

Möller, Joachim

**Working Paper**

## Time-varying coefficients in a simultaneous equation model: Theoretical concept and some simulation results

Diskussionsbeiträge - Serie I, No. 241

**Provided in Cooperation with:**

Department of Economics, University of Konstanz

*Suggested Citation:* Möller, Joachim (1988) : Time-varying coefficients in a simultaneous equation model: Theoretical concept and some simulation results, Diskussionsbeiträge - Serie I, No. 241, Universität Konstanz, Fakultät für Wirtschaftswissenschaften und Statistik, Konstanz

This Version is available at:

<https://hdl.handle.net/10419/75065>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

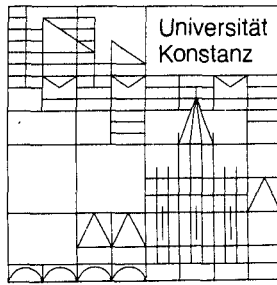
Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



**Fakultät für  
Wirtschaftswissenschaften  
und Statistik**

---

Joachim Möller

**Time-Varying Coefficients in a  
Simultaneous Equation Model**  
Theoretical Concept  
and Some Simulation Results

Diskussionsbeiträge

---

TIME-VARYING COEFFICIENTS  
IN A SIMULTANEOUS EQUATION MODEL

Theoretical Concept and Some Simulation Results

Joachim MÖLLER

Serie I – Nr.241

Ag 282/89 *JK* Weltwirtschaft  
1988

November 1988

### Abstract

The paper deals with the estimation of a simultaneous equation system with time-varying parameters. The system is formulated in state space form and it is shown that the Kalman filter algorithm is applicable. The prior information required by this estimation method can be obtained by maximization of the likelihood function via Fisher-Scoring. Applying and extending an approach due to Engle, Watson (1981) leads to the recursive calculation of the score and the information matrix for the model under consideration. Three simulation experiments are run to show that the proposed algorithm performs well with respect to the identification of time-varying parameters in a simultaneous equation model.

## 1. Introduction

The paper deals with the problems of estimating a simultaneous equation system with time-varying parameters. Compared to the estimation of single equations with time-varying coefficients some specific complications arise in this context which can be solved within a framework of recursive estimation.

The estimation problem has been considered by Chow (1984) who gives a state-space representation of the simultaneous equation model with time-varying coefficients. Furthermore, Chow outlines a Maximum Likelihood method for the problem of estimating the unknown prior statistics of the model (i.e. the covariance matrices of the state space model, the initial vector of coefficient and the corresponding covariance matrix). Information about the prior statistics will be required if the Kalman filter algorithm is to be used as a convenient tool for estimation of the time-path of the coefficient vector.

Our approach differs from that of Chow not only in the representation of the model but also in the Maximum-Likelihood procedure which is developed to estimate the unknown prior statistics. An instrumental variable approach is chosen to handle the problem of simultaneity within the context of the state space model. It can be shown that an extended version of the recursive algorithm to calculate the score and the information matrix, which was originally proposed by Engle, Watson (1981) and Watson, Engle (1983), can be applied to the problem. Then, the Fisher-Scoring iteration method can be used to calculate Maximum Likelihood estimates of the prior statistics. In our view, this procedure has numerical and computational advantages over the method proposed by Chow.

The paper is organized as follows: After the state-space representation of the simultaneous equation system with time-varying parameters in section 2, the corresponding (extended) Kalman filter algorithm is derived in section 3 and problems of estimating the time-path of the coefficient vector for given prior statistics are discussed in section 4. Section 5 outlines an iterative Maximum-Likelihood-approach for the estimation of the variance components of the state space model, based on recursive calculation of the first and the second derivatives of the likelihood function. Some possibilities for testing the hypothesis of time-varying coefficients versus the hypothesis of constant coefficients are described in section 6. In order to demonstrate the applicability of the proposed estimation method, section 7 presents the results of some simulation experiments.

## 2. State-Space Representation of a Simultaneous Equation System

The simultaneous equation system under consideration is:

$$\begin{matrix} \Gamma_t & y_t & = & B_t & x_t & + & u_t & , \\ (g \times g) & (g \times 1) & & (g \times k) & (k \times 1) & & (g \times 1) & \end{matrix} \quad (1)$$

where  $\Gamma_t$  and  $B_t$  are matrices of coefficients associated with  $g$  endogenous variables  $y_t$  and  $k$  predetermined variables  $x_t$ , respectively.  $\Gamma_t$  and  $B_t$  partly consist of time-varying

parameters and partly consist of zeros and ones which reflect structural information. The stochastic term  $u_t$  is independently normally distributed with zero mean and covariance matrix  $R$  which is assumed to be constant over time.

Let us define a matrix  $\Gamma_t^* = I_g - \Gamma_t$ , where  $I_g$  denotes a unit matrix of dimension  $g$ . Then, eq.(1) may be written as

$$\begin{aligned} y_t &= B_t x_t + \Gamma_t^* y_t + u_t \\ &= (B_t \Gamma_t^*) \begin{pmatrix} x_t \\ y_t \end{pmatrix} + u_t. \end{aligned} \quad (2)$$

Using the reduced form  $y_t = \Gamma_t^{-1} B_t x_t + \Gamma_t^{-1} u_t$  the vector of endogeneous variables may be separated into an 'explained part'  $y_t^* = \Gamma_t^{-1} B_t x_t$  and a stochastic part  $\Gamma_t^{-1} u_t$ . Of course the vector  $y_t^*$  is unknown and has to be replaced by suitably chosen instruments, a problem which we discuss later. Since  $\Gamma_t^* \Gamma_t^{-1} u_t = \Gamma_t^{-1} u_t - u_t$  it follows immediately from eq.(2) that

$$y_t = (B_t \Gamma_t^*) \begin{pmatrix} x_t \\ y_t^* \end{pmatrix} + \Gamma_t^{-1} u_t. \quad (3)$$

Let

$$\begin{aligned} \tilde{\beta}_t &\equiv \text{vec} B_t \\ \tilde{\gamma}_t &\equiv \text{vec} \Gamma_t^*, \end{aligned}$$

where  $\text{vec}$  stands for an operator which vectorizes the following matrix by columns. Note that  $B_t x_t = (x_t' \otimes I_k) \tilde{\beta}_t$  and  $\Gamma_t^* y_t^* = (y_t^{*'} \otimes I_g) \tilde{\gamma}_t$ .

