
0.32	0.46	0.66	0.48	0.10	-0.23	-0.36	-0.28	-0.08	0.11	0.20	0.14	0.71	1.00	0.71	0.14	-0.36	-0.55	-0.42	-0.11	0.17	0.30
0.23	0.32	0.46	0.66	0.48	0.10	-0.23	-0.36	-0.28	-0.08	0.11	-0.36	0.14	0.71	1.00	0.71	0.14	-0.36	-0.55	-0.42	-0.11	0.17
0.16	0.23	0.32	0.46	0.66	0.48	0.10	-0.23	-0.36	-0.28	-0.08	-0.55	-0.36	0.14	0.71	1.00	0.71	0.14	-0.36	-0.55	-0.42	-0.11
0.11	0.16	0.23	0.32	0.46	0.66	0.48	0.10	-0.23	-0.36	-0.28	-0.42	-0.55	-0.36	0.14	0.71	1.00	0.71	0.14	-0.36	-0.55	-0.42
0.08	0.11	0.16	0.23	0.32	0.46	0.66	0.48	0.10	-0.23	-0.36	-0.11	-0.42	-0.55	-0.36	0.14	0.71	1.00	0.71	0.14	-0.36	-0.55
0.05	0.08	0.11	0.16	0.23	0.32	0.46	0.66	0.48	0.10	-0.23	0.17	-0.11	-0.42	-0.55	-0.36	0.14	0.71	1.00	0.71	0.14	-0.36
0.04	0.05	0.08	0.11	0.16	0.23	0.32	0.46	0.66	0.48	0.10	0.30	0.17	-0.11	-0.42	-0.55	-0.36	0.14	0.71	1.00	0.71	0.14
0.03	0.04	0.05	0.08	0.11	0.16	0.23	0.32	0.46	0.66	0.48	0.25	0.30	0.17	-0.11	-0.42	-0.55	-0.36	0.14	0.71	1.00	0.71
0.02	0.03	0.04	0.05	0.08	0.11	0.16	0.23	0.32	0.46	0.66	0.08	0.25	0.30	0.17	-0.11	-0.42	-0.55	-0.36	0.14	0.71	1.00

Correlation matrix for the original variables when $\rho_{a_1 a_2} \approx .98$.

1.00	-0.11	0.53	-0.08	0.28	-0.06	0.15	-0.04	0.08	-0.02	0.04	0.67	0.04	0.35	0.01	0.18	-0.01	0.10	-0.01	0.05	-0.01	0.03
-0.11	1.00	-0.11	0.53	-0.08	0.28	-0.06	0.15	-0.04	0.08	-0.02	0.15	0.67	0.04	0.35	0.01	0.18	-0.01	0.10	-0.01	0.05	-0.01
0.53	-0.11	1.00	-0.11	0.53	-0.08	0.28	-0.06	0.15	-0.04	0.08	0.17	0.15	0.67	0.04	0.35	0.01	0.18	-0.01	0.10	-0.01	0.05
-0.08	0.53	-0.11	1.00	-0.11	0.53	-0.08	0.28	-0.06	0.15	-0.04	-0.26	0.17	0.15	0.67	0.04	0.35	0.01	0.18	-0.01	0.10	-0.01
0.28	-0.08	0.53	-0.11	1.00	-0.11	0.53	-0.08	0.28	-0.06	0.15	-0.18	-0.26	0.17	0.15	0.67	0.04	0.35	0.01	0.18	-0.01	0.10
-0.06	0.28	-0.08	0.53	-0.11	1.00	-0.11	0.53	-0.08	0.28	-0.06	-0.24	-0.18	-0.26	0.17	0.15	0.67	0.04	0.35	0.01	0.18	-0.01
0.15	-0.06	0.28	-0.08	0.53	-0.11	1.00	-0.11	0.53	-0.08	0.28	-0.01	-0.24	-0.18	-0.26	0.17	0.15	0.67	0.04	0.35	0.01	0.18
-0.04	0.15	-0.06	0.28	-0.08	0.53	-0.11	1.00	-0.11	0.53	-0.08	0.05	-0.01	-0.24	-0.18	-0.26	0.17	0.15	0.67	0.04	0.35	0.01
0.08	-0.04	0.15	-0.06	0.28	-0.08	0.53	-0.11	1.00	-0.11	0.53	0.16	0.05	-0.01	-0.24	-0.18	-0.26	0.17	0.15	0.67	0.04	0.35
-0.02	0.08	-0.04	0.15	-0.06	0.28	-0.08	0.53	-0.11	1.00	-0.11	0.09	0.16	0.05	-0.01	-0.24	-0.18	-0.26	0.17	0.15	0.67	0.04
0.04	-0.02	0.08	-0.04	0.15	-0.06	0.28	-0.08	0.53	-0.11	1.00	0.05	0.09	0.16	0.05	-0.01	-0.24	-0.18	-0.26	0.17	0.15	0.67
0.67	0.15	0.17	-0.26	-0.18	-0.24	-0.01	0.05	0.16	0.09	0.05	1.00	0.51	0.16	-0.39	-0.45	-0.39	-0.06	0.15	0.28	0.20	0.07
0.04	0.67	0.15	0.17	-0.26	-0.18	-0.24	-0.01	0.05	0.16	0.09	0.51	1.00	0.51	0.16	-0.39	-0.45	-0.39	-0.06	0.15	0.28	0.20
0.15	0.04	0.67	0.15	0.17	-0.26	-0.18	-0.24	-0.01	0.05	0.16	0.16	0.51	1.00	0.51	0.16	-0.39	-0.45	-0.39	-0.06	0.15	0.28
0.01	0.35	0.04	0.67	0.15	0.17	-0.26	-0.18	-0.24	-0.01	0.05	-0.39	0.16	0.51	1.00	0.51	0.16	-0.39	-0.45	-0.39	-0.06	0.15
0.18	0.01	0.35	0.04	0.67	0.15	0.17	-0.26	-0.18	-0.24	-0.01	-0.45	-0.39	0.16	0.51	1.00	0.51	0.16	-0.39	-0.45	-0.39	-0.06
-0.01	0.18	0.01	0.35	0.04	0.67	0.15	0.17	-0.26	-0.18	-0.24	-0.39	-0.45	-0.39	0.16	0.51	1.00	0.51	0.16	-0.39	-0.45	-0.39
0.10	-0.01	0.18	0.01	0.35	0.04	0.67	0.15	0.17	-0.26	-0.18	-0.06	-0.39	-0.45	-0.39	0.16	0.51	1.00	0.51	0.16	-0.39	-0.45
-0.01	0.10	-0.01	0.18	0.01	0.35	0.04	0.67	0.15	0.17	-0.26	0.15	-0.06	-0.39	-0.45	-0.39	0.16	0.51	1.00	0.51	0.16	-0.39
0.05	-0.01	0.10	-0.01	0.18	0.01	0.35	0.04	0.67	0.15	0.17	0.28	0.15	-0.06	-0.39	-0.45	-0.39	0.16	0.51	1.00	0.51	0.16
-0.01	0.05	-0.01	0.10	-0.01	0.18	0.01	0.35	0.04	0.67	0.15	0.20	0.28	0.15	-0.06	-0.39	-0.45	-0.39	0.16	0.51	1.00	0.51
0.03	-0.01	0.05	-0.01	0.10	-0.01	0.18	0.01	0.35	0.04	0.67	0.07	0.20	0.28	0.15	-0.06	-0.39	-0.45	-0.39	0.16	0.51	1.00

Correlation matrix for the transformed variables when $\rho_{a_1 a_2} \approx .98$.

Some measures of multicollinearity for the two theoretical correlation matrices with $\rho_{a_1 a_2} \approx .98$.

	Original variables				Transformed variables		
i	λ_i	%	Σ%	i	λ_i	%	Σ%
1	5.367	24.4	24.4	1	4.008	18.2	18.2
2	5.047	22.9	47.3	2	3.547	16.1	34.3
3	4.545	20.7	68.0	3	3.464	15.7	50.1
4	2.799	12.7	80.7	4	2.721	12.4	62.5
5	1.239	5.6	86.3	5	2.104	9.6	72.0
6	0.757	3.4	89.8	6	1.660	7.5	79.6
7	0.569	2.6	92.4	7	1.131	5.1	84.7
8	0.402	1.8	94.2	8	0.844	3.8	88.5
9	0.309	1.4	95.6	9	0.689	3.1	91.7
10	0.255	1.2	96.8	10	0.544	2.5	94.1
11	0.224	1.0	97.8	11	0.516	2.3	96.5
12	0.206	0.9	98.7	12	0.430	2.0	98.4
13	0.194	0.9	99.6	13	0.244	1.1	99.6
14	0.031	0.1	99.7	14	0.021	0.1	99.6
15	0.029	0.1	99.9	15	0.019	0.1	99.7
16	0.014	0.1	99.9	16	0.018	0.1	99.8
17	0.006	0.0	100.0	17	0.011	0.0	99.9
18	0.003	0.0	100.0	18	0.010	0.0	99.9
19	0.002	0.0	100.0	19	0.006	0.0	99.9
20	0.001	0.0	100.0	20	0.006	0.0	100.0
21	0.001	0.0	100.0	21	0.004	0.0	100.0
22	0.001	0.0	100.0	22	0.004	0.0	100.0
	22.000	100.0			22.000	100.0	

Sum of variance inflation factors:

$\Sigma\ 1/\lambda_i = 4310.49$ $\qquad$ $\Sigma\ 1/\lambda_i = 1198.88$

The spectral condition number:

$\lambda_1/\lambda_{22} = 6611.98$ $\qquad$ $\lambda_1/\lambda_{22} = 994.06$

Determinant of the correlation matrix:

$|\mathbf{X'X}| = \lambda_1 \cdot \ldots \cdot \lambda_{22} = 2.17814 \cdot 10^{-23}$ $\qquad$ $|\mathbf{X'X}| = \lambda_1 \cdot \ldots \cdot \lambda_{22} = 3.93816 \cdot 10^{-18}$

Haitovskys test variable:

$\chi^2_H(231) = 0.000$ $\qquad$ $\chi^2_H(231) = 0.000$

Paper B

On Identification of Transfer Function Models by Biased Regression Methods [1]

[1] Reprint from *Journal of Statistical Computation and Simulation*, vol 31, 1989, pp. 131–148.

On Identification of Transfer Function Models by Biased Regression Methods

PER–OLOV EDLUND
Stockholm School of Economics
Box 6501
S–113 83 Stockholm, SWEDEN.

ABSTRACT

This paper investigates a biased regression approach to the preliminary estimation of the Box–Jenkins transfer function weights. Using statistical simulation to generate time series, 14 estimators (various *OLS*, ridge and principal components estimators) are compared in terms of *MSE* and standard error of the weight estimators.

The estimators are investigated for different levels of multicollinearity, signal–to–noise ratio, number of independent variables, length of time series and number of lags included in the estimation.

The results show that the ridge estimators nearly always give lower *MSE* than the *OLS* estimator, and in the computationally difficult cases give much lower *MSE* than the *OLS* estimator. The principal components estimators can give lower *MSE* than the *OLS*, but also higher values. All biased estimators nearly always give much lower estimated standard error than *OLS* when estimating the weights.

KEY WORDS: Time series, Transfer function models, Identification procedure, Biased regression, Monte Carlo.

INTRODUCTION

In many statistical models used in economics and business the dependent variable is a function of not only current values but also lagged values of the independent variables. These models occur in business cycle forecasting with leading indicators as well as in sales forecasting models, where sales is modeled as a function of present and past levels of advertising expenditure and present and past price levels.

In these cases the transfer function models described by Box and Jenkins (1976)

may be used to describe and identify dynamic relationships based on theoretical models. Since theoretical models very seldom imply the exact nature of the lag structure, there clearly is a need for an empirical method to identify these structures.

The method of identification proposed by Box and Jenkins seems to work well only when there is one independent variable or when the independent variables are mutually uncorrelated. In economic time series this is unlikely to be the case.

Therefore, an identification method is needed that works well even in the case of several cross–correlated independent variables. One such method is the multiple regression approach. Even though that approach was not recommended by Box and Jenkins, studies by Pukkila (1980) and Edlund (1984) have shown that the regression approach may be used with good results if the estimation problems can be solved. The regression method also has the advantage of being quite easy to use and it is also easy to find suitable computer software.

The aim of this paper is to further investigate a biased regression procedure proposed by Edlund (1984). The procedure will be studied by means of statistical simulation.

IDENTIFICATION OF TRANSFER FUNCTION MODELS

The transfer function–noise model proposed by Box and Jenkins (1976) is,

$$y_t = c + \frac{\omega(B)}{\delta(B)} x_{t-b} + n_t$$

where y_t is the dependent variable, c is a constant term, x_t is the independent variable, n_t is an error term which represents all the 'missing' x variables plus the pure noise, B is the ordinary lag operator, $\omega(B)$ is a 'moving average' operator, $\delta(B)$ is an 'autoregressive' operator, b is a pure delay parameter which represents the number of complete time intervals before a change in x_t begins to have an effect on y_t. (The y and x variables are assumed to be differenced/transformed to be mean and variance stationary.)

The transfer function $\nu(B)$ is a rational lag structure,

$$\nu(B) = \frac{\omega(B)}{\delta(B)} B^b$$

or

$$\nu(B) = (\nu_0 + \nu_1 B + \nu_2 B^2 + \ldots) B^b$$

that may be used to represent any form of linear dynamic relationship between x_t and y_t to any specified degree of accuracy. This formulation has also been used in econometrics, see e.g. Jorgenson (1966).

The noise term, n_t, may be expressed as an ordinary $ARMA(p,q)$ model of the form

$$n_t = \frac{\theta(B)}{\phi(B)} a_t$$

where $\theta(B)$ is a moving average operator of order q, $\phi(B)$ is an autoregressive operator of order p, and a_t a white noise variable.

The single–input model is easily generalized to an m–input transfer function model,

$$y_t = c + \sum_{j=1}^{m} \frac{\omega_j(B)}{\delta_j(B)} x_{j,t-b_j} + \frac{\theta(B)}{\phi(B)} a_t$$

with m input variables $x_{1t}, x_{2t}, \ldots, x_{mt}$. The ordinary multiple regression model

$$y_t = c + \nu_{10} x_{1t} + \nu_{20} x_{2t} + \ldots + \nu_{m0} x_{mt} + a_t$$

is a special case of the more general multiple–input transfer function model.

If the true transfer function $\nu(B)$ was known, suitable operators $\delta(B)$ and $\omega(B)$ and the pure delay parameter b could be identified. In practice the transfer function has to be estimated before the operators and delay parameter can be guessed. This preliminary estimation problem has led to different methods of identification. (The identification phase of the transfer function model estimation work also includes the guessing of operators and delay parameter. In this study only the preliminary estimation of the transfer function weights, ν_j, will be considered. Good preliminary estimation improves the conditions for reliable parameter identification.)

Box and Jenkins (1976, Chap. 11) discuss three methods for estimating the transfer function weights. Two of the methods are time domain methods, the prewhitening–cross correlation method and the regression method. The third method is a frequency domain method, the cross–spectral analysis method. They found that the regression method had several deficiencies and that the prewhitening–cross correlation method was to be preferred. This may be true in the case of a single input transfer function model, but for multiple input models with correlated input variables the prewhitening–cross correlation method is not applicable. The cross–spectral

analysis method may be used both for the single and the multiple input models, for an example see Pukkila (1979). This method, however, is more demanding in terms of knowledge in statistics and in computer time, than the regression method.

Other methods have been proposed or used in practical applications. Priestley (1971) proposed a method similar to Box and Jenkins prewhitening–cross correlation method, called the covariance contraction method. A similar method was also proposed by Haugh and Box (1977). These methods are also intended for the one–input model, even though an extension of Priestleys method to multiple–input models has been given by Fask and Robinson (1977). Apart from these more theoretical approaches one may try a stepwise procedure starting with too few or too many lags and then add or delete lags to obtain a good model. Of course there is no guarantee that this approach will give the best fitting model or the most parsimonious representation of the model.

THE REGRESSION METHOD

The most natural way of obtaining preliminary estimates of the transfer function weights, ν_j, is simply to estimate the weights by ordinary least–squares regression. The simple transfer function may be written as,

$$y_t = c + \nu_0 x_t + \nu_1 x_{t-1} + \nu_2 x_{t-2} + \ldots + n_t \qquad (1)$$

where y, x, c and n are defined as above. If it is reasonable to assume that the weights ν_j are ≈ 0 for some lag $k > K$ then (1) can be estimated by *OLS* with $K+1$ independent variables (lags 0 to K of the x variable). If there are m independent x variables equation (1) can be expanded accordingly. This regression approach was tested by Pukkila (1980) for a two–input and a three–input model. He found that the *OLS* estimators yielded surprisingly good results in terms of average bias and standard errors. In his study the input variables were only moderately cross correlated. In general we may encounter several problems trying to estimate (1). First of all, we need knowledge of the lag K at which the ν_j weights may safely be assumed to be zero. Trying large values of K means that many degrees of freedom will be lost. Secondly, the correlation matrix of the independent variables (including lagged variables) may show strong multicollinearity. This collinearity may result from autocorrelated input variables or/and cross correlation between the input variables. The third major problem is the possibility of autocorrelated residuals, i.e. n_t is not white noise. This

will decrease the efficiency of the regression estimates.

The first problem can usually be solved by finding some trade–off between the risk of including too few lags (gives biased estimates) and losing too many degrees of freedom. It seems wise to start with too many rather than too few lags in the model. The multicollinearity problem is more difficult to solve. Liu and Hanssens (1982) transformed the x and y variables by a common filter. The filter was constructed to eliminate AR factors with roots close to one in the estimated ARMA processes for the x variables. Edlund (1984) used biased regression, ridge and principal components regression, to obtain more accurate estimates. Autocorrelated residuals can be dealt with by using generalized least squares (*GLS*) instead of *OLS* (see Liu and Hanssens) or by transforming the x and y variables. The latter method has been used by Box and Tiao (1975) and by Edlund (1984).

In Edlund (1984) the following two–step procedure was suggested and tested in a simulation study:

Step one: **identification, estimation and checking of the noise model and transformation of the input and output variables**

First the multiple regression model

$$y_t = c + \nu_{10}x_{1t} + \nu_{11}x_{1,t-1} + \ldots + \nu_{mK}x_{m,t-K} + n_t \qquad (2)$$

is estimated by biased regression, ridge or principal components regression. Then the estimated residuals are computed,

$$\hat{n}_t = y_t - \hat{c} - \sum_{i=1}^{m} \hat{\nu}_i(B)x_{it} = y_t - \hat{c} - \hat{\nu}_{10}x_{1t} - \hat{\nu}_{11}x_{1,t-1} - \ldots - \hat{\nu}_{mK}x_{m,t-K} \ .$$

The noise model, $\hat{n}_t = \dfrac{\theta(B)}{\phi(B)}\hat{a}_t$ is then identified and estimated using the standard Box–Jenkins procedure for ARMA models. The estimated operators are then used to transform the original x and y variables,

$$\hat{\theta}(B)y'_t = \hat{\phi}(B)y_t, \text{ all } t, \text{ and } \hat{\theta}(B)x'_{jt} = \hat{\phi}(B)x_{jt}\ , \ j = 1,\ldots,m, \text{ all } t.$$

Step two: **estimation of the impulse response function from the transformed variables y'_t and x'_{jt}**

In the second step equation (2) is re–estimated, by biased regression methods, using the transformed x and y variables,

$$y'_t = c + \nu_{10}x'_{1t} + \nu_{11}x'_{1,t-1} + \cdots + \nu_{mK}x'_{m,t-K} + a_t . \qquad (3)$$

In (3) the residuals $\{a_t\}$ should be almost white noise and by using biased regression the bad effects of multicollinearity will be decreased. If so, the estimates $\hat{\nu}_{ij}$ will be good estimates of the true values and could be used for guessing the orders of the operators in the transfer functions. The significance of the $\hat{\nu}_{ij}$ can be tested by the standard t–test. In case the estimated residuals in (3) are not white noise, step one could be repeated using the estimated values of ν_{ij} in (3) for calculating the residuals n_t. Step two is then performed again. In the end acceptable estimates of ν_{ij} will be obtained.

In order to test the proposed procedure a small simulation experiment was performed in Edlund (1984). It showed that the proposed procedure may be of value to the practitioner, especially when the multicollinearity is strong and when the signal–to–noise ratio is low (large residual variance). The ridge estimator was nearly always superior to the principal components estimator.

METHODOLOGY

In order to further investigate the properties of the two–step procedure in Edlund (1984) an extended Monte Carlo study has been carried out. This study covers more factors and estimators but also some modifications of the estimation procedure. The design of the study is described below.

