Schlicht, Ekkehart; Ludsteck, Johannes

Working Paper

Variance Estimation in a Random Coefficients Model

Munich Discussion Paper, No. 2006-12

Provided in Cooperation with:
University of Munich, Department of Economics

Suggested Citation: Schlicht, Ekkehart; Ludsteck, Johannes (2006) : Variance Estimation in a Random Coefficients Model, Munich Discussion Paper, No. 2006-12, Ludwig-Maximilians-Universität München, Volkswirtschaftliche Fakultät, München, https://doi.org/10.5282/ubm/epub.904

This Version is available at:
https://hdl.handle.net/10419/104208

Ekkehart Schlicht und Johannes Ludsteck:

Variance Estimation in a Random Coefficients Model

Munich Discussion Paper No. 2006-12

Department of Economics
University of Munich

Volkswirtschaftliche Fakultät
Ludwig-Maximilians-Universität München

Online at http://epub.ub.uni-muenchen.de/904/

Variance Estimation in a Random Coefficients Model

by

EKKEHART SCHLICHT and JOHANNES LUDSTECK[1]

Abstract

This papers describes an estimator for a standard state-space model with coefficients generated by a random walk that is statistically superior to the Kalman filter as applied to this particular class of models. Two closely related estimators for the variances are introduced: A maximum likelihood estimator and a moments estimator that builds on the idea that some moments are equalized to their expectations. These estimators perform quite similar in many cases. In some cases, however, the moments estimator is preferable both to the proposed likelihood estimator and the Kalman filter, as implemented in the program package Eviews.

[1] Ekkehart Schlicht (schlicht@lmu.de) is at the Department of Economics, University of Munich, Schackstr. 4, 80539 Munich, Germany; Johannes Ludsteck (Johannes.Ludsteck@iab.de) is at the Institut für Arbeitsmarkt- und Berufsforschung, Regensburger Straße 104, 90478 Nürnberg, Germany. Ekkehart Schlicht takes main responsibity for the theoretical part and Johannes Ludsteck for the simulations. An earlier version of the paper has been presented at the Econometric Society European Meeting in Munich 1989.

> The fact that the general conditions of life are not stationary is the source of many of the difficulties that are met with in applying economic doctrines to practical problems.
>
> ALFRED MARSHALL (1949, v.iii.24)

1 Introduction

Economists, from KEYNES (1939) to LUCAS (1976), have have been uneasy about using regression analysis for estimating behavioral coefficients from economic time-series. The assumption that the coefficients describing economic behavior remain invariant over decades, let alone centuries, seems preposterous. KEYNES (1973, 294) objected against regression analysis ("Tinbergen's method") for this, as well as other, reasons: "One of the chief dilemmas facing you is, of course, ... that the method requires not too short a series whereas it is only in a short series, in most cases, that there is a reasonable expectation that the coefficients will be fairly constant."

Further, the assumption that alternative courses of economic policy will leave the coefficients describing economic behavior unaffected points in a similar direction: "What happens if the phenomenon under investigation itself reacts on the factors by which we are explaining it?" (KEYNES 1939, 561). This point – known as the "Lucas Critique" – led to a rejection of allegedly Keynesian models in modern macroeconomics (LUCAS and SARGENT, 1979).

Some economists tried to account for these problems by allowing coefficients to change over time in a diffuse way, formalized by a random walk (COOLEY and PRESCOTT 1973, SCHLICHT 1973, ATHANS 1974). Here the standard linear model

$$y_t = a' x_t + u_t,\ a, x_t \in \mathbb{R}^n,\ u_t \sim \mathcal{N}\left(0, \sigma^2\right),\ t = 1, 2, \ldots T$$

is replaced by

$$y_t = a_t' x_t + u_t,\ u_t \sim \mathcal{N}\left(0, \sigma^2\right) \tag{1}$$

$$a_{t+1} = a_t + v_t,\quad v_t \sim \mathcal{N}(0, \Sigma) \tag{2}$$

with $y_t \in \mathbb{R}$, $x_t \in \mathbb{R}^n$ observations, $a_t \in \mathbb{R}^n$ coefficients to be estimated, $u_t \in \mathbb{R}$ normal disturbances with variance σ^2, and $v_t \in \mathbb{R}^n$ normal perturbations in the random walk