$R_x$  and  $R_y$  are two matrices of suitable size containing information about linear restrictions with respect to  $B_t$  and  $\Gamma_t$  so that

$$\begin{aligned} \tilde{\beta}_t &= R_x \beta_t \\ \tilde{\gamma}_t &= R_y \gamma_t, \end{aligned}$$

where the components of the vectors  $\beta_t$  and  $\gamma_t$  are the unknown elements of the parameter matrices.

The data matrices of the system can be combined with the information about linear restrictions:

$$\begin{aligned}\tilde{X}_t &\equiv (x_t' \otimes I_g)R_x \\ \tilde{Y}_t^* &\equiv (y_t^{*'} \otimes I_g)R_y.\end{aligned}$$

Substituting in eq.(3) gives:

$$y_t = (\tilde{X}_t \tilde{Y}_t^*) \begin{pmatrix} \beta_t \\ \gamma_t \end{pmatrix} + \Gamma_t^{-1} u_t,$$

or with the corresponding definition of  $Z_t$  and  $\delta_t$ :

$$y_t = Z_t \delta_t + \Gamma_t^{-1} u_t. \quad (4)$$

Eq.(4) is in the form of a *measurement equation* of a state space model. To close the model a state equation is required which gives the 'law of motion' for the system parameters. In the absence of further information it may be assumed that the coefficient vector  $\delta$  follows a *random walk*:

$$\delta_t = \delta_{t-1} + v_t, \quad (5)$$

where the stochastic disturbances  $v_t$  are independently normally distributed with covariance matrix  $Q$ . Note, that the model covers the possibility of constant parameters as a special case (if  $Q$  is taken as a matrix of zeros).

In the next section it is shown that the state-space model defined by eq.(4) and eq.(5) may be estimated recursively by using the Kalman filter algorithm (Kalman 1960), provided that instruments for  $y_t^*$  are available and the covariance matrices  $Q$  and  $R$  are given.

### 3. Derivation of the Kalman algorithm

The basic principle of the Kalman filter is to calculate the vector of coefficients of a state space model with information up to time period  $t$  using estimates of this vector containing information up to period  $t - 1$ . So all estimates consist of a conditional expectation component reflecting all past information and an 'innovation' component which takes into account the new information of the data set in period  $t$ .

Let  $E_{t_1} a_{t_2} \equiv a_{t_2|t_1}$  be the expectation of a variable  $a$  for time period  $t_2$  based on information up to time period  $t_1$  with  $t_1 \leq t_2$ . The recursive approach postulates a linear relationship between the conditional parameter vector  $\delta_{t|t-1}$  and the unconditional parameter vector  $\delta_{t|t}$  of the following form:

$$\delta_{t|t} = \delta_{t|t-1} + K_t \eta_t, \quad (6)$$

where the  $(g \times 1)$  innovation vector  $\eta_t$  is defined as:

$$\eta_t \equiv y_t - y_{t|t-1}. \quad (7)$$

$K_t$  denotes a matrix of 'weight factors' by which the innovation process is transformed into changes of the parameters. Following Chow (1984),  $K_t$  may be easily determined using classical regression theory:

$$K_t = E[(\delta_{t|t} - \delta_{t|t-1}) \eta_t' (\eta_t \eta_t')^{-1}]. \quad (8)$$

Since  $y_{t|t-1} = Z_t \delta_{t|t-1}$ , the innovation vector can be expressed as

$$\eta_t = Z_t(\delta_t - \delta_{t|t-1}) + \Gamma_t^{-1} u_t, \quad (9)$$

so that the covariance matrix of the innovations  $H_t \equiv \text{var}(\eta_t)$  is given as follows:

$$H_t \equiv \text{var}(\eta_t) = Z_t P_{t|t-1} Z_t' + \Gamma_{t|t-1}^{-1} R \Gamma_{t|t-1}^{-1'}, \quad (10)$$

where

$$P_{t|t-1} \equiv \text{var}(\delta_t - \delta_{t|t-1}).$$

Since  $E(\Gamma_{t|t-1}^{-1} u_t) = 0$ , the Kalman filter 'gain matrix'  $K_t$  is obtained as

$$K_t = P_{t|t-1} Z_t' H_t^{-1}. \quad (11)$$

Taking account of eq.(6) it follows that the unconditional covariance matrix of the estimated coefficient vector for time period  $t$  obeys the following recursive relationship:

$$P_{t|t} = P_{t|t-1} - K_t H_t K_t' = P_{t|t-1} - K_t Z_t P_{t|t-1}. \quad (12)$$

The covariance matrix of the coefficient vector for time period  $t$  conditional to the information set of  $t - 1$  is given by



$$P_{t|t-1} = P_{t-1|t-1} + Q. \quad (13)$$

Since the matrix  $H_t$  is positive definite, eq.(12) implies that the variance of the estimated coefficient vector becomes smaller if new data are available from the information stream.

Provided that the  $y_t^*$  are properly instrumented and suitable prior information is available, the given formulas define a complete filter algorithm. Calculating instruments in a conventional way (for example by regressing the endogeneous variables of all predetermined variables of the system), however, would be unsatisfactory, especially if the assumption of constant parameters was imposed in this step. The logic of recursive estimation prevents a better solution to this problem. The simple idea is to replace  $y_t^*$  by its conditional expectation  $y_{t|t-1} = \Gamma_{t|t-1}^{-1} B_{t|t-1} x_t$ . Since the innovation vector has a zero expectation this means that a consistent estimate for  $y_t^*$  is used which contains all information of the system up to period  $t - 1$ . Furthermore, this proceeding has an attractive economic interpretation because the usage of the conditional expectation  $y_{t|t-1}$  for the endogeneous variables on the right hand side of the equations is in accord with rational expectation theory.