Models

In the simulations three basic models have been studied. The two– and three–input models are from Pukkila (1979). They have also been used by Pukkila (1980), and the two input model has also been used by Damsleth (1979) and Edlund (1984). In addition to these models a single–input model was chosen. The models are nested, the two–input model contains the one–input model, and the three–input model

contains the two–input model. The three models exemplify different degrees of complexity. In the one–input case the results can be directly compared to those from the ordinary pre–whitening cross correlation method proposed by Box and Jenkins. The models are,

Model 1:

$$y_t = (2 + 4B + B^2)x_{1t} + n_t$$

Model 2:

$$y_t = (2 + 4B + B^2)x_{1t} + (1 - 0.6B)^{-1}B^3x_{2t} + n_t$$

Model 3:

$$y_t = (2 + 4B + B^2)x_{1t} + (1 - 0.6B)^{-1}B^3x_{2t} + 2(1 - 0.8B)^{-1}B^5x_{3t} + n_t$$

where,

$$\begin{aligned} x_{1t} &= 0.7x_{1,t-1} + a_{1t} \\ x_{2t} &= 1.25x_{2,t-1} - 0.75x_{2,t-2} + a_{2t} \\ x_{3t} &= a_{3t} + 0.8a_{3,t-1} \\ n_t &= a_t + 0.75a_{t-1} \end{aligned}$$

The processes a_{1t}, a_{2t}, a_{3t} and a_t are generated as white noise processes such that a_t is uncorrelated with a_{1t}, a_{2t} and a_{3t}, and a_{1t}, a_{2t} and a_{3t} are correlated to give different strengths of correlation between the x variables.

Criteria for comparing estimators

The performance of the estimators can be measured by two main categories of statistics, the first measuring the closeness of the estimated coefficients to the true values, the second measuring the forecasting performance of the model. Since the object of the identification phase is to identify the "true" coefficients and a suitable transfer function for them, the first category is of major interest. Therefore different measures were computed describing the *MSE* of the coefficients for the estimators and the standard error of the estimated coefficients. All mean value statistics were averaged over all coefficients and over all replications.

Selected estimators

The principal component estimator used in Edlund (1984) performed very poorly in *MSE* when the signal–to–noise ratio was high. Therefore, in this simulation study two other selection criteria for the principal components were used. The first criterion was proposed and tested by Lott (1973). He developed a criterion which maximizes the $\bar{R}^2$ for the regression of y_t on the principal components. The second criterion used is a simple t–test of the regression coefficients in the principal component regression, see e.g. Massy (1965). Components with a non–significant regression coefficient at the 5 % level were deleted from the principal components equation. As a standard of comparison the principal component estimator from Edlund (1984) was computed.

The shrinkage factor k in the ridge estimator has been computed by three different methods proposed by; Lawless and Wang (1976), Hocking, Speed and Lynn (1976) and Hoerl, Kennard and Baldwin (1975). The estimator by Lawless and Wang was used in the previous study and performed well. In addition to that estimator two other estimators were used in order to see if further reduction in *MSE* is possible. The estimators are all functions of the estimated regression coefficients and the residual standard deviation. The value of k is computed by the following formulas,

Lawless and Wang: $$k = \frac{p\hat{\sigma}^2}{\Sigma\lambda_i\hat{\alpha}_i^2}$$

Hocking, Speed and Lynn: $$k = \hat{\sigma}^2 \frac{\Sigma\lambda_i^2\hat{\alpha}_i^2}{\Sigma\lambda_i^2\hat{\alpha}_i^4}$$

Hoerl, Kennard and Baldwin: $$k = \frac{p\hat{\sigma}^2}{\Sigma\hat{\alpha}_i^2}$$

where p is the number of independent variables, $\hat{\sigma}^2$ is the estimated residual variance, λ_i is the i:th eigenvalue from the correlation matrix and $\hat{\alpha}_i$ is the *OLS* estimate of the i:th regression coefficient from regression on variables transformed by the eigenvectors computed from the correlation matrix of the x variables.

Experimental design

In this study the effects of some relevant factors on the properties of the different estimators have been analyzed.

The factors considered in this simulation study have been the degree of multicollinearity, the signal–to–noise ratio (residual variance), number of input variables, length of time series and number of lags. Since it is computationally very tedious to investigate every possible combination of levels on these factors, only two factors have been varied at a time while the other factors have been set to their default levels.

The degree of multicollinearity can be varied by either changing the autocorrelation structure for the input variables or by changing the cross correlation between the input variables. In this study only the latter method have been used. For model 1 the variance of a_{1t} has been set to 2, 3 and 4 respectively with 3 as a default value. For model 2 the following covariance matrices have been used to generate the time series,

$$\Sigma_1 = \begin{bmatrix} 2 & \sqrt{3} \\ \sqrt{3} & 4 \end{bmatrix} \quad \Sigma_2 = \begin{bmatrix} 2 & \sqrt{5} \\ \sqrt{5} & 4 \end{bmatrix} \quad \Sigma_3 = \begin{bmatrix} 2 & \sqrt{7} \\ \sqrt{7} & 4 \end{bmatrix}$$

and for model 3,

$$\Sigma_1 = \begin{bmatrix} 4 & 2 & 2 \\ 2 & 5 & 1 \\ 2 & 1 & 5 \end{bmatrix} \quad \Sigma_2 = \begin{bmatrix} 4 & 3 & 3 \\ 3 & 5 & 2 \\ 3 & 2 & 5 \end{bmatrix} \quad \Sigma_3 = \begin{bmatrix} 4 & 4 & 3 \\ 4 & 5 & 3 \\ 3 & 3 & 5 \end{bmatrix}$$

where the first Σ matrix in model 2 and 3 is equal to the covariance matrix used by Pukkila (1979, 1980). The second Σ matrix has been used as a default.

The signal–to–noise ratio has been varied by changing the value of the residual variance, σ_a^2. The variance was set to 0.01, 1 and 25 respectively, with $\sigma_a^2 = 1$ as default.

The number of input variables have been altered by using the three different models above, with model 2 as the default model.

The length of time series has been set to 50, 100 and 200 observations respectively, with 100 observations as the default length. Model 1 could be estimated with less than 50 observations but to obtain meaningful estimates, especially with long lags, the number of observations must not be too small. Model 3 can not be estimated with 50 observations unless the lags are rather limited.

The last factor considered is the number of lags included in the model. Here different sets have been used to obtain cases where there are too few, a correct

number or too many lags included. This reflects different prior beliefs about the lag structure. The default values are lags 0 to 10 for x_1 and x_2 and lags 0 to 20 for x_3.

The last three factors, number of variables, number of observations and number of lags determine the degrees of freedom in each estimated model.

For all investigated combinations the number of replications has been 50. During each replication (for all combinations) the time series for variable x_1 has been held constant. Variable x_2 and x_3 have been recomputed when the covariance matrix has changed, otherwise they have also been kept constant. The main source of stochastic variation in this study has therefore been the residual series that has been different in each replication (but constant between different combinations, apart from variations in the residual variance). Thereby it has been possible to isolate the effects of the disturbance term on the estimation method.

The processes a_t, a_{1t}, a_{2t} and a_{3t} have been generated as normal white noise processes using the Box and Muller (1958) method. The input variables have been computed from the a_{it} processes using a starting period of 50 observations. These first 50 observations have not been used in the estimations. The x and y variables have been transformed by the true noise model rather than the estimated model. This has been done to obtain a significant reduction in computing time. To investigate the effects of this simplification, checks have been done for some factor combinations without observing any significant differences in terms of the relative performance of different estimators. To obtain a starting value for the transformations the original series have been forecasted backwards using Box and Jenkins technique (1976, Ch.7).

RESULTS OF SIMULATION

The following measures have been computed for all investigated combinations of factors and for all estimators:

– $MSE(\hat{\nu}_m)_i = \sum_{j=1}^{\ell} (\hat{\nu}_{ijm} - \nu_j)^2/\ell \quad \overline{MSE(\hat{\nu}_m)} = \sum_{i=1}^{50} MSE(\hat{\nu}_m)_i/50$

where $\hat{\nu}_{ijm}$ is an estimate of the j:th transfer function weight in replication i for estimator m, 50 is the number of replications, ℓ is the total number of estimated transfer function weights. The *MSE* may be split into a bias term and a variance term. In the simulations the bias term has been small relative to the variance term. Therefore only the total *MSE* has been given.

- $\%[MSE(\hat{\nu}_m) \leq MSE(\hat{\nu}_{OLS})]$ is the proportion of times the average *MSE* for estimator m is not larger than the *MSE* for *OLS*. This measure may be used instead of testing for significant differences in *MSE* between the *OLS* and the other estimators.
- $\bar{S}(\hat{\nu}_m) = \sum_{i=1}^{50} \sum_{j=1}^{\ell} s_{ijm}/50 \cdot \ell$

  where s_{ijm} is an estimate of the standard error for estimator m when estimating weight j in replication i.
- Average number of principal components (PC) used in estimation (for the principal components estimators). This and the next measure were computed to study the differences between the PC and ridge estimators.
- Average value of the shrinkage factor, k (for the ridge estimators)
- Average standard error of estimate (residual standard deviation). This measure shows if the estimate is unbiased.

To simplify the evaluation of the 14 estimators the following ratios were calculated

$$MSE_m = \frac{\overline{MSE(\hat{\nu}_m)}}{\overline{MSE(\hat{\nu}_{OLS})}} \qquad S_m = \frac{\bar{S}(\hat{\nu}_m)}{\bar{S}(\hat{\nu}_{OLS})}$$

where m = *OLS*, *RR LW* (Lawless and Wang), *RR HSL* (Hocking, Speed and Lynn), *RR HKB* (Hoerl, Kennard and Baldwin), *PC* Σ (the h first principal components corresponding to the h largest eigenvalues covering at least 99.5 % of the variance in the correlation matrix), *PC* t>2 (principal components with a t–value larger than 2), *PC MAX* (Lott's maximum $\bar{R}^2$ criterion). *OLSt*, *RRt LW*, *RRt HSL*, *RRt HKB*, *PCt* Σ, *PCt* t>2, *PCt MAX* are estimators based on transformed variables.

In all, 10 different combinations of the five factors considered have been evaluated for all estimators and for all measures above. Since the number of tables clearly exceeds the space allowed in a journal only one set of tables will be given. The contents of the other tables will be summarized in the text. A complete set of tables can be obtained from the author.

Degree of multicollinearity

Tables I–VI show the results for different combinations of multicollinearity (different Σ:s here denoted by the correlation, ρ, between a_1 and a_2) and

Estimator	σ_a = .1 ρ=.61	ρ=.79	ρ=.94	σ_a = 1 ρ=.61	ρ=.79	ρ=.94	σ_a = 5 ρ=.61	ρ=.79	ρ=.94
OLS	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
RR LW	1.00	1.00	1.00	0.93	0.89	0.74	0.49	0.37	0.18
RR HSL	1.00	1.00	1.00	1.00	0.99	0.98	0.92	0.88	0.73
RR HKB	1.00	1.00	0.99	0.91	0.84	0.64	0.49	0.38	0.24
PC Σ	45.55	67.93	66.54	1.40	1.56	1.04	0.68	0.51	0.16
PC $t>2$	0.97	1.34	2.21	1.42	1.37	0.87	0.91	0.57	0.42
PC MAX	0.96	1.24	1.73	1.32	1.26	0.91	0.91	0.58	0.47
OLSt	0.71	0.74	0.80	0.83	0.82	0.81	0.83	0.82	0.81
RRt LW	0.71	0.75	0.80	0.81	0.79	0.69	0.50	0.38	0.19
RRt HSL	0.71	0.74	0.80	0.83	0.82	0.80	0.79	0.76	0.65
RRt HKB	0.71	0.75	0.80	0.82	0.79	0.70	0.53	0.42	0.25
PCt Σ	73.18	91.10	66.25	1.92	2.04	1.16	0.75	0.61	0.29
PCt $t>2$	0.73	0.80	0.83	1.11	1.22	1.15	0.83	0.65	0.51
PCt MAX	0.75	0.93	0.80	0.95	1.02	0.99	0.81	0.65	0.56
$\overline{MSE}(\hat{\nu}_{OLS})$	0.0002	0.0004	0.0010	0.015	0.023	0.065	0.360	0.576	1.625

Table I. Estimated ratios, MSE_m, of $\overline{MSE}(\hat{\nu}_m)$ to $\overline{MSE}(\nu_{OLS})$. Correlation ($\rho$) between a_1 and a_2 vs. residual standard deviation (σ_a). Model 2 with 100 observations, lag 0–10 for x_1 and x_2.

signal–to–noise ratio (different σ_a values).

From Table I it may be noted that:

- The transformation of variables in this case nearly always leads to a reduction in *MSE* for the ridge estimators. The relative *MSE* (MSE_m) for the ridge estimators tends to decrease as ρ increases, except when $\sigma_a = .1$. In these cases the *MSE* is very small compared to the size of the weights. Then there is very little to gain from using ridge regression and the gain decreases rather than increases as ρ increases.
- The performances of the *LW* and *HKB* estimators are very similar and in most cases these estimators are superior to the *HSL* estimator. The *HSL* estimator usually is more similar to the corresponding *OLS* estimator. The ridge estimators give lower *MSE* than the *OLS* estimators, especially in computationally difficult cases (when multicollinearity is strong and signal–to–noise ratio is low, see e.g. when $\rho = .94$ and $\sigma_a = 5$), which is to be expected.
- The principal components estimators in some cases perform much better than the *OLS* estimators, but in some cases much worse. There are no clear responses to increases in ρ except when the signal–to–noise ratio is low ($\sigma_a = 5$) where the MSE_m for all estimators decreases as ρ increases. When σ_a is small relatively few

components are deleted (see Table IV) and we may therefore suspect that there are some random effects, on the margin, in choosing which components to delete.

- The *PC t>2* and *PC MAX* estimators have a more stable performance than the *PC* Σ estimator. On the other hand the *PC* Σ estimator is clearly better than the other estimators when the signal to noise ratio is low. These effects are due to the insensitiveness of the *PC* Σ estimator to changes in the residual variance, see Table IV.

	σ_a = .1			σ_a = 1			σ_a = 5		
Estimator	ρ=.61	ρ=.79	ρ=.94	ρ=.61	ρ=.79	ρ=.94	ρ=.61	ρ=.79	ρ=.94
RR LW	0.54	0.58	0.60	0.74	0.94	1.00	1.00	1.00	1.00
RR HSL	0.54	0.58	0.60	0.88	0.94	1.00	1.00	1.00	1.00
RR HKB	0.54	0.58	0.58	0.68	0.86	0.98	1.00	1.00	1.00
PC Σ	0.00	0.00	0.00	0.12	0.06	0.34	0.96	0.98	1.00
PC t>2	0.46	0.14	0.00	0.14	0.24	0.68	0.56	0.96	1.00
PC MAX	0.46	0.16	0.06	0.22	0.26	0.64	0.62	0.96	1.00
OLSt	0.84	0.78	0.62	0.68	0.68	0.70	0.66	0.66	0.64
RRt LW	0.84	0.76	0.62	0.68	0.68	0.88	0.98	1.00	1.00
RRt HSL	0.84	0.78	0.62	0.68	0.68	0.70	0.78	0.78	0.94
RRt HKB	0.84	0.76	0.62	0.68	0.64	0.88	0.98	1.00	1.00
PCt Σ	0.00	0.00	0.00	0.04	0.04	0.32	0.78	0.90	1.00
PCt t>2	0.82	0.66	0.64	0.44	0.36	0.34	0.68	0.90	0.94
PCt MAX	0.80	0.48	0.64	0.58	0.46	0.48	0.66	0.90	0.88

Table II. Proportion of iterations where the *MSE* for an estimator is less than or equal to the *MSE* for the *OLS* estimator.

Table II shows that the ridge estimators have lower (or equal) *MSE* values than the *OLS* estimator in more than 50 % of the cases. There seems to be a tendency for the proportion to increase as ρ increases. The pattern is less clear when σ_a is small which is consistent with the MSE_m results above. As pointed out above, the performance of the *PC* estimators is more erratic, in some cases the proportion is close to one and in other cases close to zero.

Estimator	σ_a = .1 ρ=.61	ρ=.79	ρ=.94	σ_a = 1 ρ=.61	ρ=.79	ρ=.94	σ_a = 5 ρ=.61	ρ=.79	ρ=.94
OLS	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
RR LW	1.00	1.00	1.00	0.91	0.87	0.74	0.50	0.42	0.27
RR HSL	1.00	1.00	1.00	1.00	0.99	0.98	0.92	0.88	0.74
RR HKB	1.00	1.00	0.99	0.87	0.80	0.61	0.48	0.41	0.31
PC Σ	1.44	1.56	0.97	0.61	0.49	0.28	0.60	0.47	0.26
PC $t>2$	0.96	0.94	0.65	0.59	0.47	0.33	0.29	0.23	0.18
PC MAX	0.96	0.96	0.79	0.63	0.51	0.34	0.31	0.24	0.19
OLSt	0.69	0.69	0.70	0.66	0.65	0.64	0.66	0.65	0.64
RRt LW	0.69	0.69	0.70	0.64	0.62	0.57	0.43	0.36	0.22
RRt HSL	0.69	0.69	0.70	0.66	0.65	0.64	0.65	0.63	0.57
RRt HKB	0.69	0.69	0.70	0.65	0.62	0.56	0.47	0.41	0.30
PCt Σ	2.30	2.29	1.28	0.67	0.60	0.39	0.62	0.55	0.37
PCt $t>2$	0.68	0.69	0.66	0.58	0.53	0.42	0.34	0.29	0.21
PCt MAX	0.68	0.70	0.67	0.60	0.56	0.47	0.37	0.32	0.25
$S(\hat{\nu}_{OLS})$	0.018	0.021	0.034	0.155	0.195	0.326	0.771	0.974	1.630

Table III. Estimated ratios, S_m, of the standard error of transfer function weights.

The relative standard errors in Table III generally decreases as ρ increases which is also implied by the use of biased regression. The biased estimators will perform relatively better as the multicollinearity increases. This is important since it becomes increasingly more difficult to identify significant weights when the multicollinearity is strong. The standard errors for the ridge estimators decreases after the transformation. This is to be expected as the estimate of σ_a is biased upwards when the variables are not transformed.

Estimator	σ_a = .1 ρ=.61	ρ=.79	ρ=.94	σ_a = 1 ρ=.61	ρ=.79	ρ=.94	σ_a = 5 ρ=.61	ρ=.79	ρ=.94
Σ	18.00	17.00	15.00	18.00	17.00	15.00	18.00	17.00	15.00
$t>2$	20.98	20.42	17.98	15.98	15.04	14.84	9.12	8.28	7.28
MAX	21.00	20.78	19.26	16.88	15.62	15.04	9.60	8.28	7.64
Σ	21.00	20.00	17.00	21.00	20.00	17.00	21.00	20.00	17.00
$t>2$	21.04	21.62	20.32	18.88	18.04	17.12	10.80	10.18	9.82
MAX	21.00	21.00	20.80	19.46	18.94	17.94	11.96	11.04	10.76

Table IV. Average number of principal components

In Table IV the number of *PC*:s decreases as ρ increases. More components are used after transformation. This may be expected since the multicollinearity in this case decreases after transformation and more components are retained as the multicollinearity decreases.

	σ_a = .1			σ_a = 1			σ_a = 5		
Estimator	ρ=.61	ρ=.79	ρ=.94	ρ=.61	ρ=.79	ρ=.94	ρ=.61	ρ=.79	ρ=.94
LW	.00003	.00002	.00002	.00225	.00202	.00180	.05470	.04942	.04408
HSL	.00000	.00000	.00000	.00010	.00009	.00008	.00211	.00194	.00175
HKB	.00005	.00005	.00005	.00386	.00396	.00386	.06330	.05453	.03076
LW	.00006	.00005	.00004	.00408	.00372	.00335	.09615	.08808	.07973
HSL	.00000	.00000	.00000	.00018	.00017	.00015	.00334	.00312	.00289
HKB	.00005	.00005	.00005	.00354	.00359	.00355	.06865	.06170	.04053

Table V. Average value of the shrinkage factor, k.