(2) with variances $\sigma_1^2, \sigma_2^2, ..., \sigma_n^2$, *viz.* a covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_1^2 & & & 0 \\ & \sigma_2^2 & & \\ & & . & \\ & & & . \\ 0 & & & \sigma_n^2 \end{pmatrix}.$$

As a straightforward estimation procedure for the time-paths of the coefficients $a' = (a_1', a_2', \dots, a_T')$, the Kalman-filter has been proposed.[1] It calculates $a_t^{Kalman} = E\{a_t | (y_1, y_2, \dots, y_t), (x_1, x_2, \dots, x_t)\}$ for all $t = 1, 2, \dots, T$.

This paper describes another method for estimating the time-paths of the coefficients. This method – termed VC (for "varying coefficients") – offers the following *theoretical* advantages:

- The state of the coefficients at time t, *viz.* a_t, is estimated by using *all* information, rather than merely observations up to t: $a_t^{VC} = E\{a_t | (y_1, y_2, \dots, y_T), (x_1, x_2, \dots, x_T)\}$. This is preferable from a theoretical point of view.

- The estimator uses an orthogonal parametrization, instead of the usual parametrization by initial values. This effaces the problems associated with estimating, or otherwise providing, initial values.

- For the variances, a maximum likelihood estimator and a related moments estimator are derived. The moments estimator has a straightforward interpretation in small samples and coincides with the likelihood estimator in large samples. This lends intuitive appeal to the maximum likelihood estimator and statistical appeal to the moments estimator.

Further, VC provides the following *practical* advantages:

- It is implemented in software packages that are freely available and easy to use (LUDSTECK 2004, SCHLICHT 2005a,b). The more elaborate implementation (LUDSTECK 2004) relies on the commercial Mathematica program and uses a global maximization of the criterion function, rather than some gradient method as typically found in implementations of the Kalman filter. It permits, in addition, the analysis of panel data, a feature that is not available in current software packages, whether free or commercial.

[1] ATHANS (1974), SCHLICHT (1977). For the treatment of time-varying coefficients with the Kalman filter see *e.g.* HAMILTON (1994, Sect. 13.8).

- Although the VC moments estimator, the VC likelihood estimator, and the Kalman filter yield almost identical results in many cases, the VC moments estimator is, from a computational point of view, sometimes superior to the other estimators in poorly conditioned (yet, we submit, practically important) cases.

From a *descriptive* point of view, the VC method has a clear-cut interpretation as a straightforward generalization of the method of least squares:

- While the method of ordinary least squares selects estimates that minimize the sum of squares $\sum_{t=1}^{T} u_t^2$, VC selects estimates that minimize the weighted sum of squares $\sum_{t=1}^{T} u_t^2 + \gamma_1 \sum_{t=2}^{T} v_1^2 + \gamma_2 \sum_{t=2}^{T} v_2^2 + ... + \gamma_n \sum_{t=2}^{T} v_n^2$, where the weights for the changes in the coefficients $\gamma_1, \gamma_2, ..., \gamma_n$ are determined by the inverse variance ratios, *i.e.* $\gamma_i = \sigma^2/\sigma_i^2$. In other words, it balances the desiderata of a good fit and parameter stability.

- The time-averages of the regression coefficients are GLS estimates of the corresponding regression with fixed coefficients, *i.e.* $\frac{1}{T}\sum_t a_t = a_{GLS}$.

Both features permit an easier interpretation of the results in relation to the original data than seems possible from the Kalman filter perspective.

The paper is organized as follows: Sections 2 to 4 gives some notation, introduce the filter and present some preliminary results. Sections 5 to 8 derive the likelihood estimator and the moments estimator and compare them. Sections 10 to 11 illustrate the estimators by means of some simulations and compare the performance with the Kalman filter, as implemented in the popular program package EVIEWS (2005). The final section 12 summarizes the argument and findings.