#### 4. Estimating the time-path of coefficients

Two questions arise in practical application: How to get the prior information and how to properly estimate the time-path of those model coefficients for which the corresponding diagonal elements of the  $Q$ -matrix are greater than zero. In this section the focus is laid on the second question whereas the first will be discussed in the next.

If the econometrician is equipped with covariance information  $\hat{Q}$  and  $\hat{R}$ , he can choose an arbitrary initial coefficient vector  $\delta_0$  (typically set equal to zero) and a corresponding covariance matrix  $P_0$  with large diagonal elements to express the uncertainty of the initial coefficient estimate. Then the Kalman filter algorithm gives a series of unbiased and efficient estimates of  $\delta_t$  conditional on the information set which is available up to time period  $t$ . Note that such a filter applied to the case of  $Q$  being a zero matrix gives estimates of the coefficient vector which are by no means constant over time. Although the outcome of the filter process (which is the analogon of recursive least squares estimation) may be of some diagnostic value, the econometrician is generally interested in estimates of the coefficients which contain the whole sample information set. For this purpose, the *Kalman smoothing algorithm* has to be used. In contrast to the filter algorithm, the Kalman smoothing algorithm with  $Q$  being a zero matrix yields estimates for the coefficient vector and the corresponding covariance matrix which do not vary over time except for numerical inaccuracies.

For calculating the Kalman smoother, the results of the (forward) filtering process,  $\delta_{t|t}$  and  $P_{t|t}$ , have to be stored. The smoothing algorithm, then operates 'backward', i.e. from time period  $(T - 1)$  to 1 where the smoothing formulas [see, for instance, Anderson, Moore (1979)] are:

$$\delta_{t|T} = \delta_{t|t} + W_t(\delta_{t+1|T} - \delta_{t+1|t}) \quad (14)$$

$$P_{t|T} = P_{t|t} + W_t(P_{t+1|T} - P_{t+1|t}^*)W_t' \quad (15)$$

with

$$W_t = P_{t|t-1}P_{t+1}^{*-1}.$$

The starting values of the smoothing algorithm  $\delta_{T|T}$  and  $P_{T|T}$  are given by the last estimates of the forward filter, because the filter results for the last time period contain all the information of the sample and are therefore equivalent to the corresponding smoothing results.

## 5. Maximum Likelihood Estimation of the Prior Statistics

An important question is what the 'sufficient statistics', in the sense of Rao (1973), are in the model under consideration. As mentioned above, the filter algorithm requires knowledge of the matrices  $Q$  and  $R$  as well as of the initial coefficient vector  $\delta_0$  and the corresponding covariance matrix  $P_0$ . As has been shown by Sarris (1973) for the single equation case an estimate for  $\delta_0$  and  $P_0$  can be obtained by GLS for given covariance information  $R$  and  $Q$ . The results of Duncan, Horn (1972) and Sant (1977) imply that equivalent estimates are obtained from a backward filter for the sequence of time periods  $(T, T-1, \dots, 0)$  starting with an arbitrary coefficient vector  $\delta_{T+1}$  and a corresponding covariance matrix  $P_{T+1}$  with large diagonal elements. The problem of prior statistics thus reduces to the determination of  $Q$  and  $R$ , so that these matrices can be interpreted as the 'sufficient statistics'.

In the econometric context,  $Q$  and  $R$  are unknown and have to be estimated. In the following, a Maximum Likelihood method will be outlined.

In analogy to a proof given by Kirchen (1988) for the case of single equation estimation, it is shown in the appendix, that the likelihood function for a simultaneous equation system with time-varying parameters is equivalent to the likelihood function of a state space model [as already formulated by Schweppe (1965)]:

$$\ln L_t = -\frac{g}{2} \ln(2\pi) - \frac{1}{2} \ln |H_t| - \frac{1}{2} \text{spur}(H_t^{-1} \eta_t \eta_t'). \quad (16)$$

The well-known formulation of the log likelihood function of a simultaneous equation system with *constant* coefficients [see, for instance, Schmidt (1976)] is:

$$L = -\frac{Tg}{2} \ln(2\pi) + \frac{T}{2} (\ln |\Gamma|^2 - \ln |R|) - \frac{1}{2} \sum_1^T \text{spur}(R^{-1} u_t u_t') \quad (17)$$

or

$$L = \sum_1^T L_t$$

with

$$L_t = -\frac{g}{2} \ln(2\pi) + \frac{1}{2} (\ln |\Gamma|^2 - \ln |R|) - \frac{1}{2} \text{spur}(R^{-1} u_t u_t'). \quad (18)$$

It can easily be shown that eq.(17) includes eq.(18) as a special case since  $H_t = \Gamma^{-1} R \Gamma^{-1}$  and  $\eta_t = \Gamma^{-1} u_t$  if the elements of  $\Gamma$  are constant over time. But, evidently, adding a time index to  $\Gamma$  in eq.(17) will not lead to a correct formulation of the likelihood function of a time-varying parameter model.

Since the direct maximization of the likelihood function eq.(16) is not possible, an appropriate iteration procedure has to be used. For example, the Scoring method [see, for instance, Rao (1973)] is applicable in principle but involves extremely large computations in the case of simultaneous equation estimation. These computational problems can be reduced to a manageable size if a *filter based* Scoring algorithm is used. As discussed in Watson and Engle (1983) a corresponding *EM*-algorithm [see Dempster, Laird, Rubin (1977)] may be seen as an alternative. We tried both methods but finally preferred the former not only because the speed of convergence of the *EM*-algorithm was found to be very low but also because the Scoring method gives more information and allows for the possibility that parts of the covariance matrices to be estimated become zero during the iteration process.