The k value in Table V for the *LW* and *HSL* estimators decreases as ρ increases and are larger after transformation. This is also expected since when the multicollinearity decreases a smaller value of k is needed to reduce the *MSE* of the weights. The *HKB* k value is rather unchanged as ρ increases, except when the signal–to–noise ratio is low. The reason for this is that the *HKB* k value is computed without respect to the eigenvalues (or multicollinearity). For low values of σ_a the estimates of $\hat{\alpha}_i$ (the *OLS* estimate of the i:th regression coefficient from regression on variables transformed by the eigenvectors computed from the correlation matrix of the x variables) are relatively unaffected by variations in ρ and therefore the k value is more or less constant. When the multicollinearity is strong the $\hat{\alpha}_i$:s are affected and hence the k value.

Table VI shows that the degree of multicollinearity does not affect the estimate. The estimates for regression on untransformed variables are biased, as should be expected. The size of the bias is determined by the true model for the noise process. In this case we would expect the estimates from untransformed variables to be 25 % larger than the true values. The bias is not to far from the expected value. It is interesting to note that there is bias in the estimates even for transformed variables when σ_a is small.

		Original variables	Transformed variables
ρ = 0.61	σ_a = .1	0.137	0.118
ρ = 0.79	σ_a = .1	0.133	0.116
ρ = 0.94	σ_a = .1	0.126	0.113
ρ = 0.61	σ_a = 1	1.210	0.997
ρ = 0.79	σ_a = 1	1.211	0.998
ρ = 0.94	σ_a = 1	1.213	0.999
ρ = 0.61	σ_a = 5	6.040	4.972
ρ = 0.79	σ_a = 5	6.048	4.978
ρ = 0.94	σ_a = 5	6.061	4.988

Table VI. Average standard error of the estimate of σ_a.

For model 3 the MSE_m decreases as the multicollinearity increases for the ridge estimators on untransformed data. After transformation there is no clear pattern. The same is true for the *PC* estimators. The S_m decreases as the multicollinearity increases for all estimators. The standard errors of the ridge estimates are lower after transformation while the s.e. of the *PC* estimators increases. The other measures show roughly the same patterns as above.

The pattern of the MSE_m for different number of observations, n, is as expected. As the multicollinearity increases the MSE_m decreases independently of n (with a few exceptions for the *PC* estimators). The S_m patterns are the same for all n and the same as above. The number of *PC*:s, k and $\hat{\sigma}_a$ all show the same pattern as in the tables above.

The number of lags included, ℓ, seems to be an important factor. The main results are:

- When there are too few lags included, $\ell = 5$, MSE_m does not change when ρ is increased for the untransformed variables. After transformation MSE_m decreases as ρ increases (with a few exceptions). When the number of lags is correct, $\ell = 11$, the same is true for the untransformed variables, but after transformation the MSE_m increases as ρ increases. When ℓ is too large ($\ell = 22$) MSE_m decreases as ρ increases both before and after transformation for the *RR* estimators. The unpredicted behavior of MSE_m for $\ell = 5$ underlines the danger of omitting relevant variables in regression analysis. When too may lags are included the MSE_m behaves as predicted.
- The S_m seems to be independent of ρ when $\ell = 5$ or 11 but decreasing as ρ in-

creases when $\ell = 22$. The S_m is generally lower after transformation (except for the PC estimators when $\ell = 22$).

- The number of PC:s increases after transformation but show no specific pattern when $\ell = 5$ or 11. When $\ell = 22$ the number of PC:s decreases as ρ increases.
- The effects on k of increasing ρ is about the same as above for all estimators irrespective of ℓ.
- The standard error of estimate, $\hat{\sigma}_a$, is heavily biased when $\ell = 5$ even for transformed variables and independently of ρ. This is well known from ordinary regression analysis when relevant x variables are missing.

Signal–to–noise ratio

The signal–to–noise ratio has been varied by altering the residual standard deviation, σ_a; .1, 1 and 5. Since the true weights range from almost zero to four setting $\sigma_a = .1$ corresponds to a high signal–to–noise ratio and $\sigma_a = 5$ to a low ratio. The results may be summarized as follows:

- For untransformed variables the MSE_m decreases as σ_a increases. After transformation the pattern is less clear. However, in most cases the MSE_m is lowest when $\sigma_a = 5$.
- The S_m ratio decreases as σ_a increases.
- For the t>2 and MAX estimators the number of PC:s decreases as σ_a increases. The number of PC:s increase after transformation.
- The shrinkage factor, k, increases as σ_a increases.
- The estimates of σ_a are heavily biased when $\sigma_a = .1$, even after transformation. The estimates after transformation are unbiased for $\sigma_a = 1$ and 5.

There are some exceptions to the general expected results above. For model 3 MSE_m after transformation is lowest for $\sigma_a = .1$. Also the S_m ratio shows an irregular pattern after transformation. For model 1 the bias of $\hat{\sigma}_a$ is not increased when $\sigma_a = .1$. This is not so strange since the estimates has many more degrees of freedom than for model 2 and 3.

When ℓ is correct the MSE_m has its minimum when $\sigma_a = .1$. When ℓ is too small the estimates of σ_a are heavily biased when $\sigma_a = .1$ and 1.

Length of time series

The following general conclusions can be drawn regarding the effects of varying the length of the time series:

- As n is increased, the MSE_m for the ridge estimators is also increased. This means that the benefit from using ridge regression is largest when the time series are short. The *LW* and *HKB* estimators have a similar performance and are better than the *HSL* estimator. The ridge estimators are never worse than the corresponding *OLS* estimator.
- In some cases the *PC* estimators are much better than the *OLS* estimator, especially when $n = 50$.
- The S_m increases as n increases. This is also expected since the benefit from using biased regression should be largest when it is most difficult to estimate the weights.
- The number of PC:s is increased when the length of the time series increases.
- The k value decreases as n increases.
- The estimate of residual standard deviation is practically independent of n.

There are a few exceptions to these results that should be pointed out:

- When $\sigma_a = .1$ MSE_m decreases as n increases for all estimators but the $PC\,\Sigma$ estimator. In these cases the $\overline{MSE(\hat{v}_m)}$ is very small (.0002 – .0011). The MSE_m values for the $PC\,\Sigma$ estimator ranges between 16 and 104!
- Even the S_m behave differently when $\sigma_a = .1$. For the original variables there is no clear pattern, and for the transformed variables the *OLS* and the ridge estimators S_m decrease as n increases.
- For the original variables the MSE_m:s behave as expected except when $\ell = 5$ and for the *PC* estimators. After transformation there is no clear pattern, even though the MSE_m:s are lowest when $\ell = 5$.

Number of lags

Three different cases have been investigated; too few lags, right number of lags and too many (for model 2 this means $\ell = 5$, 11 and 22 lags respectively). The following results were obtained:

- For the untransformed ridge estimators, MSE_m for $\ell = 5$ and 11 were about the same. With too may lags the MSE_m was lower. After transformation this pattern disappeared. In many cases MSE_m was lowest when too few lags were

estimated.

- For the *PC* estimators no clear pattern could be found. For the untransformed variables the t>2 and *MAX* estimators in many cases showed the largest MSE_m (larger than 1) when the right number of weights were estimated. After transformation the largest MSE_m (larger than 1) was found when too many weights were estimated.
- Before transformation too few or the right number of lags offer no significant improvement over *OLS* while too many lags gives a substantial reduction in MSE_m. After transformation the MSE_m:s are reduced. Too many lags still gives the lowest MSE_m but the differences are smaller.
- The number of *PC*:s relative to ℓ shows no clear pattern.
- The shrinkage factor k is smallest when the right number of weights are estimated, except when the signal–to–noise ratio is low ($\sigma_a = 5$). In most cases the k value is largest when too many lags are included for the *LW* and *HKB* estimators. The *HSL* estimator show low values when too few weights are estimated.
- When the number of lags is right or too large the estimated standard error of estimate is unbiased for transformed variables. The estimates on untransformed variables show the expected bias. However when too few lags are included the estimates are severely biased even after transformation.

For models 1 and 3 there are some exceptions to the results above. For the untransformed variables the ridge estimators have MSE_m:s close to one in model 1 and for model 3 MSE_m is lowest when too few lags are included. The estimated S_m is lowest when the lags are too few for both models 1 and 3.

CONCLUSIONS

From the simulation study it seems quite clear that the proposed identification procedure decreases the average *MSE* and the average standard error of the estimated weights as well as giving unbiased estimates of the residual standard deviation. The ridge estimators will in most cases give lower *MSE* than the corresponding *OLS* estimator. The average *MSE* is in all but a few cases lower than the *OLS* estimator. Of the ridge estimators the Lawless and Wang (*LW*) and the Hoerl, Kennard and Baldwin (*HKB*) estimators nearly always perform better than the Hocking, Speed and Lynn (*HSL*) estimator. The main reason for this is that the *HSL* estimator gives a

very small k value. This means that the *HSL* estimates will be very close to the *OLS* estimates. Of the *LW* and *HKB* estimators it is difficult to choose the best since they are very similar and neither of them is constantly superior to the other.

The principal components (*PC*) estimators are more erratic. In some cases the *PC* estimators are much better than the *OLS* estimators, and in some cases even better than the ridge estimators. The two new estimators, the "choose all significant *PC*:s" (t>2) and Lott's maximum R^2 criterion (*MAX*) give more stable results than the old "choose h *PC*:s corresponding to the h largest eigenvalues" (Σ) estimator. However, the new estimators will in some cases give very poor estimates and can thus not be recommended. The poor performance of the Σ estimator is due to its lack of adjustment to the level of the residual variance.

The transformation procedure seems to work well, in most cases the estimated standard error of the weights will be smaller, but we will also obtain an unbiased estimate of σ_a. In some cases, depending on the interaction between the noise model and the model for the input variables, the multicollinearity may increase after transformation. Since an unbiased estimate of σ_a is very important it is still suggested that the variables are transformed by the estimated noise model.

An important question is what number of lags to include for each input variable. The simulation results confirm results from ordinary regression analysis. It is obviously more dangerous to include too few than too many lags. The proposed estimation procedure will give reasonable estimates of the weights even if we include twice as many lags as is needed. Therefore it is good strategy to start with too many weights and by deleting non–significant weights reducing the number of lags in estimation.

By altering the length of the time series and by changing the number of variables it has been shown that the estimation procedure will give reasonable estimates in all but a few cases and that the reduction in MSE and standard error of the weights are very significant in the computationally difficult cases.

Acknowledgement

This research has been supported by grants from The Swedish Research Council for Humanities and Social Sciences (HSFR) and The Bank of Sweden Tercentenary Foundation (Riksbankens Jubileumsfond).

References

Box, G.E.P. and Jenkins, G.M., (1976), Time series analysis: Forecasting and control, 2:nd ed., Holden–Day, San Francisco.

Box, G.E.P. and Muller, M.E., (1958), A note on the generation of random normal deviates, *Annals of Mathematical Statistics* **29,** 610–611.

Box, G.E.P. and Tiao, G.C., (1975), Intervention analysis with applications to economic and environmental problems, *JASA* **70,** 70–79.

Damsleth, E., (1979), Analysis of multi–input transfer function models when the inputs are correlated, Norwegian Computing Center, Oslo, publ. no 641.

Edlund, P.–O., (1984), Identification of the multi–input Box–Jenkins transfer function model, *Journal of Forecasting* **3,** 297–308.

Fask, A. and Robinson, P.B., (1977), Identification of multivariate dynamic econometric models, *American Statistical Association 1977* Proceedings of the Business and Economic Statistics Section, 284–289.

Haugh, L.D. and Box, G.E.P., (1977), Identification of dynamic regression (distributed lag) models connecting two time series, *JASA* **72,** 121–130.

Hocking, R.R., Speed, F.M. and Lynn, M.J., (1976), A class of biased estimators in linear regression, *Technometrics* **18,** 425–437.

Hoerl, A.E., Kennard, R.W. and Baldwin, K.F., (1975), Ridge regression: some simulations, *Comm. in Statistics* **4,** 105–123.

Jorgenson, D.W., (1966), Rational distributed lag functions, *Econometrica* **32,** 135–148.

Lawless, J.F. and Wang, P., (1976), A simulation study of ridge and other regression estimators, *Comm. in Statistics* **A5,** 307–323.

Liu, L.–M. and Hanssens, D., (1982), Identification of multiple–input transfer function models, *Comm. in Statistics* **B11,** 297–314.

Lott, W.F., (1973), The optimal set of principal component restrictions on a least–squares regression, *Comm. in Statistics* **2,** 449–464.

Massy, W.F., (1965), Principal components regression in exploratory statistical research, *JASA* **60,** 234–256.

Priestley, M.B., (1971), Fitting relationships between time series, *Bull. Inst. Internat. Statist.* **38,** 1–27, (paper presented at the 38:th session of the I.S.I., Washington, USA).

Pukkila, T., (1979), On the identification of transfer function noise models with several correlated input processes using frequency domain tools, Acta Universitatis Tamperensis, ser A, vol 110, Tampere, Finland.

Pukkila, T., (1980), On transfer function noise model identification, Dept. of Mathematical Sciences, Univ. of Tampere, Finland, (paper presented to the 8:th Nordic conference on Mathematical Statistics held on Ahvenanmaa 26–29.5 1980).

Paper C

Ridge Estimation of Transfer Function Weights [1]

[1] This is a revised version of Research Paper 6380, Economic Research Institute, Stockholm School of Economics, February 1989. The paper has been conditionally accepted for publication in *Communications in Statistics – Simulation and Computation.*

Ridge Estimation of Transfer Function Weights

Per–Olov Edlund

Stockholm School of Economics
Box 6501
S–113 83 Stockholm, Sweden

Key Words and Phrases: *ridge regression; equity estimator; transfer function models; identification procedure; Monte Carlo*

ABSTRACT

In this paper ridge estimators are used to obtain preliminary estimates of the impulse response weights in the Box–Jenkins transfer function model. Five ridge estimators and the Equity estimator are evaluated in a Monte Carlo study covering 5000 different transfer function models. The results show that although the differences between the estimators in most cases are rather small, the Dempster et al estimator *RIDGM* and the estimator by Lawless and Wang are best in terms of low average and median squared deviations from the true coefficient vector. The k values by these estimators are most closely correlated with the optimal k (computed from known true values) and also give the highest frequency of lowest squared error and lowest frequency of highest squared error among the investigated estimators.

Introduction

Ever since the seminal papers by Hoerl and Kennard (1970 a,b) several different estimators of the ridge shrinkage factor, k, have been proposed and evaluated, see eg Hoerl, Kennard and Baldwin (1975), Lawless and Wang (1976), Hocking, Speed and Lynn (1976), McDonald and Galarneau (1975).

Unfortunately, there seems to be no generally superior estimation method. The performance is dependent on the orientation of the parameter vector relative to the eigenvectors, the degree of multicollinearity, the signal–to–noise ratio and other factors. A number of simulation studies have been carried out to investigate the relative merits of the estimators under different conditions, see e.g. Gibbons (1981) and the studies given above.

Ridge methods have been used in preliminary estimation of weights in the Box and Jenkins (1976) transfer function model, see Edlund (1984, 1988).

In the transfer function model identification phase we usually end up with a large number of regression coefficients to be estimated. This is due to the unknown lag structures we want to identify. For example, if we have two input variables each with lags 0–10 we have 22 unknown parameters to estimate (not counting the intercept). Typically, many of the true values are zero. In these cases the ridge estimation technique may be of great value to ensure that the zero–value parameters are forced towards zero.

The goal of this study is to shed some light on the performance of six different biased estimators under the conditions typically encountered in estimating the transfer function weights.

Preliminary estimation of transfer function weights

The Box–Jenkins transfer function model

The Box–Jenkins (1976) m–input transfer function model is,

$$y_t = c + \sum_{i=1}^{m} \frac{\omega_i(B)}{\delta_i(B)} B^{b_i} x_{it} + \epsilon_t \qquad (1)$$

where y_t is the dependent variable, c is a constant term, x_{it} is the i:th independent variable, ϵ_t is an error term representing all 'missing' x variables plus the pure noise (ϵ_t may be autocorrelated), B is the ordinary lag operator, b_i is a pure delay parameter, $\omega_i(B)$ is a 'moving average' operator of order s and $\delta_i(B)$ is an 'autoregressive' operator of order r. It is assumed that the y and x variables are stationary (possibly after differencing). The above formulation has also been used in econometrics, see Jorgenson (1966). Equation (1) may be re–written in a form more similar to the ordinary multiple regression model,

$$y_t = c + \nu_{10} x_{1t} + \nu_{11} x_{1t-1} + \ldots + \nu_{m0} x_{mt} + \ldots + \epsilon_t \qquad (2)$$

where the ν–weights are computed from $\nu(B) = \omega(B) \cdot B^b / \delta(B)$. In equation (2) there are, in the general case, an infinite number of parameters to estimate. In practice only the K first ν–weights for each input variable are significant. If s, r and b are all equal to zero equation (2) becomes the ordinary multiple regression model, which is a special case of the more general transfer function model,

$$y_t = c + \nu_1 x_{1t} + \nu_2 x_{2t} + \ldots + \nu_m x_{mt} + \epsilon_t \qquad (3)$$

Identification of the transfer function model

To estimate the transfer function model efficiently we need knowledge of the values of s, r and b for all input variables. Box and Jenkins (1976) suggested that s, r and b could be guessed from the pattern of preliminary estimates of the ν–weights. There are several ways to estimate these weights, see Edlund (1988). The most simple way is to use the ordinary least–squares (*OLS*) estimates of equation (2). The major problem with *OLS* is multicollinearity between the x variables. Unless the x variables are uncorrelated white noise processes there will be both autocorrelation between lagged values of the same x variable as well as cross correlation between different x variables.

In Edlund (1984, 1988) a two–step method was proposed to deal with the multicollinearity problem and the possible autocorrelation structure in the error term. The multicollinearity problem was solved by using biased regression. In simulation studies of ridge and principal components estimators it was found that ridge regression nearly always showed lower *MSE* than *OLS* and that the principal components estimators in some cases were much worse than *OLS*.

The ordinary ridge estimator

The *OLS* estimate of the multiple regression model is in matrix form,

$$\hat{\boldsymbol{\nu}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \qquad (4)$$

Let us assume that the variables have been scaled so that $\mathbf{X}'\mathbf{X}$ is the correlation matrix and $\mathbf{X}'\mathbf{y}$ is the correlation vector between the x variables and y. The model can also be expressed in a canonical form,

$$\hat{\boldsymbol{\alpha}} = \boldsymbol{\Lambda}^{-1}\mathbf{Z}'\mathbf{y} \qquad (5)$$

where $\hat{\boldsymbol{\nu}} = \mathbf{V}\hat{\boldsymbol{\alpha}}$. $\boldsymbol{\Lambda}$ is a diagonal matrix with eigenvalues $\lambda_1 > \lambda_2 \ldots > \lambda_p$ on the main diagonal, $\mathbf{V}$ is a matrix of eigenvectors corresponding to the eigenvalues in $\boldsymbol{\Lambda}$, $\mathbf{Z}$ is a matrix with canonical variables: $\mathbf{Z} = \mathbf{XV}$.

The ordinary ridge estimator is then defined as,

$$\hat{\boldsymbol{\nu}}_{RR}(k) = (\mathbf{X}'\mathbf{X}+k\mathbf{I})^{-1}\mathbf{X}'\mathbf{y} \qquad (6)$$

where k is the shrinkage factor. In the canonical form computation of the ridge estimators is particularly simple since, $\hat{\alpha}_i(k) = \lambda_i\hat{\alpha}_i/(\lambda_i+k)$ and $\hat{\boldsymbol{\nu}}_{RR}(k) = \mathbf{V}\hat{\boldsymbol{\alpha}}_{RR}(k)$.

Ridge estimators included in the study

This study covers four ordinary ridge estimators and two related estimators. The selection of estimators has been based on two criteria. The first, and most important criterion, is the performance with respect to *MSE* relative to other estimators.

Therefore only estimators that have performed well in other simulation studies have been included. The second criterion is the complexity of computations needed to estimate the shrinkage parameter. The estimates should not be to complicated to obtain. In the following the estimators will be described.

The HKB Estimator

This estimator was suggested by Hoerl, Kennard and Baldwin (1975). The HKB estimator,

$$k_{HKB} = p\hat{\sigma}^2 / \sum_{i=1}^{p} \hat{\alpha}_i^2 \tag{7}$$

was derived from two different results. First, if the x variables are orthogonal then the k_{HKB} is the minimum *MSE* estimator. Secondly, for the generalized ridge estimator (which has individual k values for each a_i), $k_i = \sigma^2/\alpha_i^2$ will minimize the *MSE*. The k_{HKB} is the harmonic mean of the k_i:s with the unknown parameters replaced by their sample estimates. The HKB estimator has been included in many simulation studies, see e.g. Hoerl and Kennard (1976), Lawless and Wang (1976), Lawless (1978), Gibbons (1981) and Edlund (1988).