2 Notation

Define

$$y := \begin{pmatrix} y_2 \\ . \\ . \\ y_T \end{pmatrix}, \quad u := \begin{pmatrix} u_1 \\ u_2 \\ . \\ . \\ u_T \end{pmatrix}, \quad a := \begin{pmatrix} a_1 \\ a_2 \\ . \\ . \\ a_T \end{pmatrix}, \quad v := \begin{pmatrix} v_2 \\ v_3 \\ . \\ . \\ v_T \end{pmatrix}$$

$order$ $\quad T \times 1 \quad T \times 1 \quad Tn \times 1 \quad (T-1)\,n \times 1$

$$X := \begin{pmatrix} x_1' & & & 0 \\ & x_2' & & \\ & & . & \\ & & & \\ 0 & & & x_T' \end{pmatrix}, \quad P := \begin{pmatrix} -I_n & I_n & & & 0 \\ & -I_n & I_n & & \\ & & . & . & \\ & & & . & . \\ 0 & & & -I_n & I_n \end{pmatrix} \tag{3}$$

$order \qquad T \times Tn \qquad\qquad (T-1)\, n \times Tn$

and write (1), (2) as

$$y = Xa + u, \quad u \sim \mathcal{N}\left(0, \sigma^2 I_T\right) \tag{4}$$

$$Pa = v, \quad v \sim \mathcal{N}(0, V), \; V := I_{T-1} \otimes \Sigma. \tag{5}$$

Denote by $e_i \in \mathbb{R}^n$ the n-th column of an $n \times n$ identity matrix and define the $(T-1) \times (T-1)\, n$-matrix

$$F_i := I \otimes e_i' \tag{6}$$

that picks the time-path of the $i-$th disturbance $v_i = \left(v_{i,2}, v_{i,3}, ...v_{i,T}\right)'$ from the disturbance vector v:

$$v_i := F_i v.$$

Define further the $Tn \times n$ matrix

$$Z := \frac{1}{\sqrt{T}} \begin{pmatrix} I_n \\ I_n \\ . \\ I_n \end{pmatrix} \tag{7}$$

and note that the square matrix $\left(P', Z\right)$ is of full rank. Note further that

$$PZ = 0, \; Z'Z = I_n, \; P'\left(PP'\right)^{-1} P + ZZ' = I_{Tn} \tag{8}$$

where the last equality is implied by the identity

$$\begin{pmatrix} P' & Z \end{pmatrix} \left(\begin{pmatrix} P \\ Z' \end{pmatrix} \begin{pmatrix} P' & Z \end{pmatrix} \right)^{-1} \begin{pmatrix} P \\ Z' \end{pmatrix} = I_{Tn}.$$

3 Orthogonal Parametrization

For purposes of estimation we need a model that explains the observation x as a function of the random variables u and v. This would permit calculating the probability distribution of the observations x contingent on the parameters of the distributions of u and v, *viz.* σ^2 and Σ. The true model does not permit such an inference, though, because the matrix P in (3) is of rank $(T-1)\,n$ rather than of rank Tn and cannot be inverted. Hence v does not determine a unique y but rather the set of solutions

$$A := \left\{ a = P'\left(PP'\right)^{-1} v + Z\lambda \,\middle|\, \lambda \in \mathbb{R}^n \right\}. \tag{9}$$

For any v we have $a \in A \Leftrightarrow Pa = v$. Hence equation (4) and and the set (9) give equivalent descriptions of the relationship between a and v in this sense.

In view of (9), any solution a to $Pa = v$ can be written as

$$a = P'\left(PP'\right)^{-1} v + Z\lambda \tag{10}$$

for some $\lambda \in \mathbb{R}^n$. As $y = Xa + u$, equation (4) can be re-written as

$$y = u + XP'\left(PP'\right)^{-1} v + XZ\lambda. \tag{11}$$

The model (10), (11) will be referred to as the *equivalent orthogonally parametrized model*. It implies the *true model* (4), (5). It implies, in particular, that a_t is a random walk even though a_t depends, according to (10), on past *and* future realizations of v_t.