In the following, it is shown that the approach of recursive calculation of the score and the information matrix, developed by Engle, Watson (1981) is applicable to our problem.

Let us first define a vector  $\theta$  which contains the unknown elements of the covariance matrices  $R$  and  $Q$  (paying regard to symmetry). Taking eq.(17) as a starting point the derivatives of the likelihood function with respect to the  $i$ -th element of  $\theta$  can be calculated in a straightforward manner. Since

$$\frac{\partial H_t^{-1}}{\partial \theta_i} = -H_t^{-1} H_{\theta_i, t} H_t^{-1} \quad (19)$$

and

$$\frac{\partial \ln |H_t|}{\partial \theta_i} = \text{spur}(H_t^{-1} H_{\theta_i, t}). \quad (20)$$

the first derivative of the likelihood function for time period  $t$  is

$$\frac{\partial \ln L_t}{\partial \theta_i} = -\frac{1}{2} \text{spur} (H_t^{-1} H_{\theta_i,t}) - \eta'_{\theta_i,t} H_t^{-1} \eta_t + \frac{1}{2} \eta_t H_t^{-1} H_{\theta_i,t} H_t^{-1} \eta_t. \quad (21)$$

Therefore, the  $i$ -th element of the score  $s$  is

$$s_i = \sum_{t=1}^T \frac{\partial \ln L_t}{\partial \theta_i}. \quad (22)$$

Defining  $H_t^* \equiv H_t^{-1} H_{\theta_i,t}$  and denoting the derivative of of a matrix or vector  $A_t$  with respect to  $\theta_i$  as  $A_{\theta_i,t}$ , the second derivative of the likelihood function may be formulated

$$\begin{aligned} \frac{\partial^2 \ln L_t}{\partial \theta_i \partial \theta_j} = & -\frac{1}{2} \text{spur} [H_{\theta_j,t}^* (I_g - H_t^{-1} \eta_t \eta_t')] \\ & -\frac{1}{2} \text{spur} [H_t^* H_t^{-1} H_{\theta_j,t} H_t^{-1} \eta_t \eta_t'] \\ & +\frac{1}{2} \text{spur} [H_t^* H_t^{-1} (\eta_{\theta_j,t} \eta_t' + \eta_t \eta_{\theta_j,t}')] \\ & -\eta'_{\theta_i,\theta_j,t} H_t^{-1} \eta_t \\ & +\eta'_{\theta_i,t} H_t^{-1} H_{\theta_j,t} H_t^{-1} \eta_t \\ & -\eta'_{\theta_i,t} H_t^{-1} \eta_{\theta_j,t}. \end{aligned} \quad (23)$$

Taking the expectation of eq.(23) we notice that

$$E(\eta_t \eta_t') = H_t$$

and

$$E(\eta_{\theta_i,t} \eta_t') = E(\eta_t \eta_{\theta_j,t}') = 0.$$

The  $(i,j)$ -th element of the information matrix  $\mathfrak{F}$ , therefore, is

$$\mathfrak{F}_{i,j} = -E\left(\frac{\partial^2 \ln L_t}{\partial \theta_i \partial \theta_j}\right) = \sum_{t=1}^T \left[ \frac{1}{2} \text{spur} (H_t^{-1} H_{\theta_i,t} H_t^{-1} H_{\theta_j,t}) + \eta'_{\theta_i,t} H_t^{-1} \eta_{\theta_j,t} \right]. \quad (24)$$

In eq.(21) and eq.(24) the derivatives of the covariance matrix of the endogenous variables and of the innovations are required. It can be shown that these matrix derivatives  $H_{\theta_i,t}$  and  $\eta_{\theta_i,t}$  may be calculated recursively.

Using the definitions

$$V_t \equiv \Gamma_t^{*-1} R \Gamma_t^{*-1'}$$

and

$$P_{\theta_i, t}^* \equiv \frac{\partial P_{t|t-1}}{\partial \theta}$$

the complete recursive formulas for the extended Kalman filter approach to calculate the score and the information matrix are:

$$P_{\theta_i, t}^* = P_{\theta_i, t-1} + Q_{\theta_i} \quad (25)$$

$$H_{\theta_i, t} = Z_t P_{\theta_i, t}^* Z_t' + V_{\theta_i, t} \quad (26)$$

$$V_{\theta_i, t} = \Gamma_t^{*-1} R_{\theta_i} \Gamma_t^{*-1'} + \Gamma_t^{*-1} \Gamma_{\theta_i, t}^* V_t + V_t' \Gamma_{\theta_i, t}^{*'} \Gamma_t^{*-1'} \quad (27)$$

$$K_{\theta_i, t} = (P_{\theta_i, t}^* Z_t' - K_t H_{\theta_i, t}) H_t^{-1} \quad (28)$$

$$P_{\theta_i, t} = P_{\theta_i, t}^* - K_{\theta_i, t} Z_t P_t^* - K_t Z_t P_{\theta_i, t}^* \quad (29)$$

$$\eta_{\theta_i, t} = -Z_t \delta_{\theta_i, t} \quad (30)$$

$$\delta_{\theta_i, t} = \delta_{\theta_i, t-1} + K_{\theta_i, t} \eta_t + K_t \eta_{\theta_i, t} \quad (31)$$

After the  $n$ -th iteration the update of vector  $\theta$  is calculated as:

$$\theta^{n+1} = \theta^n + \lambda^n \mathfrak{S}^{n-1} s^n, \quad (32)$$

where  $\lambda^n$  is a suitable chosen constant. During the maximization procedure, the positive definiteness of the covariance matrices has to be taken into account as a restriction.