The LW Estimator

The LW estimator,

$$k_{LW} = p\hat{\sigma}^2 / \sum_{i=1}^{p} \lambda_i \hat{\alpha}_i^2 \tag{8}$$

was proposed by Lawless and Wang (1976). This estimator was derived using a Bayesian approach. It has been studied by among others Lawless (1978), Wichern and Churchill (1978), Gibbons (1981) and Edlund (1988). In the Edlund (1988) study the LW and HKB estimators both performed well relative to *OLS* and the estimator proposed by Hocking, Speed and Lynn (1976). None of the two estimators dominated the other.

The RIDGM Estimator

Dempster, Schatzoff and Wermuth (1977) suggested to choose k so that,

$$\sum_{i=1}^{p} \hat{\alpha}_i^2 / \left[\frac{\hat{\sigma}^2}{k} + \frac{\hat{\sigma}^2}{\lambda_i}\right] = p \tag{9}$$

This estimator, k_{RIDGM} is also motivated by a Bayesian interpretation of the ordinary ridge estimator. This estimator has been studied by, e.g. Gibbons (1981) where it was among the best estimators and Miller and Tracy (1984).

The COV Estimator

This estimator has been proposed both by Bulcock, Lee and Luk (1981) under the name "Normalization Ridge Regression" and by Miller and Tracy (1984) as "Correct Orthogonal Variance (COV) Ridge Estimator". They choose the k_{COV} which satisfies,

$$\sum_{i=1}^{p} \lambda_i/(\lambda_i + k)^2 = p \quad (k \geq 0) \tag{10}$$

The rationale for this estimator is that for orthogonal x variables the sum of variance inflation factors for the standardized variables ($\Sigma\ 1/\lambda_i$) are equal to p, since all λ_i:s equal 1. By choosing k_{COV} to satisfy (10) we force the sum of inflation factors for the ridge estimator to be equal to the value for orthogonal variables. This estimator differs from the other estimators by the fact that it is not dependent of sample values (if we regard the x variables to be fixed). On the other hand the value of k will not adapt to variations in the residual variance and different regression vectors. This estimator is relatively new and has not been included in other simulation studies, but the studies by Bulcock et al and Miller and Tracy indicate that this estimator may be superior to the estimators above in some cases.

The OR Estimator

The Orthogonalizing Ridge estimator, proposed by Leskinen (1980), is not an ordinary ridge estimator since only some of the α_i:s will be shrunk,

$$\alpha_{OR,i} = \begin{cases} \alpha_i, & i = 1, \ldots, q \\ \lambda_i \alpha_i/(\lambda_i + (1-\lambda_i)k), & i = q+1, \ldots, p \end{cases} \tag{11}$$

where q is the number of $\lambda_i \geq 1$. By shrinking only the α_i:s with large variance (σ^2/λ_i) the $\hat{\nu}_i$:s will not shrink to zero as $n \to \infty$. The value of k_{OR} can be determined in a number of ways. For this study a *HKB*–type of estimator has been chosen,

$$k_{OR} = (p-q)\hat{\sigma}^2 / \sum_{i=q+1}^{p} (1-\lambda_i)(\hat{\alpha}^2_{OLS,i} - \hat{\sigma}^2/\lambda_i) \tag{12}$$

The admissible region of k_{OR} is [0, 1]. Values outside this interval are set to 1.

The Equity Estimator

This estimator, proposed by Krishnamurthi and Rangaswamy (1987), is not a ridge estimator, rather it uses a square root transformation to the eigenvalues, λ_i, before inversion to reduce the negative effects of small eigenvalues. We thus obtain,

$$\hat{\nu}_{EQ} = c\,(\mathbf{V}\Lambda^{-1/2}\mathbf{V}')\mathbf{X}'\mathbf{y} \quad \text{where} \quad c = \frac{\mathbf{y}'\mathbf{X}(\mathbf{V}\Lambda^{-1/2}\mathbf{V}')'\mathbf{X}'\mathbf{y}}{\mathbf{y}'\mathbf{X}\mathbf{X}'\mathbf{y}} \tag{13}$$

c is a scalar to give the $\hat{\nu}_{EQ}$ the correct level. The performance of this estimator when

using a squared error criterion is reported by Krishnamurthi and Rangaswamy to be superior to the HKB estimator described above.

To study how well the above estimators perform relative to optimal ridge estimation, two optimal k estimators have been included. Studies by Leskinen (1980) and Gibbons (1981) show that the large potential decrease in *MSE* will not be realized for all estimators and not for all parameter settings. It is also useful to be able to study the potential reduction in *MSE* as a function of the parameters in the study.

The first estimator, k_{opt}, is the k value minimizing the expected *MSE* given knowledge of the true parameters,

$$MSE(k) = \sigma^2 \sum_{i=1}^{p} \lambda_i/(\lambda_i+k)^2 + k^2 \sum_{i=1}^{p} \alpha_i^2/(\lambda_i+k)^2 \qquad (14)$$

The second estimator, k_{min}, is the k value minimizing the observed value of,

$$\sum_{i=1}^{p} (\hat{\alpha}_i-\alpha_i)^2. \qquad (15)$$

Monte Carlo simulations

In this simulation study the design parameters have been varied over the whole range of possible conditions for the investigated model type. Most other simulation studies of ridge estimators have been more limited in the choice of design parameter values. Specific features of this study are the choice of structure in the correlation matrix and the selection of the true parameter vector. The design of this study resembles the design used by Krishnamurthi and Rangaswamy (1987).

Models considered

The basic model is,

$$y_t = \frac{\omega_1(B)}{\delta_1(B)} B^{b_1} x_{1t} + \frac{\omega_2(B)}{\delta_2(B)} B^{b_2} x_{2t} + \epsilon_t \qquad (16)$$

or

$$y_t = \sum_{i=0}^{\infty} \nu_{1i} x_{1,t-i} + \sum_{i=0}^{\infty} \nu_{2i} x_{2,t-i} + \epsilon_t \qquad (17)$$

where

$$x_{1t} = \frac{\theta_1(B)}{\phi_1(B)} a_{1t} \quad x_{2t} = \frac{\theta_2(B)}{\phi_2(B)} a_{2t} \qquad (18)$$

and,

$$\epsilon_t \in \text{iid } N(0,\sigma_\epsilon^2),\ a_{1t} \in \text{iid } N(0,\sigma_{a_1}^2),\ a_{2t} \in \text{iid } N(0,\sigma_{a_2}^2).$$

The polynomials $\omega_i(B)$, $\delta_i(B)$, $\theta_i(B)$ and $\phi_i(B)$ are all of order 0, 1 or 2. For practical purposes the ν weights have been truncated after lag 10 for both variables which means that a total of 22 regression coefficients have been estimated.

Simulation procedure

Step 1. Choose the number of observations, n. $n \in U(50, 150)$. This range cover values likely to be encountered in practical business and economic applications. Since we want to identify two lag structures, 50 observations seems to be a lower limit.

Step 2. Generate a correlation matrix of size 22×22. This is accomplished in several substeps. First *ARMA* processes for x_1 and x_2 are generated. The order of the *AR* and *MA* processes are $U(0, 2)$. Given the order the parameters are generated so that all outcomes have equal probability and the processes are invertible. Covering nine different *ARMA* models from *ARMA*(0,0) (a white noise process) to *ARMA*(2,2) it is believed that enough variation in the input processes is generated. Next the theoretical autocorrelation functions are computed. Then the correlation between a_{1t} and a_{2t} is generated from $U(-1, 1)$. Using the cross correlation coefficient and the computed ψ weights ($\psi(B) = \theta(B)/\phi(B)$) for the two input processes, the theoretical cross correlation function is obtained. The correlation functions are then used to create the correlation matrix.

Using theoretical rather than sample correlation matrices reduces the computing time considerably. An experiment comparing the two approaches shows no significant differences for the studied statistics. The only major difference seems to be a higher average degree of multicollinearity for the sample matrices. A set of tables comparing the two approaches may be obtained from the author.

Step 3. Eigenvalues and eigenvectors were computed for the correlation matrix and sorted in descending order for the eigenvalues.

Step 4. The orders of ω(B) and δ(B) are generated from $U(0, 2)$. The δ(B) polynomials are generated in the same way as the ϕ(B) and θ(B) polynomials in step 2. The ω weights are generated from $U(-2, 2)$. The pure delay b_i is generated from $U(0, 3)$. Then the ν weights are computed and normalized so that $\boldsymbol{\nu}'\boldsymbol{\nu} = 1.0$.

Step 5. The true standard deviation of the error term, σ, is generated from $U(0.01, 0.25)$. The interval has been chosen to give t values for the estimated non–zero regression coefficients that are not too far from standard critical values. In some studies the signal–to–noise ratio $\boldsymbol{\nu}'\boldsymbol{\nu}/\sigma^2$ has been varied from "the very insignificant to the completely meaningless" (Draper and Van Nostrand (1979)). The sample value of σ^2 was generated from a χ^2_{n-p} distribution using $\hat{\sigma}^2 = \sigma^2\chi^2_{n-p}/(n-p)$.

Step 6. The true $\boldsymbol{\alpha}$ vector is computed from $\boldsymbol{\alpha} = \mathbf{V}'\boldsymbol{\nu}$. Then sample values for $\boldsymbol{\alpha}$

are generated from $\hat{\alpha}_i \in N(\alpha_i, \sigma^2/\lambda_i)$.

Step 7. In the final step estimates of all estimators are obtained. For the *RIDGM*, *COV* and *OPT* estimators the k value was obtained using a Newton–Raphson algorithm. For the *MIN* estimator a quadratic interpolation algorithm was used with the k_{opt} value as a starting point.

Steps 1 trough 7 were repeated 5000 times. Appropriate statistics were computed and stored for later analysis. There were some deviations from the described procedure:

- Some of the generated correlation matrices were not positive definite, i.e. some eigenvalues were negative, due to rounding errors. This occurred when the generated input processes were close to non–stationarity. These matrices were deleted and new matrices were generated.
- To avoid cases with too high expected *MSE*, cases with $E(MSE_{OLS}) > 50$ were excluded from the simulation.
- In some cases k_{RIDGM} becomes very large giving estimated coefficients practically equal to zero. To avoid those rather extreme cases k_{RIDGM} was limited to the interval [0, 5].

The simulation program was written in Microsoft Fortran v 4.01 and run on an AT–compatible micro computer. The random number generator was multiplicative congruential $x_i = a \cdot x_{i-1} \, mod \, m$ with $a = 1{,}220{,}703{,}125$ and $m = 2^{35}$, see Jennergren (1984).

The above procedure will ensure that the model parameters are distributed evenly over the parameter space and that the analysis of the performance of various ridge estimators is not seriously affected by the choice of simulation model.

Performance measures

Since the estimated ν coefficients are to be used for identification of the order of the ω(B) and δ(B) operators and the pure delay b it seems reasonable to put emphasis on the use of a criteria that measure the closeness between the estimated and the true ν weights. The most commonly used criterion is then the squared error of the coefficients $\Sigma (\hat{\nu}_i - \nu_i)^2$. To compare different ways of estimating k some measure of the average k value and standard deviation seems appropriate. Below is a list of the estimated statistics:

1. For each sample, i, and for each estimator, j, the following statistics were computed:

$$SE(j)_i = \sum_{h=1}^{2\,2} (\hat{\alpha}_h - \alpha_h)^2 \quad \text{and} \quad M(j)_i = SE(j)_i / SE(OLS)_i.$$

$M(j)_i$ was then averaged over all samples to obtain

$$\bar{M}(j) = \sum_{i=1}^{m} M(j)_i / m$$

where m denotes the sample size for each category in the diagrams below.

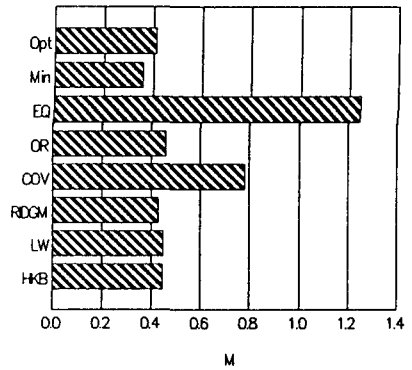
The reason for averaging over the relative numbers $M(j)$ rather than the original values $SE(j)$ is that the latter are affected by the residual variance and the eigenvalues who vary from sample to sample. The $\bar{M}(j)$ statistic therefore measures the average performance relative to *OLS*. The median of $M(j)_i$ the minimum and maximum values as well as the standard error of $\bar{M}(j)$ were also computed.

2. The mean k value for the ridge estimators and the standard error of the mean k value.
3. Correlation between the estimated k values and k_{opt} and k_{min}.
4. Number of times the estimator has the minimum and the maximum *SE*.
5. Number of times the estimator has lower *SE* than *OLS*.

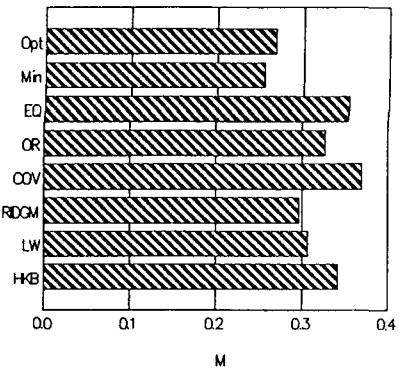
In addition to the estimated statistics the following values were collected for each sample; $\phi = \lambda_{max}/\lambda_{min}$ (the spectral condition number), σ and the correlation between the ν vector and the eigenvector corresponding to the smallest eigenvalue $\rho_{\nu, v_{\min}}$.

Results of the simulations

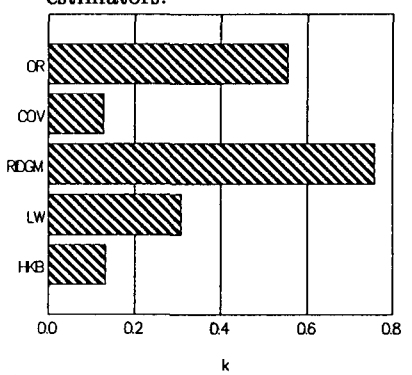
The main results have been summarized in a set of graphs. (A complete set of tables can be obtained from the author). Figure 1 shows results from the total sample of 5000 replications. As can be seen from Figures 1a and 1b the distribution of M values is skew for all estimators, but most severely for the *COV* and *EQ* estimators which have some very large M values. The median values do not show much dispersion among the estimators and the two best, the *RIDGM* and the *LW* estimators, are not too far from the optimal estimators *OPT* and *MIN*. The average k values differ a lot (Fig. 1c). The average k_{RIDGM} is about twice as large as the average k_{LW} but they still have very similar average and median M values. The *RIDGM* and *LW* estimators, who perform well relative to the other estimators are most closely correlated with the k_{OPT} estimator (Fig. 1d) and have the largest number of times "minimum *SE*" (28.8 % and 20.9 % respectively) and smallest



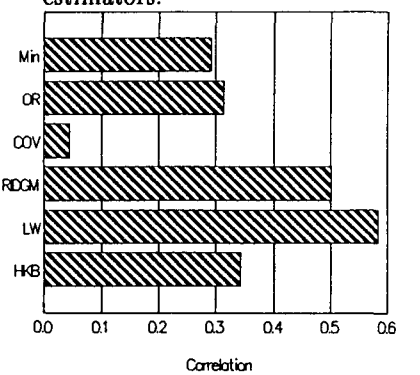
(a) Average M values for the estimators.



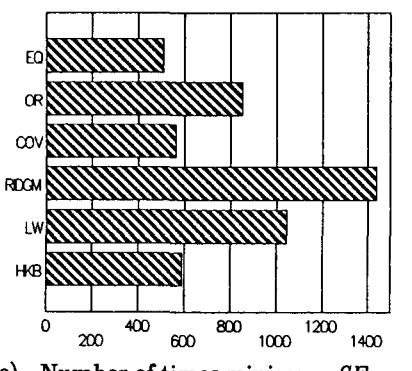
(b) Median M values for the estimators.



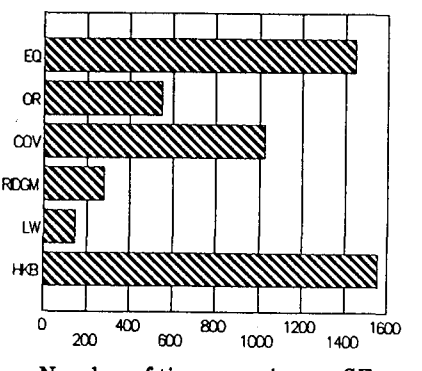
(c) Average k values for the estimators.



(d) Correlation between k values for different estimators and k_{OPT}

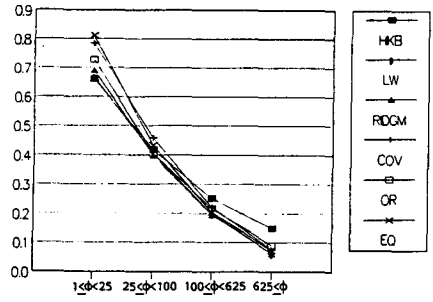


(e) Number of times minimum SE.

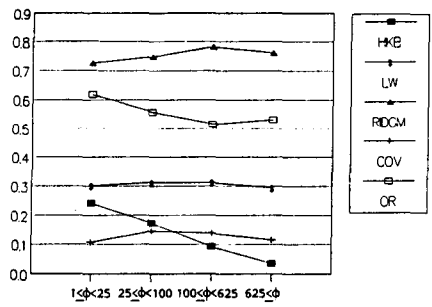


(f) Number of times maximum SE.

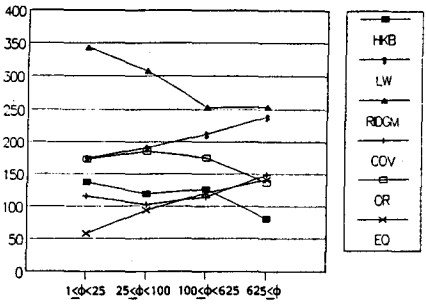
FIG 1. Results of the simulations for all 5000 replications.



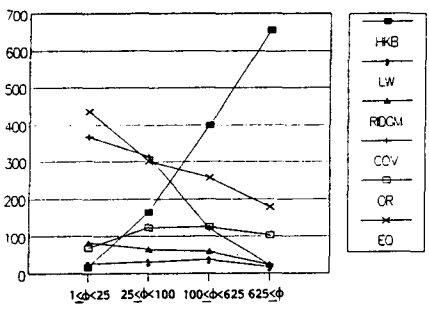
(a) Median M values for the estimators.



(b) Average k values for the estimators.



(c) Number of times minimum SE.



(d) Number of times maximum SE.

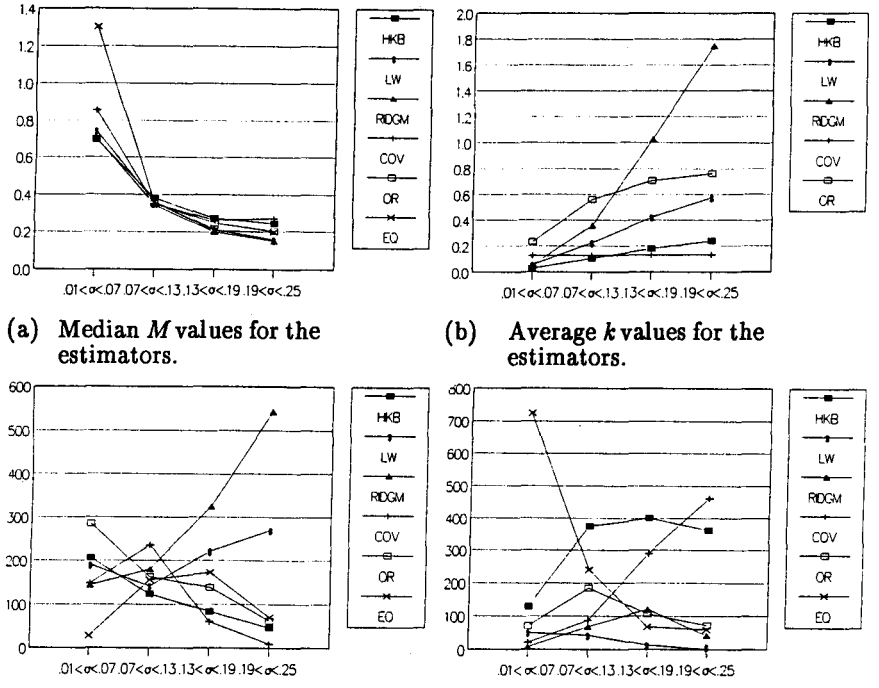
FIG 2. Results of the simulations for different levels of multicollinearity, ϕ. 1000 replications in each group.

number of "maximum SE" (5.6 % and 2.8 % respectively). All estimators give lover SE than OLS in about 70 % of the replications. From this overall comparison the $RIDGM$ and LW estimators seem to be the most reliable estimators.