Equation (11) permits calculating the density of y dependent upon the parameters of the distributions of u and v and the formal parameters λ. In a second step, all these parameters – σ_u^2, Σ, and λ – can be determined by the maximum likelihood principle. This will give our likelihood estimates. Our moments estimates – to be introduced later – will build on the equivalent orthogonally parametrized model as well.

The orthogonal parametrization entails some advantages with respect to symmetry and mathematical transparency, as compared to more usual procedures, such as parametrization by initial values. By assuming some initial values $a_1 = \bar{a}_1$, the system (2) can be solved recursively, giving a as a function of v and $\bar{a}_1$, and the analysis would then proceed in a similar way as indicated above. Theoretically speaking, and with regard to likelihood estimation, all parameterizations are equivalent, but initial values are more cumbersome to implement than the formal parameters λ. It is for this reason that, in the context of Kalman filtering, initial values are estimated as posterior

means, *viz.* by running the filter back and forth in order to determine the necessary initial values $\bar{a}_1$ iteratively (AKAIKE, 1989, 61-2). The orthogonal parametrization proposed by SCHLICHT (1985, 55-8) will permit us to write down an explicit likelihood function and estimate all relevant parameters in a unified one-shot procedure.

Although the orthogonal parametrization may appear not very intuitive at first sight, it has a straightforward interpretation: The formal parameter vector $\lambda \in \mathbb{R}^n$ expresses additive shifts in the level of the parameters a that leave the disturbance vector v unaffected.

The formal parameter vector λ relates directly to the coefficient estimates of a standard GLS regression. Equation (11) can be interpreted as a standard regression for this parameter vector:

$$y = XZ\lambda + w \tag{12}$$

with $w := XP'\left(PP'\right)^{-1} v + u$ distributed normally:

$$w \sim \mathcal{N}(0, W), \quad W := XBX' + \sigma^2 I_T \tag{13}$$

with

$$B := P'\left(PP'\right)^{-1} V \left(PP'\right)^{-1} P \tag{14}$$

The maximum likelihood (Aitken, GLS) estimate $\hat{\lambda}$ satisfies

$$Z'X'W^{-1}\left(y - XZ\hat{\lambda}\right) = 0 \tag{15}$$

or

$$\hat{\lambda} = \left(Z'X'W^{-1}XZ\right)^{-1} Z'X'W^{-1}y$$

This can be related to the corresponding standard linear regression with constant coefficients. Define the matrix of observations in conventional format as

$$X^* := \sqrt{T}XZ = X\begin{pmatrix} I_n \\ I_n \\ . \\ I_n \end{pmatrix} = \begin{pmatrix} x_1' \\ x_2' \\ . \\ x_T' \end{pmatrix}. \tag{16}$$

This matrix is assumed to be of full rank:

$$r\left(X^*\right) = r\left(XZ\right) = n \tag{17}$$

Inserting (16) into (12) gives rise to

$$y = \frac{1}{\sqrt{T}} X^* \lambda + w$$

and $\hat{\beta} = \frac{1}{\sqrt{T}} \hat{\lambda}$ turns out to be the standard GLS estimator for the regression $y = X^* \beta + w$.

4 The Filter

This section derives the filter for given variances σ^2 and Σ.

For given λ and X, the vectors y and a can be viewed as realizations of random variables determined jointly by the system (10), (12) as brought about by the disturbances u and v:

$$\begin{pmatrix} y \\ a \end{pmatrix} = \begin{pmatrix} XZ \\ Z \end{pmatrix} \lambda + \begin{pmatrix} I_T & XP'\left(PP'\right)^{-1} \\ 0 & P'\left(PP'\right)^{-1} \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}$$