It should be mentioned that computation of the information matrix is computationally expensive. Berndt *et.al.* (1974) have shown that a consistent estimate of the information matrix can be obtained from the scores. If the row vector of first derivatives of the likelihood function with respect to theta in time period  $t$  is denoted by  $s_t$  and a matrix  $S$  is formed which contains  $s_t$  as its  $t$ -th row, then an estimate for the information matrix is given by  $S'S$ .

## 6. Tests

Since the method to determine the prior statistics of our model is Maximum Likelihood the standard Lagrange multiplier, Wald and likelihood ratio tests [see Engle (1984)] can be applied in a natural way.

The Lagrange-multiplier expression

$$\xi^n = s^{n'} \mathfrak{S}^{n-1} s^n, \quad (33)$$

is a test whether the first derivative of the likelihood function is equal to zero and therefore may be used as a stopping criterium in the iteration process.

An important problem is to test the hypothesis of time-varying coefficients against the null of constant coefficients. For this purpose let us define  $\theta^0$  as the vector of estimated elements of the covariance matrices with all elements of  $Q$  set equal to zero a priori and  $\hat{\theta}$  as the vector of elements of  $Q$  and  $R$  which are the result of unconstrained maximization. Then, the likelihood ratio test

$$-2 [\ln L(\theta^0, y) - \ln L(\hat{\theta}, y)] \quad (34)$$

has a limiting  $\chi^2$  distribution with  $m$  degrees of freedom where  $m$  is the number of unknown elements of the  $Q$  matrix.

The Wald test which, following Engle (1984), can be interpreted as an asymptotic approximation of the familiar  $t$  and  $F$ -test in econometrics, is easily calculated as

$$(\hat{\theta} - \theta^0)' \mathfrak{S}(\hat{\theta}) (\hat{\theta} - \theta^0). \quad (35)$$

This test assumes that a central limit theorem applies to  $\hat{\theta}$  and that the information matrix is consistently estimated. The test statistic again has a limiting  $\chi^2$  distribution with  $m$  degrees of freedom.

The Scoring method also gives an estimate of the variance and standard error of the elements of  $\hat{\theta}$ , so

$$\text{var}(\hat{\theta}) = [\mathfrak{S}(\hat{\theta})]^{-1}. \quad (36)$$

Another important point is to test the assumptions underlying the state space model. Schneider (1987), for example, stresses the fact that the series of innovations should be white

noise. Several test statistics could be used to test the normality of the innovation process. For example, it could be tested whether the sequence of autocovariances is significantly different from zero or not. Non-parametric tests could be used which focus on the sign changes of the innovations. Also the Bera-Jarque test on normality which is based on the third and fourth moments of the empirical distribution is applicable. Finally, the hypothesis of constant variance may be tested, for example, by using ARCH-type tests.

## 7. Some simulation experiments

It is an important practical question whether the theoretically outlined procedure for estimating the prior statistics is applicable to data sets econometricians are typically concerned with. To shed light on the small sample properties of the proposed algorithm, some simulation experiments were carried through.

For the sake of simplicity a system of two simultaneous equations with six unknown parameters and  $T = 70$  observations was considered. The system is exactly identified. Real economic time series (quarterly growth rate of wage income and the unemployment rate) were used as exogeneous variables. The simulated model is:

$$y_{1t} = \delta_{1t} + \delta_{2t}x_{1t} + \delta_{5t}y_{2t} + u_{1t} \quad (37)$$

$$y_{2t} = \delta_{3t} + \delta_{4t}x_{2t} + \delta_{6t}y_{1t} + u_{2t}, \quad (38)$$

where

$x_{1t}$  : growth rate of wage income

$x_{2t}$  : log of the unemployment rate

Hence the coefficient matrices are:

$$B_t = \begin{pmatrix} \delta_{1t} & \delta_{2t} & 0 \\ \delta_{3t} & 0 & \delta_{4t} \end{pmatrix} \quad \text{and} \quad \Gamma_t^* = \begin{pmatrix} 0 & \delta_{5t} \\ \delta_{6t} & 0 \end{pmatrix}$$

The vector of endogeneous variables at time period  $t$  is  $y_t = (y_{1t}, y_{2t})'$  and the data matrix  $Z_t$  of the state space model is

$$Z_t = \begin{pmatrix} 1 & x_{1t} & 0 & 0 & y_{1t}^* & 0 \\ 0 & 0 & 1 & x_{2t} & 0 & y_{2t}^* \end{pmatrix}$$

The covariance matrices of the measurement and the state equation are assumed to have diagonal structure:  $\text{var}(u_t) = R = \text{diag}(r_{11} \ r_{22})$  and  $\text{var}(v_t) = Q = \text{diag}(q_{11} \ q_{22} \ q_{33} \ q_{44} \ q_{55} \ q_{66})$ .

In the simulation experiments emphasis is laid on the following questions: (i) Does the algorithm outlined above provide consistent estimates of the covariance structure of the model? (ii) In how many cases a time-varying coefficient will be found while the true coefficient is not time-varying (type I error) ? (iii) In how many cases will a time-varying coefficient not be detected (type II error) ? (iv) Are the estimates of the standard error of the variance parameters obtained from the information matrix reliable?

*Experiment 1 : Simulation results for the constant parameter case*

In a first experiment the case where the parameters of the true model do not vary over time is considered. This allows us to examine the extent to which the null hypothesis 'constant parameters over time' is rejected although it is true.

---

insert table 1 about here

---

The true parameters of the model and the simulation results are given at the top of *table 1*. As starting values for all elements of the diagonal matrices  $R$  and  $Q$  the arbitrary value of 0.1 was chosen. The coefficient vector  $\delta$  for time period 0 was estimated via backward filtering with the coefficient vector  $\delta$  for time period  $T + 1$  set to zero and the corresponding covariance matrix  $P_{T+1}$  set to a diagonal matrix with large elements ( $10^6$ ).<sup>1</sup> The number of replications was limited to 100 for each simulation experiment.