To further investigate the behavior of the estimators the replications have been grouped according to three important factors; the degree of multicollinearity, the signal–to–noise ratio and the orientation of the coefficient vector ν. In each case the sample has been divided into four groups, each containing the first 1000 replications falling within the class limits (since the number of replications available in each group varies). Thus only 4000 of the 5000 replications have been used in Figures 2–4.

Degree of multicollinearity

The spectral condition number $\phi = \lambda_{max}/\lambda_{min}$ is a measure of the multicollinearity in the correlation matrix. As ϕ increases we would expect the average SE



(a) Median M values for the estimators.

(b) Average k values for the estimators.

(c) Number of times minimum SE.

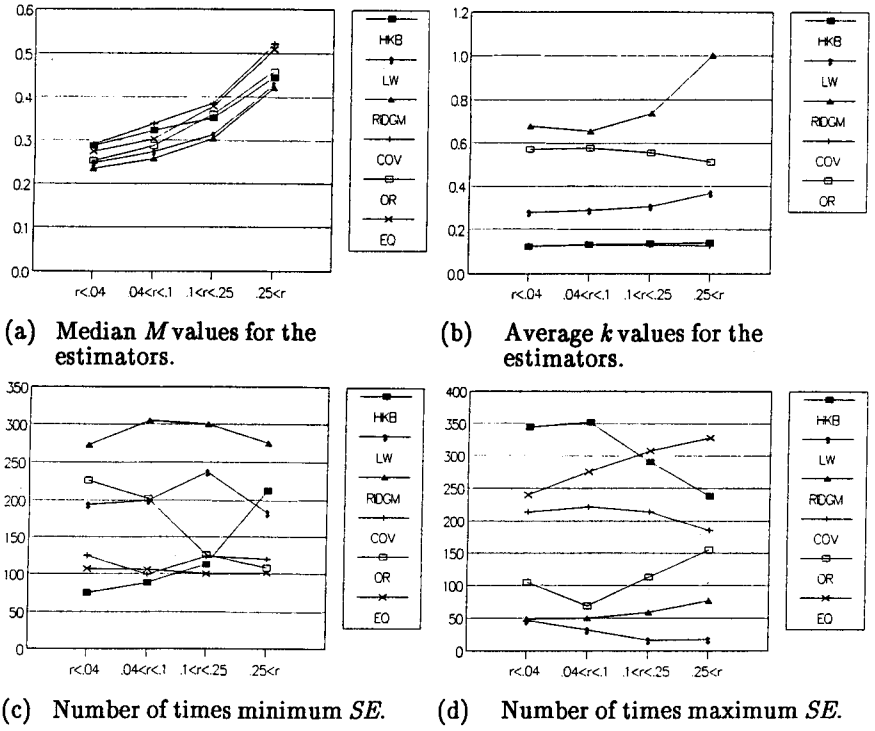
(d) Number of times maximum SE.

FIG 3. Results of the simulations for different levels of residual standard deviation, σ. 1000 replications in each group.

of the *OLS* estimator to increase and the M values for the other estimators to decrease as they become relatively more efficient to *OLS*. As can be seen in Figure 2 this is true. The median M values all decrease as ϕ increases (Fig. 2a). Although the differences are rather small, the *RIDGM* and *LW* estimators are among the top three for all values of ϕ. The *RIDGM* estimator has the largest number of minimum SE (Fig. 2c) and the *LW* the smallest number of maximum SE (Fig. 2d) for all ϕ. All estimators improve their number of times better than *OLS* as ϕ increases, from 42 % when $\phi < 25$ to about 91 % when $\phi \geq 625$.

Signal–to–noise ratio

The signal–to–noise ratio, $\nu'\nu/\sigma^2$, was generated from the interval 16 to 10000. Since $\nu'\nu$ was held constant the signal–to–noise ratio is represented by the residual standard deviation, σ. Here we would expect the average SE for *OLS* to increase as σ



(a) Median M values for the estimators.

(b) Average k values for the estimators.

(c) Number of times minimum SE.

(d) Number of times maximum SE.

FIG 4. Results of the simulations for different orientations of the coefficient vector, r. 1000 replications in each group.

increases and the other estimators to improve relative to OLS as σ increases. As can be seen in Figure 3a this holds for all estimators.

The *Equity* estimator was in the median worse than the OLS when the residual standard deviation was low. Apart from the COV estimator the k values increase as σ increases. This should be expected since a larger σ (or rather $\hat{\sigma}^2$) means a larger k (except for k_{COV}). All estimators are superior to the OLS in around 31 % for the smallest σ up to 98 % for the largest σ. The $RIDGM$ and LW perform better relative to the other estimators as σ increases and are definitely the best when $\sigma \geq .13$ (Fig. 3c).

Orientation of the coefficient vector

The orientation of the coefficient vector is important when we want to study the estimators behavior relative to OLS. When ν is strongly correlated to the eigenvector,

v, corresponding to largest (smallest) eigenvalue the ridge estimators will have their best (worst) case relative to *OLS*. In this study the absolute value of the correlation between ν and v_{min} has been computed. In Figure 4a we can observe the expected behavior of increasing median M as the correlation increases. The *RIDGM* and *LW* estimators are still among the top estimators, irrespectively of the value of the correlation.

Conclusions

From the results above it is quite clear that ridge regression may offer a substantial decrease in *MSE* in most cases which will help in identifying a suitable transfer function model. This is especially true in cases where estimation is most difficult.

In most cases the best estimators gave *SE*s not too far from the optimal estimators. Even though the differences between the estimators in most cases are small the *RIDGM* and *LW* estimators by Dempster et al (1977) and Lawless and Wang (1976) seem to be the most reliable estimators. Both these estimators are developed using a Bayesian approach and seem to be particularly useful for this kind of model (where many true coefficients may be zero). They both give the minimum *SE* a large number of times (together about 50 % of the cases) and the maximum *SE* a small number of times (about 8 %).

The estimators proposed by Hoerl et al and Leskinen are not far from the two best estimators, while the *COV* and *Equity* estimators in some cases give very large M values. For the *Equity* estimator this is in contrast to the results of Krishnamurthi and Rangaswamy (1987) who found that the *Equity* estimator were superior to the *HKB* estimator in nearly all cases. The differences may, at least partially, be explained by differences in models, allowed interval of spectral condition number and residual variance. The *Equity* estimator performs better as the *MSE* of *OLS* increases. The Krishnamurthi and Rangaswamy study includes larger *MSE* values for the *OLS* estimator than the present study. They also use a different *SE* measure that gives lower weight to large M values when *MSE* is small.

Acknowledgement

This research has been supported by grants from The Swedish Research Council for Humanities and Social Sciences (HSFR) and The Bank of Sweden Tercentenary Foundation (Riksbankens Jubileumsfond).

Bibliography

Box, G.E.P. and Jenkins, G.M. (1976). *Time series analysis: Forecasting and control.* 2:nd ed. San Francisco: Holden–Day.

Bulcock, J.W., Lee, W.F. and Luk, W.S. (1981). Ordinary least squares, ridge regression and normalization ridge regression: Some comparisons. Paper presented at the Institute of Theology, University of Uppsala, Sweden, May 1981.

Dempster, A.P., Schatzoff, M. and Wermuth, N. (1977). "A simulation study of alternatives to ordinary least squares," *JASA, 72,* 77–91.

Draper, N.R. and Van Nostrand, R.C. (1979). "Ridge regression and James–Stein estimation: review and comments," *Technometrics, 21,* 451–466.

Edlund, P.–O. (1984). "Identification of the multi–input Box–Jenkins transfer function model," *Journal of Forecasting, 3,* 297–308.

Edlund, P.–O. (1988). "On identification of transfer function models by biased regression methods," Forthcoming in *Journal of Statistical Computation and Simulation.*

Gibbons, D.G. (1981). "A simulation study of some ridge estimators," *JASA, 76,* 131–139.

Hocking, R.R., Speed, F.M. and Lynn, M.J. (1976). "A class of biased estimators in linear regression," *Technometrics, 18,* 425–437.

Hoerl, A.E., and Kennard, R.W. (1970a). "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics, 12,* 55–67.

Hoerl, A.E., and Kennard, R.W. (1970b). "Ridge regression: Applications to nonorthogonal problems," *Technometrics, 12,* 69–82.

Hoerl, A.E., and Kennard, R.W. (1976). "Ridge regression: iterative estimation of the biasing parameter," *Comm. in Statistics, A5,* 77–88.

Hoerl, A.E., Kennard, R.W. and Baldwin, K.F. (1975). "Ridge regression: some simulations," *Comm. in Statistics, 4,* 105–123.

Jennergren, L.P. (1984). *Discrete–events simulations models in PASCAL/MT+ on a microcomputer.* Lund: Studentlitteratur.

Jorgenson, D.W. (1966). "Rational distributed lag functions," *Econometrica, 32,* 135–148.

Krishnamurthi, L. and Rangaswamy, A. (1987). "The equity estimator for marketing research," *Marketing Science, 6,* 336–357.

Lawless, J.F. (1978). "Ridge and related estimation procedures: Theory and practice," *Comm. in Statistics, A7,* 139–164.

Lawless, J.F. and Wang, P. (1976). "A simulation study of ridge and other regression estimators," *Comm. in Statistics, A5,* 307–323.

Leskinen, E. (1980). On the biased estimation in regression: studies on ridge–type estimation. Jyväskylän Yliopisto, Jyväskylä, Finland.

McDonald, G.C. and Galarneau, D.I. (1975). "A Monte Carlo evaluation of some ridge–type estimators," *JASA*, *70*, 407–416.

Miller, T.I. and Tracy, R.L. (1984). "Further results concerning ridge regression," *Comm. in Statistics*, *13*, 251–262

Wichern, D.W. and Churchill, G.A. (1978). "A comparison of ridge estimators," *Technometrics*, *20*, 301–311.

Paper D

On Identification of Transfer Function Models for Business Cycle Forecasting [1]

[1] This is a revised version of a paper presented at "The Ninth International Symposium on Forecasting" in Vancouver, Canada, June 18–21 1989.

On Identification of Transfer Function Models for Business Cycle Forecasting [2]

Per–Olov Edlund
Stockholm School of Economics
Box 6501
S–113 83 STOCKHOLM
Sweden

ABSTRACT

The first phase of transfer function model identification is preliminary estimation of transfer function weights. Previous studies by the author have shown that ordinary least–squares estimates in most cases can be improved significantly by using ridge regression techniques. In particular the use of the RIDGM and Lawless and Wang estimators are recommended. The main object of this paper is to compare the ridge estimators to other relevant estimators on real data. For this purpose business cycle data from Sweden have been collected and transfer function models have been identified by ridge regression, prewhitening and cross–spectral analysis. The results indicate that the ridge estimators compare favorable to the prewhitening estimator. The prewhitening technique advocated by practitioners seems to generate too many significant weights which complicates the search for a suitable model. The theoretically correct cross–spectral estimator gives too few significant weights and fails to identify two of the three models.

Key words: *ridge regression, transfer function models, prewhitening, cross–spectral analysis, identification procedure, business cycle, leading indicators*

---

2 This research has been supported by grants from The Swedish Research Council for Humanities and Social Sciences (HSFR) and The Bank of Sweden Tercentenary Foundation (Riksbankens Jubileumsfond).

1 Introduction

In business cycle forecasting with leading indicators the dependent variable is a function of lagged values of the indicators. Since the theoretical model very seldom imply the exact nature of the lag structures, there clearly is a need for an empirical method to identify the lag structures.

The transfer function model approach described by Box and Jenkins (1976) is particularly well suited to describe, identify and estimate the dynamic relationships based on theoretical business cycle models. The dynamic relationship, the transfer function, is not limited in shape as the shapes estimated by Koyck schemes, Almon lags and other traditional econometric approaches to dynamic modeling. There are also different ways of identifying, or preliminary estimating the transfer function.

In this paper only the identification phase of transfer function modeling will be considered, with a special emphasis on preliminary estimation of the transfer function weights.

Box and Jenkins proposed the prewhitening cross–correlation method for preliminary estimation of the transfer function weights. This method seems to work well only when there is one independent variable or when the independent variables are mutually uncorrelated. In business cycle modeling with more than one leading indicator, as well as in economic time series in general, this is unlikely to be the case.

There are some other methods for preliminary estimation of transfer function weights, among them a cross–spectral method proposed by Box and Jenkins (1976, p 413–6) and a two–step ridge regression approach proposed by the author (Edlund 1984, 1989a,b).

These methods seem to works well even in the case of several cross–correlated independent variables. The regression method has the advantage of being quite easy to use and it is also easy to find suitable computer software.

The main purpose of this paper is to compare the ridge regression approach to the prewhitening and spectral methods of Box and Jenkins on identification of different business cycle models.

Section 2 gives a short overview of business cycle forecasting. Then the transfer function model is presented in Section 3 and different ways to preliminary estimate the impulse response weights are described. In Section 4 the results of a small empirical study of building transfer function models for the Swedish index of industrial production are presented. Section 5 contains a brief summary of the results and some concluding remarks.

2 Business cycle forecasting

Business cycles play an important role in government decision making as well as in corporate decision making. Traditionally interest has been focused on predicting turning points in the economy, but also the value of the reference series used to measure economic activity is of interest, e.g. what value will GNP reach at the next peak.

There are various approaches to forecasting the business cycle. In Sweden the official forecasts are based both on survey data of opinions, expert forecasts and on economic time series. The forecasts are based both on qualitative judgment and on some relatively simple econometric models. In the U.S.A. several large econometric models are in use, but their forecasting abilities have not always been impressive compared to simpler time series models like the *ARIMA* models. To overcome the criticism of economists who prefer a theory based model to a simple time series model but yet let the data help in specifying the model the transfer function approach may be of interest. This class of models are particularly suitable for identifying and estimating business cycle models with leading indicators as explanatory variables.

Business cycle theories comprehend many economic time series as e.g. production, employment, investment. Many of these series show a cyclical pattern that leads the chosen business cycle series. The theories do not explicitly state the dynamic relationship between the leading indicator series and the business cycle series, even though there are considerable empirical evidence about the dynamic structures and approximate lead times. In practice the leading indicators are used without much reference to the theories. OECD, e.g. combine several leading indicators into a composite index that is related to the reference series. This index will contain a mixture of different dynamic structures that may not be clear to the user.

There are certain assumptions that are made when using leading indicators for business cycle forecasting. First we assume that the dynamic relationships are stable over time. This assumption is not well supported for Swedish data (see Westlund and Claesson 1989). A second important assumption is that the dynamic structure is the same both for upswings and down–swings in the economy. Violation of this assumption leads to the need for two different lag structures. This assumption can be tested by recursive estimation of the dynamic structure. In this study we will assume that the relationships are stable, and similar, enough to be identified and estimated by the ordinary transfer function techniques.

Koch and Rasche (1988) examine a number of leading indicators included in the U.S. Commerce Department "Composite Index of Leading Indicators" using the transfer function approach described in the next Section. In this paper a similar approach will be used to build a model for the Swedish Index of Industrial Production, but here the main purpose of the study lies in evaluating different methods of preliminary estimating the lag structure between the leading indicators and the reference series.

3 Transfer function model identification

The transfer function–noise model proposed by Box and Jenkins (1976) is,

$$y_t = c + \frac{\omega(B)}{\delta(B)} x_{t-b} + n_t$$

where y_t is the dependent variable, c is a constant term, x_t is the independent variable, n_t is an error term which represents all the 'missing' x variables plus the pure noise, B is the ordinary lag operator, $\omega(B)$ is a 'moving average' operator, $\delta(B)$ is an 'autoregressive' operator, b is a pure delay parameter which represents the number of complete time intervals before a change in x_t begins to have an effect on y_t. (The y and x variables are assumed to be differenced/transformed to be mean and variance stationary.)

The transfer function $\nu(B)$ is a rational lag structure,

$$\nu(B) = \frac{\omega(B)}{\delta(B)} B^b \quad \text{or} \quad \nu(B) = (\nu_0 + \nu_1 B + \nu_2 B^2 + \ldots) B^b$$

that may be used to represent any form of linear dynamic relationship between x_t and y_t to any specified degree of accuracy. The weight ν_i is called the impulse response weight at lag i.

The noise term, n_t, may be expressed as an ordinary $ARMA(p,q)$ model of the form

$$n_t = \frac{\theta(B)}{\phi(B)} a_t$$

where $\theta(B)$ is a moving average operator of order q, $\phi(B)$ is an autoregressive operator of order p, and a_t a white noise variable.

The single–input model is easily generalized to an m–input transfer function model,

$$y_t = c + \sum_{j=1}^{m} \frac{\omega_j(B)}{\delta_j(B)} x_{j,t-b_j} + \frac{\theta(B)}{\phi(B)} a_t$$

with m input variables $x_{1t}, x_{2t}, \ldots, x_{mt}$. The ordinary multiple regression model,

$$y_t = c + \nu_{10} x_{1t} + \nu_{20} x_{2t} + \ldots + \nu_{m0} x_{mt} + a_t$$

is a special case of the more general multiple–input transfer function model.

To identify the order of the operators $\delta(B)$, $\omega(B)$ and b two steps are performed. First the impulse response weights, ν_i, are preliminary estimated. Then the pattern of the weights is used to guess the order of the operators (see Box and Jenkins 1976 Fig. 10.6, p 349). The Corner method can also be used to guess the orders (e.g., see Tsay 1985).

When Box and Jenkins popularized the transfer function model (and the univariate *ARIMA* model) they also suggested different methods for preliminary estimation of the transfer function weights. For the one–input model they preferred the prewhitening cross–correlation approach described below. They showed that the weights also could be estimated by ordinary least–squares (*OLS*) but that these estimates have several deficiencies. Later Pukkila (1980) showed that the *OLS* estimator may be used even when there are more than one input variable with reasonable results. The *OLS* estimator has been further improved by using a two–step ridge regression approach (see Edlund 1984, 1987, 1989a,b). For the multiple–input case Box and Jenkins gave a cross–spectral estimator (Box and Jenkins 1976, Ch 11). This estimator has been investigated and made operational by Pukkila (1979).

Priestley (1971) proposed the covariance contraction method which is similar to the prewhitening cross–correlation method but all variables are prewhitened separately. Later Haugh and Box (1977) presented a method similar to Priestley's method. Both methods are suitable for the one–input model, but Fask and Robinson (1977) extended Priestley's method to multiple–input models. Tsay (1985) suggested the use of a *VAR* model which would avoid prewhitening and allow for non–stationary variables. This approach would also be useful for testing the hypothesis of unidirectional causality. The *VAR* approach is also suitable for identification of multiple–input models.

Another approach not requiring prewhitening of the variables is the maximum entropy–generalized least–squares (*ME–GLS*) estimator proposed by Rahiala (1986). He begins by estimating the residual model in the transfer function model from estimates of the cross–spectra of the variables by a maximum entropy method. Then

the impulse response weights are estimated by an ordinary *GLS* estimator.

In this study only the prewhitening cross–correlation method, the cross–spectral method and the two–step ridge regression method will be considered.

3.1 The prewhitening cross–correlation method

Box and Jenkins noted that the most straight–forward way to preliminary estimate the impulse response weights would be to use a simple *OLS* estimator using present and lagged values of the input variables, x_i, as independent variables and y as the dependent variable. They argue that the simple *OLS* estimator do not provide efficient estimates, are cumbersome to solve and require knowledge of the lag K at which the weights are effectively zero (p 379).

To solve the efficiency problem, at least partially, and to decrease the effects of choosing K to small, Box and Jenkins suggest that both the input and the output series are transformed by the linear filter that makes the input series white noise. Then the correlation matrix of lagged input variables will be (almost) diagonal and the estimates will not suffer from multicollinearity and possibly left out lags.

While this method works well for the single–input model it is not directly applicable for the multiple–input case. If the input variables are cross–correlated the estimates will suffer from multicollinearity. Damsleth (1979) reports experiences from using this approach for a two–input model. To decrease the cross–correlation problem it would be possible to compute the partial cross–correlation when holding the effects of the other input variables constant.

According to representatives from the Gwilym Jenkins & Partners Ltd consulting firm (with considerable experience from using the transfer function model) it is quite safe to do the prewhitening for one input variable at a time, neglecting the effects of the other input variables (see also McLeod 1982). The same approach is also implemented in the the AUTOBOX software package from AFS Inc. This will be the approach used in the empirical part of this paper.