The marginal distribution of y is as given by (12) and (13). The conditional distribution of a for given y is

$$a \sim \mathcal{N}\left(Z\lambda + BX'W^{-1}\left(y - XZ\lambda\right), A\right) \tag{18}$$

with

$$A := B - BW^{-1}B'$$

Hence a likelihood estimator for a can be derived by plugging the Aitken estimator $\hat{\lambda}$ from (15) into (18) and calculating the mean:

$$\hat{a} := Z\hat{\lambda} + BX'W^{-1}\left(y - XZ\hat{\lambda}\right). \tag{19}$$

Note that $Z'B = 0$ implies

$$Z'\hat{a} = \hat{\lambda}$$

and hence

$$\frac{1}{T} \sum_{t=1}^{T} a_{i,t} = \beta_i, \quad i = 1, 2, ..., n \tag{20}$$

with $a_{i,t}$ as the i-th element of a_t and β_i as the i-th element of the Aitken estimate in the GLS regression $y = X^* \beta + w$. Note further that the variance-covariance matrix of w, as given in equation (13), tends to $\sigma^2 I_T$ if the the variances σ_i^2 go to zero. In this

sense, the standard regression model is covered as a special limiting case of the model discussed here: The time-averages of the estimated coefficients $a_{i,t}$ are equal to the corresponding Aitken estimates β_i. For zero variance in the coefficients, equation (11) turns into a standard unweighted linear regression.

The estimator (19) can be characterized in an easy way. For given observations X and y and any given $\hat{a}$, the estimated disturbances are

$$\hat{u} = y - X\hat{a} \tag{21}$$

$$\hat{v} = P\hat{a} \tag{22}$$

and the weighted sum of squares of these disturbances is

$$Q := \frac{1}{\sigma^2}\hat{u}'\hat{u} + \hat{v}'V^{-1}\hat{v}. \tag{23}$$

The following theorem states that the likelihood estimator $\hat{a}$ can be obtained by minimizing this expression.

Theorem 1. *Minimizing the sum of squares (23) with respect to a yields a unique solution that is numerically identical to estimator $\hat{a}$ given in (19).*

Proof. Consider first the necessary conditions for a minimum of (23). Inserting (21) and (22) into (23) and putting the derivative with respect to a to zero gives the necessary condition

$$\left(X'X + \sigma^2 P'V^{-1}P\right)a = X'y \tag{24}$$

1. It will be shown first that the system matrix

$$M := \left(X'X + \sigma^2 P'V^{-1}P\right) \tag{25}$$

is of full rank. This will establish uniqueness. For $S = \frac{1}{\sigma^2}V$ we have

$$\begin{aligned} r(M) &= r\left(\begin{pmatrix} X', & P'S^{-\frac{1}{2}} \end{pmatrix}\begin{pmatrix} X \\ S^{-\frac{1}{2}}P \end{pmatrix}\right) \\ &= r\begin{pmatrix} X', & P'S^{-\frac{1}{2}} \end{pmatrix}. \end{aligned}$$

If $\begin{pmatrix} X', & P' \end{pmatrix}$ were not of full rank, there would exist vectors $c_t \in \mathbb{R}^n$, $T = 1,2,\ldots T$, not

all of them zero, such that

$$X'c_1 = P'S^{-\frac{1}{2}} \begin{pmatrix} c_2 \\ c_3 \\ . \\ c_T \end{pmatrix} \tag{26}$$

Pre-multiplication by Z' gives

$$Z'X'c_1 = 0.$$

Together with (17) this implies $c_1 = 0$. As $P'S^{-\frac{1}{2}}$ has full rank, (26) would imply all c_t to be zero – a contradiction. Hence the system matrix is of full rank.

2. As the system matrix is positive definite, this establishes that the second order condition is fulfilled and that the solution to (24) gives indeed a global minimum.