The results of the simulation experiment as shown in *table 1* are rather satisfactory. A type II error does not occur with respect to the variances of the measurement equation. The bias of these estimates is negative for the element  $R_{11}$  and positive for  $R_{22}$  with an order of magnitude between 3.5 and 4%. These results may be compared to the variance estimates of 2-Stage-Least-Squares (2SLS) estimates applied to the same data. The results of the 2SLS-method imply a considerable (positive) bias for the estimated variances for the measurement equation, indicating that this method is less efficient.

The estimation results for the parameter vector  $\delta$  do not differ much between the two estimation methods and come close to the true parameters. Also the standard deviations of the empirical distribution of the parameter estimates are roughly comparable.

---

<sup>1</sup> It should be mentioned that there is no guarantee that the algorithm will converge in a given number of steps. In order to keep the simulation experiment tractable it was decided to skip the cases where the algorithm failed to converge in 12 iteration steps. In some cases the algorithm did not converge because of the accumulation of numerical inaccuracies. These cases were also excluded from the considerations.



The mean standard error as estimated from the information matrix is 0.034 for  $\hat{R}_{11}$  and 0.091 for  $\hat{R}_{22}$ . Taking the root mean square error of the estimated variance elements we get a somewhat smaller estimate of the variation of  $\hat{R}_{11}$  (0.015) and a somewhat larger estimate for  $\hat{R}_{22}$  (0.138). Hence, there is no indication that the square roots of the inverted information matrix do systematically deviate from the standard error of the model variances.

Although the estimated elements of the covariance matrix of the state equation are greater than zero in several cases, only in one case do these estimates exceed twice the root of the diagonal element of the inverted information matrix. If the usual t-test at the 5% level is used, two incorrect rejections of the null hypothesis are found in 100 replications. Neglecting the fact, that the number of replications is not sufficient to give precise information, these results indicate that applying a t-test might be too conservative a procedure in small samples.

*Experiment 2 : Identifying a single time-varying parameter*

In a second simulation experiment we consider the same model as in experiment 1 but with smaller variances in the measurement equation and with one element of the parameter vector ( $\delta_5$ ) varying over time [ $Q_{55} = 5.0 \times 10^{-3}$ ].

As reported in *table 2*, no type II errors with respect to the elements of the  $Q$ -matrix occur, i.e. in all 100 simulations the hypothesis  $Q_{55} = 0$  is rejected. This implies that the time-varying coefficient is detected in all cases<sup>2</sup>. Furthermore, a type I error is not found either, i.e. in no case a coefficient is incorrectly considered as time-varying if a t-test at the 5% level is applied. (The last finding confirms our tentative conclusion from above that the conventional t-test is too conservative).

---

insert table 2 about here

---

The simulation results as shown in *table 2* indicate that the variance of the measurement equations are slightly overestimated, whereas  $\hat{Q}_{55}$  appears somewhat smaller in mean ( $4.224 \times 10^{-3}$ ) than the true value ( $5.000 \times 10^{-3}$ ). The estimates of the coefficient vector  $\delta$  in time period 0 are very close to the true values for equation 1, whereas there is a certain deviation (especially for the coefficient  $\delta_6$ ) in the second equation which has a larger variance. As can be seen from *table 2* the standard deviations of  $Q$  and  $R$  as calculated from estimates of the information matrix are rather close to the standard deviations of the empirical distribution.

---

<sup>2</sup> This is not surprising, since the t-statistic for  $\hat{Q}_{55}$  in mean exceeds the 1 percent level.

The results of 2SLS estimates are the following: The variance of the first equation  $R_{11}$  (in which the time-varying coefficient occurs) is strongly overestimated by 2SLS (with a large standard deviation between different simulations). The estimated variance of equation 2 also has a positive bias. Although the estimates of the coefficient vector in mean do not deviate systematically from the true values and are roughly comparable to those of the Kalman method, the estimates of the parameter vector in equation 1 exhibit a large variation between different simulations as compared to the results of the Kalman method.

### *Experiment 3 : Identifying multiple time-varying parameters*

In the third experiment a model is simulated with one element of the  $B$ -matrix ( $\delta_4$ ) and one element of the  $\Gamma$ -matrix ( $\delta_5$ ) varying over time. The variances of the measurement equation are chosen as in experiment 2. Again each iteration procedure is started with arbitrary values and no a priori information is imposed.

---

insert table 3 about here

---

As reported in *table 3*, in this experiment only one error of type I occurs, whereas type II errors are more frequent: in four cases the variance  $R_{11}$  is not significant from zero when a t-test at the 5% level is applied, in seven resp. three cases the same occurs with respect to  $Q_{44}$  and  $Q_{55}$ , whereas the variance element  $R_{22}$  is always found to be significantly different from zero.

The variance elements of the measurement equations  $R_{11}$  and  $R_{22}$  are slightly underestimated in the first case and somewhat overestimated in the second. Likewise, there is no systematic deviation for the elements of the  $Q$ -matrix. As should be expected, the variance elements of the measurement equation are strongly overestimated by the 2SLS method. Again, the estimates of the parameter vector in mean are rather close to the true values for both methods, but the 2SLS results are much more volatile.

In this simulation experiment all estimates of the standard deviations of the elements of the  $Q$  and  $R$  matrices are underestimated when calculated from the information matrix.

## **8. Conclusion**

It was shown that a simultaneous equation model with time-varying parameters can be formulated in state space form. If the systematic part of the endogenous variables on the right hand side of the equations is replaced by its conditional expectation, the extended Kalman-filter algorithm may be used for recursive estimation. It can be demonstrated that

an (approximate) Maximum-Likelihood procedure to estimate the unknown prior statistics based on the Fisher-scoring method is applicable to the state space model.