3.2 The cross–spectral analysis method

In an Appendix (p 413–6) Box and Jenkins briefly discuss identification of multiple–input transfer function models. They consider the use of cross–spectral analysis to estimate the impulse response weights. By taking a Fourier transform of the impulse response function a frequency response function is obtained. The frequency response function is estimated from the cross spectra of the time series.

Then using an inverse transformation estimates of the impulse response weights are obtained. This method is easily extended to multiple–input models.

This method do not require prewhitening, even though the estimates of spectra are improved if the series are white noise and aligned (i.e. not leading/lagging each other). Suitable filters for the series can be found by studying the autocorrelation and the cross–correlation functions. Before estimating the frequency response function the spectral estimates are 'recolored' so they correspond to the original variables.

Pukkila (1979) gives a description of the estimation procedure and discusses identification of three different transfer function models. In Appendix 1 the estimation procedure is outlined.

3.3 The ridge regression method

As argued above, the most straight–forward way of obtaining preliminary estimates of the transfer function weights, ν_j, is to estimate the weights by ordinary least–squares regression. The simple transfer function may be written as,

$$y_t = c + \nu_0 x_t + \nu_1 x_{t-1} + \nu_2 x_{t-2} + \dots + n_t$$

where y, x, c and n are defined as above. If it is reasonable to assume that the weights ν_j are ≈ 0 for some lag $k > K$ then the above equation can be estimated by *OLS* with $K+1$ independent variables (lags 0 to K of the x variable). The equation above is easily expanded if there are m independent input variables.

If we try to estimate the regression equation we encounter several problems. First, we need knowledge of the lag K at which the ν_j weights may safely be assumed to be zero. Using too many lags means that many degrees of freedom will be lost, using too few lags gives us the missing variables problem. Second, the correlation matrix of the independent variables (including lagged variables) may show strong multicollinearity. This collinearity may result from autocorrelated input variables or/and cross correlation between the input variables. The third major problem is the possibility of autocorrelated residuals, i.e. n_t is not white noise. This will decrease the efficiency of the regression estimates.

The first problem can usually be solved by finding some trade–off between the risk of including too few lags (gives biased estimates) and losing to many degrees of freedom. It seems wise to start with too many rather than too few lags in the model.

The multicollinearity problem is more difficult to solve. Liu and Hanssens (1982) transformed the x and y variables by a common filter. The filter was constructed to eliminate AR factors with roots close to one in the estimated $ARMA$

processes for the input variables. Edlund (1984, 1989a,b) used biased regression, ridge and principal components regression, to obtain more accurate estimates.

Autocorrelated residuals can be dealt with by using generalized least squares (*GLS*) instead of *OLS* (see Liu and Hanssens 1982, Rahiala 1986) or by transforming the x and y variables. The latter method has been used by Box and Tiao (1975) and by Edlund (1984, 1989a,b) who suggested and used the following two–step procedure in his simulation studies:

Step one: **identification, estimation and checking of the noise model and transformation of the input and output variables**

First the multiple regression model

$$y_t = c + \nu_{10}x_{1t} + \nu_{11}x_{1,t-1} + \ldots + \nu_{mK}x_{m,t-K} + n_t$$

is estimated by ridge regression, k computed according to Lawless and Wang (1976) or *RIDGM* (Dempster, Schatzoff and Wermuth, 1977). Then the estimated residuals are computed,

$$\hat{n}_t = y_t - \hat{c} - \sum_{i=1}^{m} \hat{\nu}_i(B)x_{it} = y_t - \hat{c} - \hat{\nu}_{10}x_{1t} - \hat{\nu}_{11}x_{1,t-1} - \ldots - \hat{\nu}_{mK}x_{m,t-K} \,.$$

The noise model, $\hat{n}_t = \frac{\hat{\theta}(B)}{\hat{\phi}(B)}\hat{a}_t$ is then identified and estimated using the standard Box–Jenkins procedure for *ARMA* models. The estimated operators are then used to transform the original x and y variables,

$$\hat{\theta}(B)y'_t = \hat{\phi}(B)y_t, \text{ all } t, \text{ and } \hat{\theta}(B)x'_{jt} = \hat{\phi}(B)x_{jt}, \; j = 1,\ldots,m, \text{ all } t.$$

Step two: **estimation of the impulse response function from the transformed variables y'_t and x'_{jt}**

In the second step the first equation is re–estimated, by ridge regression methods, using the transformed x and y variables,

$$y'_t = c + \nu_{10}x'_{1t} + \nu_{11}x'_{1,t-1} + \ldots + \nu_{mK}x'_{m,t-K} + a_t \,.$$

In the last equation the residuals $\{a_t\}$ should be almost white noise and by using biased regression the bad effects of multicollinearity will be decreased.

In case the estimated residuals are not white noise, step one could be repeated using the estimated values of $\hat{\nu}_{ij}$ in the last equation for calculating the residuals n_t. Step two is then performed again.

4 Empirical results

This section describes identification of transfer function models for the Swedish index for industrial production (*IIP*) based on Swedish financial data. The models have been identified by the three methods presented above. The models have been chosen to represent plausible relationships (with a satisfactory forecasting ability), as well as different number of variables. Another major concern has been availability of the data. The choice of models may be seen as a search for tentative models based on different number of leading indicators.

The following estimation procedures have been used:

Step 1: First the series have been plotted against time, see Fig 1–4. Then with the aid of autocorrelation functions suitable order of differencing have been decided for the variables.

Step 2: Using AUTOBOX PLUS (v 2.0 from AFS, Inc) univariate *ARIMA* models for all series have been identified and estimated. Then for each model, *Step* 3–6 are repeated:

Step 3: For the prewhitening estimator AUTOBOX has been used to prewhiten each input series (one at a time) and the output series, and then compute the estimated cross correlation function and the impulse response weights.

Step 4: For the cross–spectral estimator a *C* program has been developed (see Appendix 1 for details of the estimation method) and used to filter the series, compute relevant spectra, then re–color the spectra and estimate the impulse response weights and their standard errors. Different values of the spectral computation number and lag window length have been tried to derive reasonable estimates. The filters have been adopted from the univariate models from AUTOBOX (in *Step* 2).

Step 5: The ridge regression estimator has been computed in two steps. First the impulse response weights have been estimated by the two ridge estimators. Then the residuals have been computed and modeled using AUTOBOX. The estimated residual model has then been used to transform all variables. Back–casting has been used to improve the estimates of the starting values of each series. Then ridge estimation has again been used to estimate the impulse response weights. Again the residuals have been checked for significant autocorrelations.

Step 6: The estimated impulse response weights and the identified models have been compared to an estimated final transfer function model.

Fig. 1. Swedish Index of Industrial Production, *IIP*, for the period January 1960 – December 1979, 240 monthly observations.

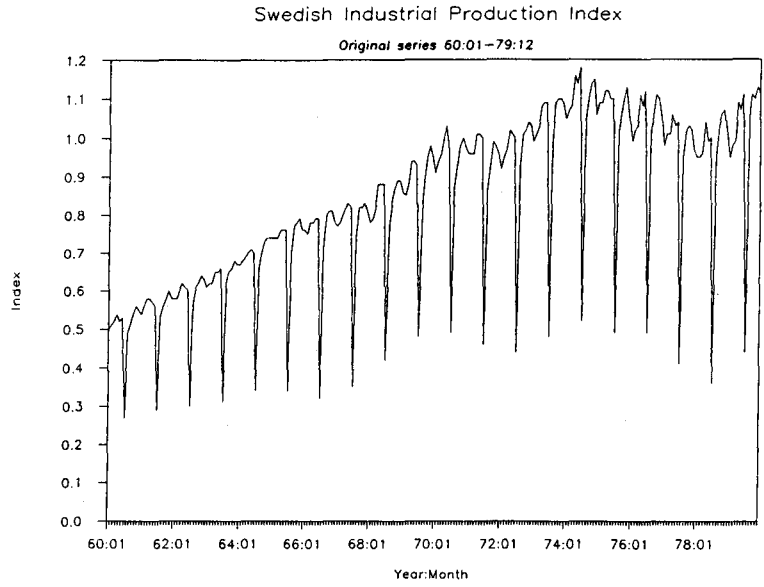


Fig. 2. Deflated Money Supply, *M*1, for the period January 1960 – December 1979, 240 monthly observations.

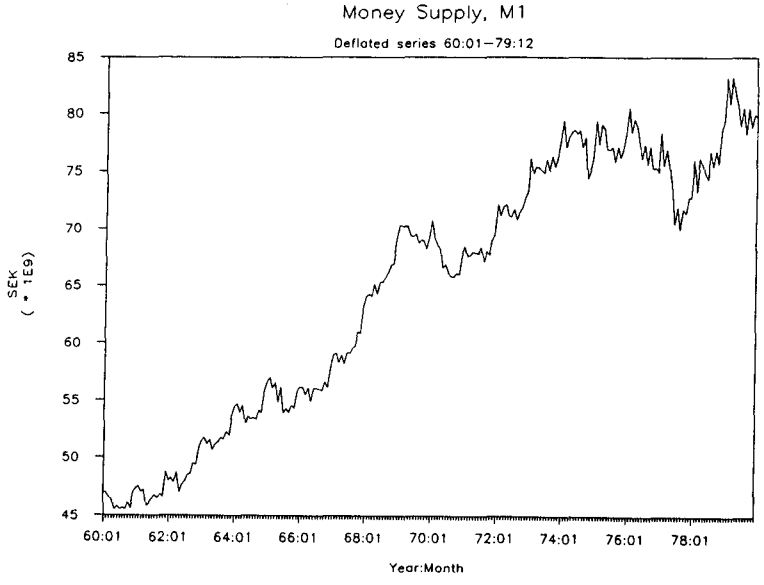


Fig. 3. Yield of long term government bonds for the period January 1960 – December 1979, 240 monthly observations.

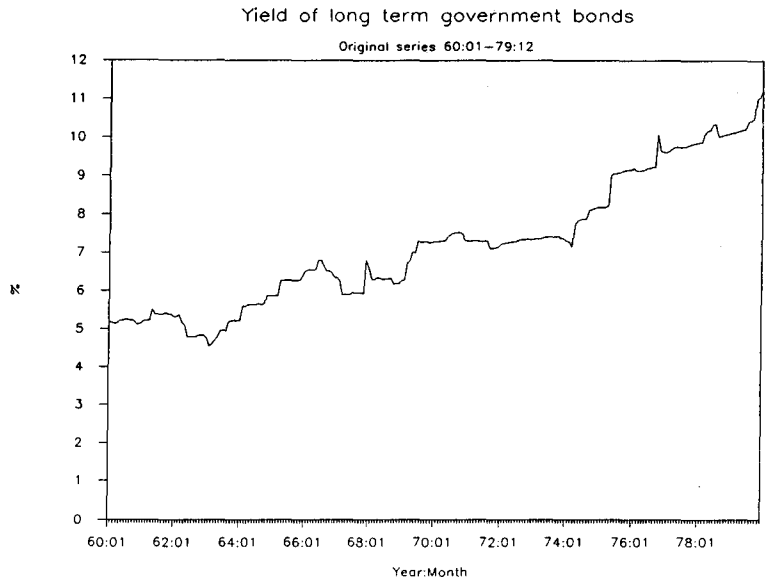
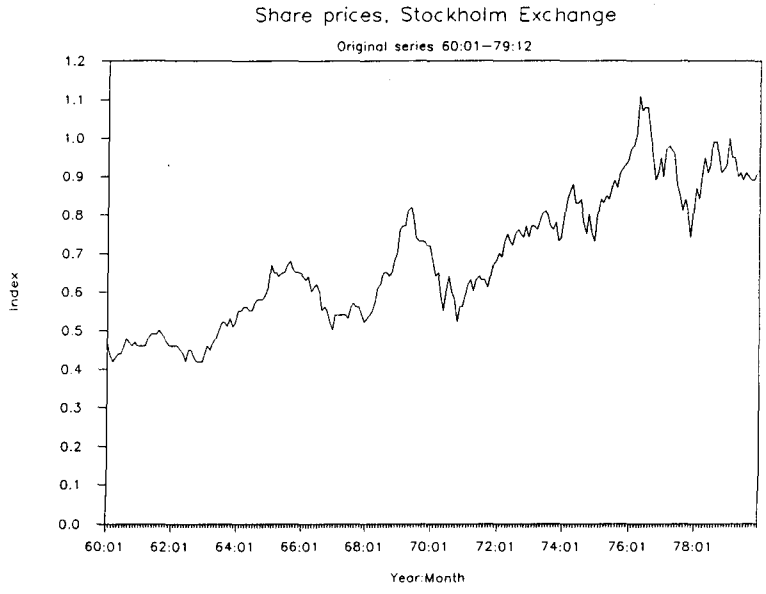


Fig. 4. Share prices at the Stockholm Exchange for the period January 1960 – December 1979, 240 monthly observations.



4.1 Leading indicators

The leading financial indicators chosen constitute the set of financial variables used by OECD to forecast the Swedish business cycle. Data have been collected from the Main Economic Indicators database at Statistics Sweden. The period January 1960 – December 1979 (240 monthly observations) was chosen as it appears to be a rather stable period. Results from OECD (1987) and Westlund and Claesson (1989) indicate that there is a considerable variation in the lead times at the turning points (The period 1960–1988 was studied) and estimated business cycle models (transfer functions) vary considerable both with respect to structure and estimated values for different subperiods. The following variables have been used:

$X_{1,t}$ = *M*1 (with minor adjustments) deflated by CPI (12 months)
$X_{2,t}$ = Yield of long term government bonds (15 months)
$X_{3,t}$ = Share prices at the Stockholm Exchange (8 months)

The values in parenthesis are lags at which the OECD found the maximum cross–correlation between the *IIP* and the indicator. It should be noted, that the OECD cross–correlation functions are computed from detrended and smoothed series. According to OECD (1987) *M*1 may be seen as an "expansionary accomodating stance of monetary policy". The *Yield* variable is an indicator of the "stimulus to consume or invest" and the *Share prices* variable act as a "business confidence proxy".

It seems reasonable to build models relating relative changes in the variables to the relative change in the *IIP*, Y_t. Therefore the first regular difference of the logarithm of each variable has been used. It should be noted that both Y_t and $X_{1,t}$ are seasonal series but that the other two series do not show seasonality.

Table 1 shows the estimated autocorrelation and partial autocorrelation functions for the $x_{i,t} = \nabla ln(X_{i,t})$ series. The following univariate *ARIMA* models were identified, estimated and checked:

$$\nabla_{12} y_t = (1-0.597B)(1-0.372B^{24})a_t, \quad \hat{\sigma}_a = 0.03111$$

$$(1+0.145B-0.304B^2-0.297B^4)\nabla_{12} x_{1t} = (1-0.755B^{12})a_{1t}, \quad \hat{\sigma}_{a_1} = 0.00847$$

$$x_{2t} = 0.00324 + a_{2t}, \quad \hat{\sigma}_{a_2} = 0.02001$$

$$x_{3t} = a_{3t}, \quad \hat{\sigma}_{a_3} = 0.03892$$

Table 1. Estimated autocorrelation and partial autocorrelation functions for $\nabla ln(Y_t)$, $\nabla ln(X_{1t})$, $\nabla ln(X_{2t})$ and $\nabla ln(X_{3t})$. $^+$ denotes a coefficient significant at the 5 % level.

Lag	$\nabla ln(Y)$ acf	pacf	$\nabla ln(X_1)$ acf	pacf	$\nabla ln(X_2)$ acf	pacf	$\nabla ln(X_3)$ acf	pacf
0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1	−.446$^+$	−.446$^+$	−.376$^+$	−.376$^+$	−.044	−.044	.057	.057
2	−.026	−.280$^+$	.368$^+$	.263$^+$	−.009	−.011	−.100	−.104
3	−.051	−.258$^+$	−.259$^+$	−.072	.007	.006	.100	.114
4	−.015	−.252$^+$	.287$^+$	.130$^+$	.022	.022	.006	−.019
5	.009	−.236$^+$	−.539$^+$	−.448$^+$	.077	.080	−.108	−.087
6	.058	−.145$^+$	.468$^+$	.238$^+$	−.014	−.006	.077	.081
7	.009	−.080	−.511$^+$	−.256$^+$	−.014	−.014	−.035	−.069
8	−.014	−.049	.274$^+$	−.039	.061	.059	−.034	.011
9	−.050	−.091	−.241$^+$	.064	−.071	−.070	.009	−.015
10	−.024	−.148$^+$	.293$^+$	−.048	−.023	−.034	.081	.082
11	−.418$^+$	−.929$^+$	−.338$^+$	.041	−.013	−.016	.050	.058
12	.941$^+$	.262$^+$	.748$^+$	.572$^+$	−.065	−.068	.016	.006
13	−.418$^+$	.160$^+$	−.338$^+$	.069	.008	−.004	−.061	−.063
14	−.028	.078	.262	−.260$^+$	.002	.015	−.060	−.064
15	−.046	.117	−.219	−.020	.012	.019	−.053	−.047
16	−.016	.125	.205	−.046	−.088	−.088	−.123	−.126
17	.009	.134$^+$	−.498$^+$	−.070	.102	.116	−.036	−.014
18	.052	.064	.378$^+$	−.069	.004	.007	−.051	−.076
19	.009	.026	−.479$^+$	−.147$^+$	.107	.109	−.189$^+$	−.174$^+$
20	−.012	.015	.183	−.134$^+$	−.084	−.072	−.010	−.007
21	−.048	.003	−.247	−.132$^+$	−.062	−.070	−.037	−.113
22	−.022	.012	.240	.022	−.056	−.095	−.097	−.065
23	−.389$^+$	.080	−.304	.028	−.011	−.027	−.038	−.069
24	.875$^+$	−.024	.648$^+$	.154$^+$	−.026	−.034	.102	.088

4.2 Money supply

For the first model, with one input, money supply, $M1$ has been chosen as leading indicator. This gives us the following model:

$$IIP_{SWE,t} = f(M1_{SWE,t}, \ldots, M1_{SWE,t-K}).$$

Prewhitening

The estimated *ARIMA* model for $M1$ in Section 4.1 was used to prewhiten the *IIP*. Table 2 shows the estimated impulse response weights for lag 0 to 23. The second largest coefficient and the only significant coefficient with the expected sign is at lag 7. This is a considerably shorter lead than the OECD maximum cross–

Table 2. Estimated impulse response weights, $\hat{\nu}$, between prewhitened *IIP* and filtered *Money supply*, *M*1. Weights larger (absolute value) than 1.64σ are marked with $^{+}$ and larger than 1.96σ with $^{++}$.

***Money supply*, x_{1t}**								
Lags 0–7	-.118	.104	.219	**-.883**$^{++}$	.209	.558	-.354	**.682**$^{+}$
Lags 8–15	-.485	.019	.226	.194	.252	.017	.209	**-.626**$^{+}$
Lags 16–23	-.114	.558	-.213	.163	-.037	.036	-.307	-.283

correlation at lag 12. There may be two explanations to this: the OECD cross–correlation function is estimated from detrended and smoothed series, not prewhitened series, and/or the pattern may have changed during the later part of the period. The tentatively identified model may therefore be expressed as:

$$y_t = \omega_0 x_{1,t-7} + N_t \quad \text{where } \nabla_{12} N_t = (1-\theta B)(1-\Theta B^{24}) a_t$$

with $\hat{\omega}_0 = 0.68$ as a preliminary estimate of ω_0.

Cross spectral

The cross spectral analysis was performed with $F = 60$ (spectral computation number) and a Parzen lag window with $L = 45$. To improve the spectral estimates both variables were transformed by autoregressive filters to become white noise. The *M*1 variable was lagged 7 months (aligned) to improve the estimates of the coherency function. The results are given in Table 3. The pattern of the weights is similar to the pattern in Table 2 and there is only one significant weight (at $\alpha = 0.1$) with the expected sign. In this case we would identify the same model as in the prewhitening case, but with $\hat{\omega}_0 = 0.63$ as our preliminary estimate of ω_0. Our tentative model therefore is:

$$y_t = \omega_0 x_{1,t-7} + N_t$$

where $\nabla_{12} N_t = (1-\theta B)(1-\Theta B^{24}) a_t$.

Ridge regression

The ridge estimates *LW* (Lawless and Wang) and *RIDGM* have been computed using the two–step method described in Section 3.3. In the first step the estimated residuals from both the *LW* and *RIDGM* estimators were best described by a model of the type:

Table 3. Estimated impulse response weights, $\hat{\nu}$, between *IIP* and *Money supply*, *M*1 using cross spectral analysis. Weights larger (absolute value) than 1.64σ are marked with $^{+}$ and larger than 1.96σ with $^{++}$.