3. We show now that (19) implies (24). Pre-multiplication of (19) by $\left(X'X + \sigma^2 P'V^{-1}P\right)$ gives

$$\begin{aligned} \left(X'X + \sigma^2 P'V^{-1}P\right)\hat{a} &= \left(X'X + \sigma^2 P'V^{-1}P\right)Z\hat{\lambda} \\ &\quad + X'XBX'W^{-1}\left(y - XZ\hat{\lambda}\right) \\ &\quad + \sigma^2 P'V^{-1}PBX'W^{-1}\left(y - XZ\hat{\lambda}\right). \end{aligned}$$

Since $PZ = 0$, the first term reduces to $X'XZ\hat{\lambda}$. Since $W = XBX' + \sigma^2 I_T$ we have $XBX'W^{-1} = I_T - \sigma^2 W^{-1}$ and the second term reduces to

$$X'\left(I_T - \sigma^2 W^{-1}\right)\left(y - XZ\hat{\lambda}\right) = X'\left(y - XZ\hat{\lambda}\right) - \sigma^2 X'W^{-1}\left(y - XZ\hat{\lambda}\right).$$

Because $P'S^{-1}PB = P'\left(PP'\right)^{-1}P$ and (8), the third term reduces to

$$\sigma^2\left(I_{Tn} - ZZ'\right)X'W^{-1}\left(y - XZ\hat{\lambda}\right).$$

By using the definition of $\hat{\lambda}$ in (15), this simplifies further to

$$\sigma^2 X'W^{-1}\left(y - XZ\hat{\lambda}\right).$$

Collecting terms gives

$$\left(X'X + \sigma^2 P'V^{-1}P\right)\hat{a} = X'y \tag{27}$$

which is identical to the definition of a given in (24).

The normal equation (27) can be equivalently expressed in terms of the variance

ratios

$$r_i := \frac{\sigma_i^2}{\sigma^2}, \quad r = \begin{pmatrix} r_1 \\ r_2 \\ . \\ r_n \end{pmatrix}. \tag{28}$$

Define the matrix of variance ratios

$$R := \mathrm{diag} r = \frac{1}{\sigma^2}\Sigma$$

and the covariance matrix

$$S := I_{T-1} \otimes R = \frac{1}{\sigma^2} V. \tag{29}$$

Using these expressions, the normal equation (27) can be written as

$$\left(X'X + P'S^{-1}P\right)\hat{a} = X'y. \tag{30}$$

or, for short

$$M\hat{a} = X'y \tag{31}$$

with M of order $nT \times nT$ as the system matrix, as in (25).

The estimates for $\hat{a}$ and for the corresponding disturbances $\hat{u}$ and $\hat{v}$ are

$$\begin{aligned} \hat{a} &= M^{-1}X'y \\ \hat{u} &= \left(I - XM^{-1}X'\right)y \\ \hat{v} &= PM^{-1}X'y. \end{aligned}$$

All these are functions of the variance ratios r and the observations $\left(y, X\right)$.

Next turn to the covariance of the estimate $\hat{a}$. As the estimate satisfies the normal equation (31) it follows together with (1) and (2) that

$$\begin{aligned} \hat{a} &= M^{-1}X'(Xa+u) \\ &= M^{-1}\left(X'Xa + X'u + P'S^{-1}Pa - P'S^{-1}Pa\right) \\ &= a + M^{-1}\left(X'u - P'S^{-1}v\right). \end{aligned} \tag{32}$$

Given a realization of the time-path of the coefficients a, the estimator $\hat{a}$ is normally distributed with mean a and covariance

$$E\left\{(\hat{a}-a)^2\right\} = \sigma^2 M^{-1}. \tag{33}$$

The system matrix (25) is determined by the observations X and the variance ratios

r. If we know in addition the variance of the error term σ^2, equation (33) gives the precision of our estimate. The next step is to determine the variance σ^2 and the variances Σ.

5 Maximum Likelihood Estimation of the Variances

This section derives a maximum-likelihood estimator for the variances.

The likelihood function for (12), (13) is

$$L\left(\sigma^2, \Sigma \right) = -\log\det W - \hat{w}' W^{-1} \hat{w} \tag{34}$$

with

$$\hat{w} = \left(y - XZ\hat{\lambda} \right)$$

The construction of the matrix W according to (13) and (14) is involved and requires the inversion of large full matrices. A considerable simplification can be achieved by expressing the likelihood in terms of the system matrix (25), the sum of squares (23), and the variances σ^2, Σ.