In some simulation experiments a sample size was chosen which can be considered as realistic for econometric models with quarterly time-series data. The results indicate that the proposed procedure is very successful in identifying time-varying parameters of a small simultaneous model. If a conventional t-test is applied to the variance elements (using the square roots of the inverted information matrix as estimates for the standard error) errors of type I (rejecting the null hypothesis of constant coefficients although it is true) occur in very few cases. Although the number of replications is too small to give precise information there is some evidence that the t-test is too conservative.

## Appendix

Let  $f(Y_t)$  be the joint density function  $f(Y_t) = f(y_1, \dots, y_T)$  and  $L(Y_T)$  be the log likelihood function. The system state in period 0 is taken as known  $\delta_0 = \bar{\delta}_0$ , so that the corresponding covariance matrix  $P_0$  is a zero matrix. Then

$$L(y_1, \dots, y_T) = -\frac{gT}{2} \ln(2\pi) - \frac{1}{2} \sum_1^T (\ln |H_t| + \eta_t' H_t^{-1} \eta_t), \quad (\text{A-1})$$

where  $\eta_t = y_t - y_{t|t-1}$  and  $H_t = Z_t P_{t|t-1} Z_t' + \Gamma_{t|t-1}^{-1} R \Gamma_{t|t-1}^{-1}$ .

**Proof:**

From the definition of the conditional density follows for  $t \geq 2$ :

$$f(Y_t) = f(Y_{t-1})f(y_t|Y_{t-1}).$$

Hence a recursive formula for the log likelihood function is

$$L(Y_t) = L(Y_{t-1}) + \ln[f(y_t|Y_{t-1})].$$

The expression  $f(y_t|Y_{t-1}) = f(y_t|y_{t-1})$  is a normally distributed multivariate density with mean vector  $y_{t-1}$  and covariance matrix

$$\begin{aligned} \text{var } y_{t|t-1} &= E_{t|t-1}[(y_t - y_{t|t-1})(y_t - y_{t|t-1})'] \\ &= Z_t E_{t|t-1}[(\delta_t - \delta_{t|t-1})(\delta_t - \delta_{t|t-1})'] Z_t' + E_{t|t-1}(\Gamma_{t|t-1}^{-1} u_t u_t' \Gamma_{t|t-1}^{-1}) \\ &= Z_t P_{t|t-1} Z_t' + \Gamma_{t|t-1}^{-1} R \Gamma_{t|t-1}^{-1}. \end{aligned} \quad (\text{A-2})$$

Thus with  $H_t = \text{var}(y_{t|t-1})$  it follows for the density

$$\begin{aligned} \ln f(y_{t|t-1}) &= -g/2 \ln(2\pi) - 0.5 \ln |H_t| - 0.5 (y_t - y_{t|t-1}) H_t^{-1} (y_t - y_{t|t-1})' \\ &= -g/2 \ln(2\pi) - 0.5 (\ln |H_t| + \eta_t H_t^{-1} \eta_t'). \end{aligned}$$

Under our assumptions the dependent variable in period 0  $y_0$  is considered as non stochastic and the log likelihood for period 1 is

$$L(y_1) = -g/2 \ln(2\pi) - 0.5 \ln |\text{var}(y_1)| - 0.5 (y_1 - y_{1|0}) \text{var}(y_{1|0})^{-1} (y_1 - y_{1|0}). \quad (\text{A-3})$$

Since  $\text{var}(y_{1|0}) = H_1$ , eq. A-1 follows directly from eq. A-2 and A-3.

## REFERENCES

- Anderson, B.D.O. and J.B. Moore, 1979, Optimal Filtering (Prentice-Hall, Inc., Englewood Cliffs, N.J.).
- Berndt, E.K., B.H. Hall, R.E. Hall and J.A. Hausman, 1974, Estimation and inference in nonlinear structural models, *Annals of Economic and Social Measurement*, **3**, 653-666.
- Chow, G.C., 1984, Random and changing coefficient models, in: Griliches, Z. and M.D. Intriligator (eds.), *Handbook of Econometrics*, Vol. II (North Holland, Amsterdam), 1213-1245.
- Dempster, A.P., N.M. Laird and D.B. Rubin, 1977, Maximum Likelihood from incomplete data via the EM-algorithm, *Journal of the Royal Statistical Society*, **39**, 1-38.
- Duncan, D.B., and S.D. Horn, 1972, Linear dynamic recursive estimation from the viewpoint of regression analysis, *Journal of the American Statistical Association*, **67**, 815-821.
- Engle, R.F., 1984, Wald, Likelihood, and Lagrange Multiplier Tests in Econometrics, in: Griliches, Z. and M.D. Intriligator (eds.), *Handbook of Econometrics*, Vol. II (North Holland, Amsterdam), 775-826.
- Engle, R.F. and M.W. Watson, 1981, A one-factor multivariate time series model of metropolitan wage rates, *Journal of the American Statistical Association*, **76**, 774-781.
- Kalman, R.E., 1960, A new approach to linear filtering and prediction problems, *Transactions of ASME (American Society of Mechanical Engineers)*, Series D: *Journal of Basic Engineering*, **82**, 35-45.
- Kirchen, A., 1988, Schätzung zeitveränderlicher Strukturparameter in ökonometrischen Prognosemodellen (athenäum, Frankfurt am Main).
- Möller, J. and B. Wais, 1987, Kalman-Verfahren in der Ökonometrie - Schätzungen einer Geldangebotsfunktion mit zeitvariablen Koeffizienten unter Verwendung optimaler Filtereingangsinformation, *Allgemeines Statistisches Archiv*, **71**, 267-283.
- Rao, C.R., 1973, *Linear statistical inference and its applications*, 2nd edition (John Wiley and Sons, Inc., New York).
- Sant, D., 1977, Generalized Least Squares applied to time-varying parameter models, *Annals of Economic and Social Measurement*, **6**, 301-314.