***Money supply*, x_{1t}**								
Lags 0–7	-.005	.007	.436	**-.639**$^{+}$	-.087	.429	.102	**.632**$^{+}$
Lags 8–15	-.585	-.098	.324	.332	.288	-.306	-.096	-.032
Lags 16–23	-.073	.073	-.163	.387	-.065	-.268	.089	.319

$$n_t = (1-\theta B)(1-\Theta B)a_t.$$

The same model was also identified for all two– and three–input models below. Since the estimates of θ and Θ did not differ significantly between the models and estimators, the variables were in all cases transformed by the estimated model:

$$(1-0.66B)(1-0.34B^{24})y_t' = y_t \quad \text{and} \quad (1-0.66B)(1-0.34B^{24})x_{it}' = x_{it}.$$

In the second step one significant autocorrelation coefficient was found at lag 36 for both estimators. The estimated k values in the second step was 0.0731 and 5.1524 (!) for the *LW* and *RIDGM* estimators respectively.

As can be seen in Table 4 the estimated weights for the *RIDGM* are clearly much more shrunk towards 0 than the *LW* weights (even though the pattern is the same). Again the weight at lag 7 is significant with the expected sign, but the weights at lags 5 and 6 are also significant. In this case we may therefore tentatively identify the model:

$$y_t = (\omega_0 - \omega_1 B - \omega_2 B^2)x_{1,t-5} + N_t$$

where $\nabla_{12}N_t = (1-\theta B)(1-\Theta B^{24})a_t$.

Estimated models

Using AUTOBOX the proposed models were estimated. For the Ridge model, the ω_0 and ω_1 were non–significant, so the model was reduced to the same model as identified by the other two methods:

$$y_t = .632x_{1,t-7} + N_t.$$

The residual model, was expanded to account for autocorrelation in the a_t series at lags 12 and 36 to:

$$\nabla_{12}N_t = (1-.731B)(1-.259B^{12}-.501B^{24})a_t \quad \hat{\sigma}_a = .02850.$$

Table 4. Estimated impulse response weights, $\hat{\nu}$, between *IIP* and *Money supply*, *M*1 using ridge regression, *LW* and *RIDGM*. Weights larger (absolute value) than 1.64σ are marked with $^{+}$ and larger than 1.96σ with $^{++}$.

***Money supply*, x_{1t}**								
LW:								
Lags 0–7	.061	.059	.017	-.066	.038	**.144**$^{++}$	**.120**$^{+}$	**.204**$^{++}$
Lags 8–15	-.010	.019	.067	.054	.068	-.006	.005	.002
Lags 16–23	-.017	.006	.021	.011	.078	.028	-.083	.104
RIDGM:								
Lags 0–7	.020	.023	.015	.007	**.029**$^{+}$	**.054**$^{++}$	**.052**$^{++}$	**.067**$^{++}$
Lags 8–15	.025	.026	**.032**$^{+}$	.026	.028	.015	.015	.012
Lags 16–23	.002	-.001	-.000	-.003	.012	.003	-.019	.014

The Θ_1 and Θ_2 parameters are strongly correlated, but deleting one of them would give autocorrelated residuals. The estimated model, has two just significant cross–correlations at lags 4 and 6 between the estimated residuals and the prewhitened input variable. This indicates a possible two–way causality. The reduction in residual variance compared to the univariate model for *IIP* is approximately 16 %, i.e. most of the information in *M*1 is already contained in the *IIP* series. When we compare the different estimators, it is interesting to note that all of them had a significant weight at lag 7 but the prewhitening and the cross–spectral estimators turned up significant weights with wrong signs.

4.3 Money supply and Yield of long term government bonds

As the second model for *IIP* we choose a two–input model with money supply, M1, and Yield of long term government bonds as leading indicators. This gives us the following model:

$$IIP_{SWE,t} = f(M1_{SWE,t}\cdots,M1_{SWE,t-K}, YB_{SWE,t}\cdots YB_{SWE,t-K}).$$

Prewhitening

Since we prewhiten each input variable and *IIP* separately the estimated weights for *M*1 are the same as in the previous Section. The *Yield* variable is not seasonal,

and using the developed model for x_{2t} in Section 4.1 the seasonality in y_t is not considered. The estimated weights are given in Table 5. They do not indicate any significant relationship with y_t. Therefore x_{2t} was differenced seasonally and a model was estimated for the over–differenced series. This model:

$$\nabla_{12} x_{2t} = (1-.968B^{12})a_{2t}$$

was then used to prewhiten both x_{2t} and y_t.

As can be seen in Table 5 estimates more in line with the other estimators (Tables 6 and 7) are found. The significant coefficient at lag 20 is not too far from the OECD maximum cross–correlation at lag 15. The following tentative model may now be identified:

$$y_t = \omega_{10} x_{1,t-7} + \omega_{20} x_{2,t-20} + N_t$$

with $\nabla_{12} N_t = (1-\theta B)(1-\Theta B^{24})a_t$ as before.

Cross spectral

The *Yield* variable was transformed in the same way as the *Money* and *IIP* variables to be white noise. The estimated weights are shown in Table 6. The significant weight for *M*1 at lag 7 has now vanished. The *Yield* variable has a significant weight at lag 1 month. This is far away from the OECD lag of 15 months. Our model is then reduced to:

$$y_t = \omega_{20} x_{2,t-1} + N_t$$

with $\nabla_{12} N_t = (1-\theta B)(1-\Theta B^{24})a_t$ as before.

Ridge regression

The *LW* and *RIDGM* estimates in Table 7 for *M*1 is close to the bivariate model above. It therefore seems reasonable to use the estimated model from Section 4.2. For the *Yield* variable the weight at lag 1 is most significant but the significant weights at lag 20 and 21 are closer to the OECD lag.

We therefore suggest:

$$y_t = \omega_{10} x_{1,t-7} + (\omega_{20} - \omega_{21} B) x_{2,t-20} + N_t$$

with $\nabla_{12} N_t = (1-\theta B)(1-\Theta B^{24})a_t$ as a starting model.

Table 5. Estimated impulse response weights, $\hat{\nu}$, between prewhitened *IIP* and filtered *Money supply*, *M*1 and between prewhitened *IIP* and *Yield of long term government bonds* (prewhitening with and without respect to seasonality in *IIP*). Weights larger (absolute value) than 1.64σ are marked with $^{+}$ and larger than 1.96σ with $^{++}$.

***Money supply*, x_{1t}**								
Lags 0–7	-.118	.104	.219	**-.883**$^{++}$	.209	.558	-.354	**.682**$^{+}$
Lags 8–15	-.485	.019	.226	.194	.252	.017	.209	**-.626**$^{+}$
Lags 16–23	-.114	.558	-.213	.163	-.037	.036	-.307	-.283
***Yield of long term government bonds*, x_{2t}** (prewhitening without regard to seasonality)								
Lags 0–7	-1.115	.284	-1.364	-.140	1.434	.529	.039	-1.095
Lags 8–15	1.018	-1.205	.741	.870	-1.048	.214	-1.526	-.085
Lags 16–23	1.354	.484	.153	-.987	.808	-1.408	.869	.529
***Yield of long term government bonds*, x_{2t}** (prewhitening with regard to seasonality)								
Lags 0–7	-.048	-.246	.303	**-.321**$^{+}$	.047	.144	.179	-.156
Lags 8–15	-.288	.115	.215	.084	-.093	-.183	.136	-.279
Lags 16–23	-.044	.136	.274	-.047	**-.513**$^{++}$	-.086	**.350**$^{+}$	.010

Table 6. Estimated impulse response weights, $\hat{\nu}$, between *IIP* and *Money supply* and *Yield of long term government bonds* using cross spectral analysis. Weights larger (absolute value) than 1.64σ are marked with $^{+}$ and larger than 1.96σ with $^{++}$.

***Money supply*, x_{1t}**								
Lags 0–7	-.082	.062	.377	**-.637**$^{+}$	-.099	.538	.153	.408
Lags 8–15	-.536	-.172	.274	.451	.253	-.091	-.216	.030
Lags 16–23	-.015	.045	-.176	.288	.071	-.343	.018	.323
***Yield of long term government bonds*, x_{2t}**								
Lags 0–7	-.062	**-.247**$^{+}$	.134	-.072	.166	-.163	-.052	.132
Lags 8–15	-.001	.006	.079	-.095	-.059	-.148	.093	-.110
Lags 16–23	.160	-.084	.078	.155	-.175	-.064	.121	-.022

Table 7. Estimated impulse response weights, $\hat{\nu}$, between *IIP* and *Money supply* and *Yield of long term government bonds* using ridge regression, *LW* and *RIDGM*. Weights larger (absolute value) than 1.64σ are marked with $^{+}$ and larger than 1.96σ with $^{++}$.

Money supply, x_{1t}								
LW:								
Lags 0–7	.056	.063	.022	-.076	.037	**.146**$^{++}$	**.115**$^{+}$	**.205**$^{++}$
Lags 8–15	-.005	.034	.090	.062	.079	.007	.011	.006
Lags 16–23	-.019	.006	.018	-.004	.054	.004	-.115	.089
RIDGM:								
Lags 0–7	.033	.039	.020	-.014	.040	**.097**$^{++}$	**.087**$^{++}$	**.128**$^{++}$
Lags 8–15	.027	.039	.059	.043	.049	.015	.018	.013
Lags 16–23	-.003	-.001	.003	-.005	.026	.002	-.055	.037
Yield of long term government bonds, x_{2t}								
LW:								
Lags 0–7	-.034	**-.064**$^{++}$	.034	.008	.040	.002	-.003	.036
Lags 8–15	-.021	.018	.028	.026	-.022	.005	-.016	-.001
Lags 16–23	.007	.018	**.051**$^{+}$	.020	**-.053**$^{+}$	**-.050**$^{+}$	.021	-.011
RIDGM:								
Lags 0–7	-.019	**-.034**$^{++}$	.015	.007	.020	.001	-.004	.017
Lags 8–15	-.003	.015	.016	.017	-.006	.006	-.011	-.004
Lags 16–23	.002	.011	**.026**$^{+}$	.008	**-.028**$^{+}$	**-.028**$^{+}$	.004	-.011

Estimated models

As suggested by the identification above three models were estimated. The model suggested by the cross–spectral estimator was not significant, i.e. the model reduced to a univariate model for *IIP*. In the model suggested by the ridge estimators, ω_{20} was not significant. This gives us the following two models:

$$y_t = .689x_{1,t-7} - .129x_{2,t-20} + N_t,$$

$$\nabla_{12}N_t = (1-.748B)(1-.337B^{12}-.599B^{24})a_t \quad \hat{\sigma}_a = .02788$$

and

$$y_t = .707x_{1,t-7} - .135x_{2,t-21} + N_t,$$

Table 8. Estimated impulse response weights, $\hat{\nu}$, between prewhitened *IIP* and filtered *Money supply, M*1 and between prewhitened *IIP* and *Yield of long term government bonds* (prewhitening with respect to seasonality in *IIP*) and between prewhitened *IIP* and filtered *Share prices* (prewhitening with respect to seasonality in *IIP*). Weights larger (absolute value) than 1.64σ are marked with $^{+}$ and larger than 1.96σ with $^{++}$.

Money supply, x_{1t}								
Lags 0–7	-.118	.104	.219	**-.883**$^{++}$	.209	.558	-.354	**.682**$^{+}$
Lags 8–15	-.485	.019	.226	.194	.252	.017	.209	**-.626**$^{+}$
Lag 16–23	-.114	.558	-.213	.163	-.037	.036	-.307	-.283
Yield of long term government bonds, x_{2t} (prewhitening with respect to seasonality)								
Lags 0–7	-.048	-.246	.303	**-.321**$^{+}$	.047	.144	.179	-.156
Lags 8–15	-.288	.115	.215	.084	-.093	-.183	.136	-.279
Lags 16–23	-.044	.136	.274	-.047	**-.513**$^{++}$	-.086	**.350**$^{+}$	.010
Share prices, x_{3t} (prewhitening with respect to seasonality)								
Lags 0–7	-.094	-.083	.109	.046	**-.169**$^{+}$	.020	.094	.079
Lags 8–15	**.170**$^{+}$	**-.198**$^{++}$	-.107	**.250**$^{++}$	-.136	-.081	**.154**$^{+}$	.021
Lags 16–23	**-.195**$^{++}$	-.018	**.196**$^{++}$	-.010	**.153**$^{+}$	-.135	-.145	**.175**$^{+}$

$$\nabla_{12}N_t = (1-.759B)(1-.329B^{12}-.613B^{24})a_t \quad \hat{\sigma}_a = .02778$$

Again, an extra parameter has been added to the residual model at lag 12. The two seasonal *MA* coefficients are strongly correlated, but excluding one of them introduces autocorrelation in the residuals.

Both models show a few significant correlations between a_t and the prewhitened x_{it}, but on the whole they seem to be adequate models for the data series. The reduction in residual variance compared to the univariate model for *IIP* is approximately 20 %, i.e. a minor reduction compared with the model with just *M*1 as input variable. Comparing the different estimators, it is interesting to note that the cross spectral estimator missed the significant weights at lag 20 and 21.

4.4 Money supply, Yield of long term government bonds and Share prices

The third model for *IIP* is a three–input model with money supply, *M*1, Yield of long term government bonds and Share prices at the Stockholm Exchange as leading

Table 9. Estimated impulse response weights, $\hat{\nu}$, between *IIP* and *Money supply*, *M*1, *Yield of long term government bonds* and *Share prices* using cross spectral analysis. Weights larger (absolute value) than 1.64σ are marked with $^{+}$ and larger than 1.96σ with $^{++}$.

***Money supply*, x_{1t}**								
Lags 0–7	.027	-.053	.403	**-.648**$^{+}$	.014	.524	-.020	.488
Lags 8–15	-.508	-.230	.102	.559	.521	-.389	-.275	.240
Lags 16–23	-.048	-.033	-.322	.473	.020	-.430	.129	.261
***Yield of long term government bonds*, x_{2t}**								
Lags 0–7	-.136	**-.295**$^{+}$	.160	.022	.165	-.226	-.048	.146
Lags 8–15	.045	-.062	.022	.036	-.110	-.201	.118	-.048
Lags 16–23	.154	-.198	.114	.136	-.152	-.046	.048	.055
***Share prices*, x_{3t}**								
Lags 0–7	-.038	.073	-.094	.004	.054	-.007	-.053	-.067
Lags 8–15	**.201**$^{++}$	-.012	**-.141**$^{+}$	.082	-.035	.118	**-.149**$^{+}$	.008
Lags 16–23	.033	-.047	.012	-.054	**.147**$^{+}$	.029	-.113	.055

indicators. This gives us the following model

$$IIP_{SWE,t} = f(M1_{SWE,t},\ldots,M1_{SWE,t-K}, YB_{SWE,t},\ldots,YB_{SWE,t-K}, SP_{SWE,t},\ldots,SP_{SWE,t-K}).$$

Prewhitening

As with the *Yield* variable the *Share prices* variable has been seasonally differenced to take care of the seasonality in *IIP*. The following model:

$$\nabla_{12}x_{3t} = (1-.965B^{12})a_{3t}$$

was used to prewhiten x_{3t} and y_t. The results are given in Table 8.

With the OECD lag of 8 months for the *Share* variable the weights at lag 8 and 11 are most interesting. Since the weight at lag 11 is much larger than the weight at lag 8 the following model is proposed:

$$y_t = \omega_{10}x_{1,t-7} + \omega_{20}x_{2,t-20} + \omega_{30}x_{3,t-11} + N_t$$

with $\nabla_{12}N_t = (1-\theta B)(1-\Theta B^{24})a_t$ as before.

Table 10. Estimated impulse response weights, $\hat{\nu}$, between *IIP* and *Money supply, M*1, *Yield of long term government bonds* and *Share prices* using ridge regression, *LW* and *RIDGM*. Weights larger (absolute value) than 1.64σ are marked with $^{+}$ and larger than 1.96σ with $^{++}$.

Money supply, x_{1t}								
LW:								
Lags 0–7	.064	.081	.044	-.064	.040	**.135**$^{++}$	.113	**.201**$^{++}$
Lags 8–15	-.002	.024	.065	.062	.064	-.009	-.001	.008
Lags 16–23	-.029	-.004	-.002	-.017	.044	.004	-.111	-.074
RIDGM:								
Lags 0–7	.035	.041	.025	-.006	.038	**.084**$^{++}$	**.078**$^{++}$	**.111**$^{++}$
Lags 8–15	.025	.030	.043	.034	.036	.007	.010	.010
Lags 16–23	-.008	-.005	-.004	-.010	.019	.003	-.044	.031
Yield of long term government bonds, x_{2t}								
LW:								
Lags 0–7	-.027	**-.059**$^{++}$	.028	.005	.035	-.001	-.006	.027
Lags 8–15	-.015	.018	.021	.016	-.022	.003	-.015	-.000
Lags 16–23	.003	.008	.040	.024	**-.049**$^{+}$	**-.051**$^{+}$	.016	-.008
RIDGM:								
Lags 0–7	-.017	**-.030**$^{++}$	.010	.003	.014	-.001	-.006	.011
Lags 8–15	-.002	.012	.012	.012	-.006	.004	-.009	-.003
Lags 16–23	.001	.007	.021	.008	**-.024**$^{+}$	**-.025**$^{+}$	.002	-.010
Share prices, x_{3t}								
LW:								
Lags 0–7	.022	-.004	-.021	-.020	.001	.011	.002	.008
Lags 8–15	.002	.013	.015	.009	-.010	.018	-.009	.010
Lags 16–23	-.013	-.006	.002	-.004	**.029**$^{++}$	**.031**$^{++}$	-.010	-.000
RIDGM:								
Lags 0–7	.011	-.001	-.009	-.009	.002	.006	.004	.006
Lags 8–15	.001	.005	.010	.007	-.003	.009	-.001	.007
Lags 16–23	-.005	-.002	.000	-.002	**.013**$^{+}$	**.014**$^{+}$	-.006	-.004

Cross spectral

The cross spectral estimates of the weights for the *M*1 and *Yield* variables in Table 9 are approximately the same as in the two–input model. The *Share* variable has one significant weight with the expected sign at lag 8 which is the same as the OECD lag. Therefore the following model is proposed:

$$y_t = \omega_{20}x_{2,t-1} + \omega_{30}x_{3,t-8} + N_t$$

where $\nabla_{12}N_t = (1-\theta B)(1-\Theta B^{24})a_t$ as before.

Ridge regression

Looking at Table 10 the ridge estimates for *M*1 and *Yield* suggest that the previously identified model still may be appropriate. For the *Share* variable the weights are more difficult to explain. There are no significant weights around the OECD lag at 8 months, the significant weights at lags 20–21 are difficult to explain. In this case the proposed model is:

$$y_t = \omega_{10}x_{1,t-7} + (\omega_{20}-\omega_{21}B)x_{2,t-20} + (\omega_{30}-\omega_{31}B)x_{3,t-20} + N_t$$

with $\nabla_{12}N_t = (1-\theta B)(1-\Theta B^{24})a_t$ as usual.

Estimated models

Using AUTOBOX the proposed models were estimated. In the model identified using prewhitening the weight for the *Share* variable was not significant. The prewhitening model was reduced to the two–input model. In the cross–spectral model none of the weights were significant. Estimation of the ridge regression model showed that the second weights for the *Yield* variable and the *Share* variable were insignificant. Re–estimation of the model with only one weight for each variable gave the following results:

$$y_t = .826x_{1,t-7} - .132x_{2,t-20} + .074x_{3,t-20} + N_t,$$
$$\nabla_{12}N_t = (1-.772B)(1-.339B^{12}-.605B^{24})a_{1t} \quad \hat{\sigma}_a = .02732$$

The model seems to be an adequate model for the data series. The reduction in residual variance compared to the univariate model for *IIP* is approximately 23 %.

5 Conclusions

It is not an easy task to build a transfer function model for the Swedish Index of Industrial Production. It appears that the dynamic structures are not as stable as we had hoped for. The obtained delays in the models are not in concordance with the OECD maximum cross–correlation lags. These deviations may, at least in part, be explained by the different transformations used by the OECD (detrending and smoothing).