- Sarris, A.H., 1973, A Bayesian approach to estimation of time-varying regression coefficients, *Annals of Economic and Social Measurement*, **2**, 501-524.
- Schmidt, P. 1976, *Econometrics* (Marcel Dekker, Inc., New York).
- Schneider, W., 1986, *Der Kalman-Filter als Instrument zur Diagnose und Schätzung variabler Parameter in ökonomischen Modellen* (Physica-Verlag, Heidelberg).
- Schweppe, F.C., 1965, Evaluation of likelihood functions for Gaussian signals, *IEEE Transactions on Information Theory*, **11**, 61-70.
- Watson, M.W. and R.F.Engle, 1983, Alternative algorithms for the estimation of dynamic factor, MIMIC and varying coefficient regression models, *Journal of Econometrics*, **23**, 385-400.

TABLE 1  
Simulation Results (Experiment 1)

	coefficients						variances		
	equation 1			equation 2			eq.1	eq.2	
	$\delta_1$	$\delta_2$	$\delta_5$	$\delta_3$	$\delta_4$	$\delta_6$	$R_{11}$	$R_{22}$	
	<i>true model</i>								
parameter	-1.000	0.500	0.600	1.000	0.500	0.400	0.200	0.500	
$\hat{Q}_{ii}$	0.0	0.0	0.0	0.0	0.0	0.0	-	-	
	<i>simulation results (2SLS-method)</i>								
parameter	-0.991	0.497	0.598	1.015	0.519	0.354	0.613	0.867	
s.d.	0.099	0.069	0.018	0.084	0.074	0.191	0.076	0.056	
	<i>simulation results (Kalman-method)</i>								
parameter	-0.991	0.501	0.596	1.011	0.516	0.359	0.194	0.537	
s.d.	0.107	0.114	0.029	0.114	0.073	0.179	0.015	0.138	
$\hat{Q}_{ii}$	0.095	0.229	0.022	0.301	0.000	0.007	-	-	
s.d.(1)	0.465	1.546	0.116	1.172	-	0.041	-	-	
s.d.(2)	-	-	-	-	-	-	0.034	0.091	
err. I	0	1	0	0	0	0	-	-	
err. II	-	-	-	-	-	-	0	0	

Notes:

- results are based on 100 replications;
- $\hat{Q}_{ii}$  gives the mean diagonal element of the covariance matrix of the state equation times  $10^3$ ;
- s.d. gives the standard deviation of the empirical distribution;
- s.d.1 gives the standard deviation of the empirical distribution of the estimated variance element  $\hat{Q}_{ii}$ ;
- s.d.2 gives the mean square root of the corresponding diagonal element of the inverted information matrix;
- err. I: gives the number of cases where the estimated variance element  $\hat{Q}_{ii}$  divided by the square root of the diagonal element of the inverted information matrix exceeds the critical value of the t-distribution on the 5% level although the true element  $Q_{ii}$  is equal to zero;
- err. II: gives the number of cases where the estimated variance elements  $R_{ii}$  or  $\hat{Q}_{ii}$  divided by the root of the diagonal element of the inverted information matrix do not exceed the critical value of the t-distribution on the 5% level although the true element  $Q_{ii}$  is greater than zero;
- the mean square roots of the diagonal elements of the inverted information matrix are not reported for the  $\hat{Q}_{ii}$  because the  $Q_{ii}$  are estimated as zero in several cases.

TABLE 2  
Simulation Results (Experiment 2)

	coefficients						variances	
	equation 1			equation 2			eq.1	eq.2
	$\delta_1$	$\delta_2$	$\delta_5$	$\delta_3$	$\delta_4$	$\delta_6$	$R_{11}$	$R_{22}$
	<i>true model</i>							
parameter	-1.000	0.500	0.600	1.000	0.500	0.400	0.020	0.050
$\hat{Q}_{ii}$	0	0	5.000	0	0	0	-	-
	<i>simulation results (2SLS-method)</i>							
parameter	-1.012	0.525	0.625	1.003	0.506	0.375	1.202	0.273
s.d.	0.304	0.231	0.161	0.045	0.068	0.239	1.371	0.236
	<i>simulation results (Kalman-method)</i>							
parameter	-1.011	0.497	0.596	1.003	0.523	0.335	0.024	0.069
s.d.	0.060	0.052	0.081	0.069	0.026	0.065	0.009	0.011
$\hat{Q}_{ii}$	0.029	0.010	4.224	0.129	0.005	0.092	-	-
s.d.(1)	0.119	0.027	1.357	0.540	0.016	0.360	-	-
s.d.(2)	-	-	1.259	-	-	-	0.007	0.012
err. I	0	0	-	0	0	0	-	-
err. II	-	-	0	-	-	-	1	0

Notes see table 1



TABLE 3  
Simulation Results (Experiment 3)

	coefficients						variances	
	equation 1			equation 2			eq.1	eq.2
	$\delta_1$	$\delta_2$	$\delta_5$	$\delta_3$	$\delta_4$	$\delta_6$	$R_{11}$	$R_{22}$
	<i>true model</i>							
parameter	-1.000	0.500	0.600	1.000	0.500	0.400	0.020	0.050
$\hat{Q}_{ii}$	0	0	2.500	0	2.500	0	-	-
	<i>simulation results (2SLS-method)</i>							
parameter	-0.956	0.482	0.602	1.020	0.501	0.401	0.857	0.853
s.d.	0.258	0.198	0.117	0.271	0.247	0.516	0.960	0.617
	<i>simulation results (Kalman-method)</i>							
parameter	-1.024	0.513	0.605	1.007	0.528	0.336	0.018	0.032
s.d.	0.072	0.052	0.071	0.063	0.068	0.086	0.009	0.017
$\hat{Q}_{ii}$	0.073	0.021	3.298	0.047	1.769	0.140	-	-
s.d.(1)	0.282	0.076	1.802	0.169	0.748	0.671	-	-
s.d.(2)	-	-	1.131	-	0.628	-	0.006	0.008
err. I	0	0	-	0	-	1	-	-
err. II	-	-	3	-	7	-	4	0

Notes see table 1