The estimated *ARIMA* model for *IIP* and the preferred one, two and three–input models are:

$$\nabla_{12} y_t = (1-.597B)(1-.372B^{24})a_t, \quad \hat{\sigma}_a = .03111$$

$$y_t = .632x_{1,t-7} + N_t, \quad \hat{\sigma}_a = .02850$$

$$y_t = .707x_{1,t-7} - .135x_{2,t-21} + N_t, \quad \hat{\sigma}_a = .02778$$

$$y_t = .826x_{1,t-7} - .132x_{2,t-20} + .074x_{3,t-20} + N_t, \quad \hat{\sigma}_a = .02732$$

with the noise model being of the form $\nabla_{12} N_t = (1-\theta B)(1-\Theta_1 B^{12}-\Theta_2 B^{24})a_t$. The reduction of residual variance is moderate as new variables are introduced, approximately 23 % for the three–input model. This may be an indication that the input variables have a rather predictable pattern that is included in the historical values of *IIP*, or it may be a sign of a poor model. The lead times are long enough to allow for reasonably safe predictions of the *IIP*; the forecast error is approximately 2.7 %. The *M*1 has a relatively strong influence on the *IIP*; a 1 % increase in the real money supply will result in 0.8 % increase in the production volume, while the *IIP* is much less sensitive to changes in the Yield of long term government bonds and Share prices.

There is a large variation in the number of significant weights for the estimators (Tables 8, 9 and 10):

Estimator	***M*1**	***Yield***	***Share***	**Total**
Prewhitening	3	3	9	15
Cross spectral	1	1	4	6
Ridge regression				
Lawless & Wang	2	3	2	7
RIDGM	3	3	2	8

What conclusions can be drawn from this table? For this limited sample of transfer functions it seems as if the prewhitening estimator will give too many significant weights, which may be suspected as the estimator does not consider the simultaneous

relationships between the input variables. The cross spectral estimator on the other hand did not find the significant weights for the *Yield* and *Share* variables. In this study the ridge estimators have been somewhere in between and found all significant models. Among the ridge estimators, the *LW* gives more reasonable estimates of the weights while the *RIDGM* shrinks the estimates too much compared to the other estimators. The estimated k value for the *RIDGM* estimator varied between 5.15 and 1.79, while the *LW* k value was around .07 which is a more reasonable level.

From this limited experience it would be recommendable to use more than one method for identification to find the weights that are to be accounted for in the model and those who are spurious.

Appendix 1

Estimation of transfer function weights by cross spectral methods

The following description of the cross spectral estimator is based on Box and Jenkins (1976, p 413–6) and Pukkila (1979).

We begin with the stationary m–input transfer function model,

$$y_t = \nu_1(B)x_{1,t} + \cdots + \nu_m(B)x_{m,t} + n_t \tag{1}$$

where $\nu_j(B)$ is the generating function of the impulse response weights relating $x_{j,t}$ to the output. After multiplying Eq (1) throughout by $x_{1,t-k},\ldots,x_{m,t-k}$ in turn, and taking expectations an equation system of generating functions is obtained (see Eq (A11.1.11) in Box and Jenkins). After substituting $B = e^{-i\lambda}$ and dividing by 2π the spectral equations are obtained,

$$\begin{cases} p_{x_1y}(\lambda) = H_1(\lambda)p_{x_1x_1}(\lambda) + H_2(\lambda)p_{x_1x_2}(\lambda) + \cdots + H_m(\lambda)p_{x_1x_m}(\lambda) \\ p_{x_2y}(\lambda) = H_1(\lambda)p_{x_2x_1}(\lambda) + H_2(\lambda)p_{x_2x_2}(\lambda) + \cdots + H_m(\lambda)p_{x_2x_m}(\lambda) \\ \quad\vdots \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \vdots \\ p_{x_my}(\lambda) = H_1(\lambda)p_{x_mx_1}(\lambda) + H_2(\lambda)p_{x_mx_2}(\lambda) + \cdots + H_m(\lambda)p_{x_mx_m}(\lambda) \end{cases} \tag{2}$$

where $p_{x_ix_j}$ is the cross spectra between x_i and x_j. $H_k(\lambda)$ is the frequency response function where,

$$H_k(\lambda) = \sum_{j=0}^{\infty} \nu_{kj} e^{-i\lambda j}, \quad k = 1,\ldots,m \tag{3}$$

The impulse response weights can now be computed from,

$$\nu_{kj} = \frac{1}{2\pi} \int_{-\pi}^{\pi} H_k(\lambda) e^{i\lambda j} d\lambda, \quad k = 1,...,m, \; j = 0, 1, 2, ... \tag{4}$$

The system of cross spectral equations above can be solved to obtain $H_k(\lambda)$.
Let,

$$\begin{bmatrix} p_{x_1x_1}(\lambda) & p_{x_1x_2}(\lambda) & \cdots & p_{x_1x_m}(\lambda) \\ p_{x_2x_1}(\lambda) & p_{x_2x_2}(\lambda) & \cdots & p_{x_2x_m}(\lambda) \\ \vdots & & & \vdots \\ p_{x_mx_1}(\lambda) & p_{x_mx_2}(\lambda) & \cdots & p_{x_mx_m}(\lambda) \end{bmatrix} =$$

$$\begin{bmatrix} c_{x_1x_1}(\lambda) & c_{x_1x_2}(\lambda) & \cdots & c_{x_1x_m}(\lambda) \\ c_{x_2x_1}(\lambda) & c_{x_2x_2}(\lambda) & \cdots & c_{x_2x_m}(\lambda) \\ \vdots & & & \vdots \\ c_{x_mx_1}(\lambda) & c_{x_mx_2}(\lambda) & \cdots & c_{x_mx_m}(\lambda) \end{bmatrix} - i \begin{bmatrix} q_{x_1x_1}(\lambda) & q_{x_1x_2}(\lambda) & \cdots & q_{x_1x_m}(\lambda) \\ q_{x_2x_1}(\lambda) & q_{x_2x_2}(\lambda) & \cdots & q_{x_2x_m}(\lambda) \\ \vdots & & & \vdots \\ q_{x_mx_1}(\lambda) & q_{x_mx_2}(\lambda) & \cdots & q_{x_mx_m}(\lambda) \end{bmatrix}$$

$$= \boldsymbol{C}(\lambda) - i\boldsymbol{Q}(\lambda)$$

where $c_{x_ix_j}(\lambda)$ is the co–spectrum of x_i and x_j and $q_{x_ix_j}(\lambda)$ is the quadrature spectrum of x_i and x_j. Further let $\boldsymbol{c}(\lambda)$ be a column vector with co–spectra of x_i and y and $\boldsymbol{q}(\lambda)$ be a column vector with quadrature spectra of x_i and y. Then let,

$$\boldsymbol{D}(\lambda) = [\boldsymbol{C}(\lambda) + \boldsymbol{Q}(\lambda)\boldsymbol{C}^{-1}(\lambda)\boldsymbol{Q}(\lambda)]^{-1} \tag{5}$$

$$\boldsymbol{E}(\lambda) = -\boldsymbol{D}(\lambda)\boldsymbol{Q}(\lambda)\boldsymbol{C}^{-1}(\lambda)$$

Now,

$$\boldsymbol{H}(\lambda) = [\boldsymbol{D}(\lambda)\boldsymbol{c}(\lambda) - \boldsymbol{E}(\lambda)\boldsymbol{q}(\lambda)] - i[\boldsymbol{D}(\lambda)\boldsymbol{q}(\lambda) + \boldsymbol{E}(\lambda)\boldsymbol{c}(\lambda)] \tag{6}$$

where $\boldsymbol{H}(\lambda)$ is a vector with $H_1(\lambda),..., H_m(\lambda)$.

Cross spectral estimates of ν_{ij} (Eq (4)) are obtained from,

$$\hat{\nu}_{ij} = \frac{1}{2F}\hat{u}_{ij}(0) + \frac{1}{F}\sum_{k=1}^{F-1}\hat{u}_{ij}(\lambda_k) + \frac{1}{2F}\hat{u}_{ij}(\pi), \tag{7}$$

$$i = 1,...,m\,; \; j = 0, 1, 2$$

where $\lambda_k = \pi k/F$, $k = 0, 1,...,F$ (F is the spectral computation number) and $\hat{u}_{ij}(\lambda)$ is computed from

$$\hat{u}_{ij}(\lambda) = [\sum_{k=1}^{m}(\hat{d}_{ik}(\lambda)\hat{c}_{ky}(\lambda) - \hat{e}_{ik}(\lambda)\hat{q}_{ky}(\lambda))]\cos(\lambda j)$$

$$+ [\sum_{k=1}^{m}(\hat{d}_{ik}(\lambda)\hat{q}_{ky}(\lambda) + \hat{e}_{ik}(\lambda)\hat{c}_{ky}(\lambda))]\sin(\lambda j), \; i = 1,...,m. \tag{8}$$

$\hat{c}_{ij}(\lambda)$, $\hat{q}_{ij}(\lambda)$, $\hat{d}_{ij}(\lambda)$ and $\hat{e}_{ij}(\lambda)$ are elements from the estimated c, q, D and E matrices as defined above.

References

Box, G.E.P. and Jenkins, G.M. (1976). *Time series analysis: Forecasting and control.* 2:nd ed. San Francisco: Holden–Day.

Box, G.E.P. and Tiao, G.C. (1975). "Intervention analysis with applications to economic and environmental problems," *JASA*, *70*, 70–79.

Damsleth, E. (1979). Analysis of multi–input transfer function models when the inputs are correlated. Publ. no 641. Norwegian Computing Center, Oslo, Norway.

Dempster, A.P., Schatzoff, M. and Wermuth, N. (1977). "A simulation study of alternatives to ordinary least squares," *JASA*, *72*, 77–91.

Edlund, P.–O. (1984). "Identification of the multi–input Box–Jenkins transfer function model," *Journal of Forecasting*, *3*, 297–308.

Edlund, P.–O. (1987). On identification of transfer function models. Research Report 1987:2. Dept. of Statistics, University of Stockholm, Stockholm, Sweden.

Edlund, P.–O. (1989a). "On identification of transfer function models by biased regression methods," *J Statist Comput Simul*, *31*, 131–148.

Edlund, P.–O. (1989b). Ridge estimation of transfer function weights. Research Paper 6380, February 1989, The Econonomic Research Institute at the Stockholm School of Economics.

Fask, A. and Robinson, P.B. (1977). "Identification of multivariate dynamic econometric models," *American Statistical Association* 1977 Proceedings of the Business and Economic Statistics Section, 284–289.

Haugh, L.D. and Box, G.E.P. (1977). "Identification of dynamic regression (distributed lag) models connecting two time series," *JASA*, *72*, 121–130.

Koch, P.D. and Rasche, R.H. (1988). "An examination of the Commerce department leading–indicator approach," *Journal of Business & Economic Statistics*, *6*, 167–187.

Lawless, J.F. and Wang, P. (1976). "A simulation study of ridge and other regression estimators," *Comm. in Statistics*, *A5*, 307–323.

Liu, L.–M. and Hanssens, D. (1982). "Identification of multiple–input transfer function models," *Comm. in Statistics*, *B11*, 297–314.

McLeod, G. (1982). *Box–Jenkins in practice, Volume 1*, Lancaster: G Jenkins & Partners Ltd.

OECD (1987). OECD leading indicators and business cycles in member countries 1960–1985. OECD Dept. of Economics and Statistics, Main Economic

Indicators, Sources and Methods, no 39, January 1987.

Priestley, M.B. (1971). "Fitting relationships between time series," *Bull. Inst. Internat. Statist., 38*, 1–27, (paper presented at the 38:th session of the I.S.I., Washington, USA).

Pukkila, T. (1979). On the identification of transfer function noise models with several correlated input processes using frequency domain tools. Acta Universitatis Tamperensis, ser A, vol 110, Tampere, Finland.

Pukkila, T. (1980). On transfer function noise model identification, Dept. of Mathematical Sciences, Univ. of Tampere, Finland, (paper presented to the 8:th Nordic conference on Mathematical Statistics held on Ahvenanmaa 26–29.5 1980).

Rahiala, M. (1986). "Identification and preliminary estimation in linear transfer function models," *Scand J Statist, 13*, 239–255.

Tsay, R.S. (1985). "Model identification in dynamic regression (distributed lag) models," *Journal of Business & Economic Statistics, 3*, 228–237.

Westlund, A.H. and Claesson, L. (1989). On the estimation of time–varying parameters in transfer function models, (paper presented at the International Conference on "Economic Structural Change; Analysis and Forecasting", Stockholm).

EFI publications since1986:

Published in the language indicated by the title

Albert, L., Fredriksson, O. Effektivisering av golvbranschen. En studie av struktur och lönsamhet samt några ideer till förändringar. Research report.

Almkvist, B., Företagspåverkan och insatsvilja. Research report

Andersson, B., FoU-verksamhet och resursallokering. Research report

Aspling, A., Företagsdemokrati och MBL.

Bergman, L., Åkerman, J., Förbränning eller återvinning? En samhällsekonomisk analys av olika sätt att behandla hushålls-
avfall. Research report

Björkegren, D., Företagsledarutbildning - en fallstudie.

Brunsson , N., Politik och ekonomi. En kritik av rationalitet som samhällsföreställning. Doxa Ekonomi.

Brytting, T., Löwstedt, J., Organisationsfrihet, visst finns den!

Davidsson, P., Tillväxt i små företag - en pilotstudie om tillväxtvilja och tillväxtförutsättningar i små företag.

Hammarkvist, K-O., Distribution i förändring - tankar kring informationsteknologi och effektivitet.

Holmberg, I., Företagsledares mandat - ett koncernledningsuppdrag påbörjas. EFI/Studentlitteratur

Hörngren, L., On Monetary Policy and Interest Rate Determination in an Open Economy.

Hörngren, L., Vredin, A., The Foreign Exchange Risk Premium: A review and some evidence from a currency basket system. Research report

Krusell, P., The Demand for Money in Sweden 1970 - 1983. Research report

Lundgren, A., Handlingar och strukturer. En organisationsteoretisk studie av styrelse-arbetets villkor i stora företag.

Löwstedt, J., Automation eller kunskapsproduktion?

Mufindiprojektet, working paper:

Håkansson L., (ed) Planning Processes, Environment and Socio-Economic Impact.

Rombach, B., Rationalisering eller prat. Kommuners anpassning till en stagnerande ekonomi. Doxa Ekonomi

Rydqvist, K., The Pricing of Shares with Different Voting Power and the Theory of Oceanic Games.

Thimrén, C., Imperfectly Competitive Markets for Exhaustible Resources - A Game-Theoretic Approach. Research report

Westlund, A., Törnqvist, B., On the Identification of Time for Structural Changes by Mosumsq and Cusumsq Procedures. Research report

1987

Andersson, T., Ternström, B., Private Foreign Investments and Welfare Effects: A comparative study of five countries in southeast Asia. Research report

Andersson, T., Ternström, B., External Capital and Social Welfare in South-East Asia. Research report

Benndorf, H., Marknadsföringsplanering och samordning mellan företag i industriella system. EFI/MTC

Bergman, L., Mäler, K-G., Ståhl, I., Överlåtelsebara utsläppsrätter. En studie av kolväteutsläpp i Göteborg. Research report

Björkegren, D., Mot en kognitiv organisationsteori. Research report

Claesson, K., Effektiviteten på Stockholms Fondbörs

Davidsson, P. Growth Willingness in Small Firms. Entrepreneurship - and after? Research report

Engshagen, I., Finansiella nyckeltal för koncern versus koncernbolag. En studie om finansiella måltal. Research report

Fredriksson, O., Holmlöv, PG., Julander, C-R., Distribution av varor och tjänster i informationssamhället.

Hagstedt, P., Sponsring - mer än marknadsföring. EFI/MTC

Järnhäll, B., On the Formulation and Estimation of Models of Open Economies. Research report

Kylén, B., Digitalkartans ekonomi. Samhällsekonomiska modeller för strategiska val. Delrapport 2 i "Nordisk Kvantif"

Lundgren, S., Elpriser: Principer och praktik. Research report

Schwarz, B., Hederstierna, A., Socialbidragens utveckling - orsaker eller samvariationer? Research report

Schwartz, E., Begreppet federativ organisation belyst i ett organisationsteoretiskt,
juridiskt och empiriskt perpektiv. Research report.

Westlund, A.H. Öhlén, S., Business Cycle Forecasting in Sweden; A problem analysis. Research report

1988

Andréasson, I-M., Costs of Controls on Farmers' Use of Nitrogen. A study applied to Gotland.

Björkegren, D., Från sluten till öppen kunskapsproduktion. Delrapport 3 i forskningsprojektet Lärande och tänkande i organisationer. Research report

Björkegren, D., Från tavistock till Human-relation. Delrapport 4 i forskningsprojektet Lärande och tänkande i organisationer. Research report

Björklund, L., Internationell projektförsäljning. En studie av ad hoc-samverkan mellan företag vid internationella projekt. Research report

Bojö, J., Mäler, K-G., Unemo, L., Economic Analysis of Environmental Consequences of Development Projects. Research report

Brynell, K., Davidsson, P., Det nya småföretagandet? - en empirisk jämförelse mellan små high-tech företag och konventionella småföretag. Research report

Dahlgren, G., Witt, P., Ledning av fusionsförlopp. En analys av bildandet av Ericsson Information Systems AB.

Forsell, A., Från traditionell till modern sparbank. Idé organisation och verksamhet i omvandling. Research report

Hirdman, V., The Stability and the Interest Sensitivity of Swedish Short Term Capital Flows. Research report

Hultén, S., Vad bestämmer de svenska exportmarknadsandelarnas utveckling?

Häckner, J., Biobränslenas konkurrenskraft i ett framtida perspektiv. Research report

Jennergren, P.,Näslund, B., If the Supreme Court had Known Option Theory: A Reconsideration of the Gimo Case. Research report

Jennergren, P., Näslund, B., The Gimo Corporation Revisited: an exercise in Finance Theory. Research report

Jennergren, P., Näslund, B., Valuation of Debt and Equity Through One- and Two-State Contingent Claims Models - with an Application to a Swedish Court Case. Research report

Jonung, L., Laidler, D., Are Perceptions of Inflation Rational? Some Evidence for Sweden. Research report.

Lagerstam, C., Business Intelligence. Teori samt empirisk studie av elektronik- och bioteknikbranschen i Japan. Research report

Liljegren, G., Interdependens och dynamik i långsiktiga kundrelationer. Industriell försäljning i nätverksperspektiv.

Näslund, B., Några synpunkter på börsfallet i oktober 1987. Research report

Olsson, C., The Cost-Effectiveness of Different Strategies Aimed at Reducing the Amount of Sulphur Deposition in Europe. Research report

Philips, Åke, Eldsjälar. En studie av aktörsskap i arbetsorganisatoriskt utvecklingsarbete.

Sellstedt, Bo, Produktionsstrategier. En diskussion med utgångspunkt från litteraturen. Research report.

Skogsvik, K., Prognos av finansiell kris med redovisningsmått. En jämförelse mellan traditionell och inflationsjusterad redovisning.

Wahlund, R., Skatteomläggningen 1983 - 1985. En studie av några konsekvenser. Research report

Wahlund, R., Varför och hur olika svenska hushåll sparar. Research report

Vredin, A., Macroeconomic Policies and the Balance of Payments.

Åkerman, J., Economic Valuation of Risk Reduction: The Case of In-Door Radon. Research report.

1989

Andersson, T., Foreign Direct Investment in Competing Host Countries. A Study of Taxation and Nationalization.

Björkegren, D., Skönhetens uppfinnnare. Research report

Carlsson, A., Estimates of the Costs of Emission control in the Swedish Energy Sector. Research report

Davidsson, P., Continued Entrepreneurship and Small Firm Growth.

DeJuan, A., Fiscal Attitudes and Behavior. A study of 16 - 35 years old Swedish citizens. Research report

Gadde, L-E., Håkansson, H., Öberg, M., Stability and Change in Automobile Distribution. Research report

Jonung, L., The Economics of Private Money. Private Bank Notes in Sweden 1831 - 1902. Research report

Kyle'n, B., Hur företagschefer beslutar innan de blir överraskade. Ett försök till förklaring av svarsmönster i svagsignalsitutationer.

Lagerstam, C., Jämförelse av skydd med samlad valuta/aktieoption och skydd med separata aktie- och valutaoptioner samt härledning av optionspriser. Research report

Larsson, B., Koncernföretaget. Ägarorganisationer eller organisation för ägare?

Löwstedt, J., Föreställningar, ny teknik och förändring. Tre organisationsprocesser ur ett kognitivt aktörsperspektiv. Doxa Förlag

Schuster, W., Ägandeformens betydelse för ett företag - en studie av ICA-rörelsen.

Schwartz, B., Företaget som medborgare. Samhällskontakter och reklam som legitimeringsinstrument. Research report

Spets, G., Vägen till nej. Anade och oanade konsekvenser av en OS-satsning. Research report.