Theorem 2: *Disregarding constants, the likelihood function (34) can be written equivalently in terms of the variances σ^2, Σ as*

$$\begin{aligned} \mathscr{L}\left(\sigma^2, \quad \Sigma \right) \quad = \quad & -\left(\log\det M + (T-1)\log\det\Sigma - (Tn - T - n)\log\sigma^2 \right) \\ & -\left(\frac{1}{\sigma^2}\hat{u}'\hat{u} + \hat{v}' V^{-1} \hat{v} \right) \end{aligned} \tag{35}$$

Proof. Consider the first term first. Define

$$N := XP' \left(PP' \right)^{-1} S^{\frac{1}{2}}.$$

With (13), (14), and (29) we have

$$W = \sigma^2 \left(NN' + I_T \right). \tag{36}$$

Consider now

$$\begin{aligned} \frac{\det M}{\det PP'} \quad &= \quad \det M \det \left(PP' \right)^{-1} \\ &= \quad \det M \det \begin{pmatrix} \left(PP' \right)^{-1} & 0 \\ 0 & I_n \end{pmatrix} \end{aligned}$$

$$
\begin{aligned}
&= \det M \det \left(\begin{pmatrix} (PP')^{-1} & 0 \\ 0 & I_n \end{pmatrix} \begin{pmatrix} (PP')^{-1} & 0 \\ 0 & I_n \end{pmatrix} \begin{pmatrix} P \\ Z' \end{pmatrix} \begin{pmatrix} P', & Z \end{pmatrix} \right) \\
&= \det M \det \left(\begin{pmatrix} P', & Z \end{pmatrix} \begin{pmatrix} (PP')^{-1} & 0 \\ 0 & I_n \end{pmatrix} \begin{pmatrix} (PP')^{-1} & 0 \\ 0 & I_n \end{pmatrix} \begin{pmatrix} P \\ Z' \end{pmatrix} \right) \\
&= \det \left(\begin{pmatrix} (PP')^{-1} & 0 \\ 0 & I_n \end{pmatrix} \begin{pmatrix} P \\ Z' \end{pmatrix} M \begin{pmatrix} P', & Z \end{pmatrix} \begin{pmatrix} (PP')^{-1} & 0 \\ 0 & I_n \end{pmatrix} \right) \\
&= \det \begin{pmatrix} (PP')^{-1} PMP' (PP')^{-1}, & (PP')^{-1} PMZ \\ Z'MP' (PP')^{-1} & Z'MZ \end{pmatrix} \\
&= \det \begin{pmatrix} (PP')^{-1} PX'XP' (PP')^{-1} + S^{-1}, & (PP')^{-1} PX'XZ \\ Z'X'XP' (PP')^{-1} & Z'X'XZ \end{pmatrix} \\
&= \det \begin{pmatrix} (PP')^{-1} PX'XP' (PP')^{-1} + S^{-1}, & (PP')^{-1} PX'XZ \\ Z'X'XP' (PP')^{-1} & Z'X'XZ \end{pmatrix} \\
&= \det \begin{pmatrix} S & 0 \\ 0 & I_n \end{pmatrix}^{-1} \det \begin{pmatrix} N'N + I_{(T-1)n}, & N'XZ \\ Z'X'N & Z'X'XZ \end{pmatrix} \\
&= \det S^{-1} \det \left(N'N + I_{(T-1)n} \right) \cdot \\
&\qquad \det \left(Z'X'XZ - N'XZ \left(N'N + I_{(T-1)n} \right)^{-1} Z'X'N \right).
\end{aligned}
$$

Because

$$Z'X'XP' = 0$$

we have

$$Z'X'N = 0$$

and we obtain the result

$$\frac{\det M}{\det PP'} = \det S^{-1} \det \left(N'N + I_{(T-1)n} \right) \det \left(Z'X'XZ \right). \tag{37}$$

As $N'N$ and NN' have all non-zero eigenvalues in common, and adding the identity matrix shifts the eigenvalues by unity, the products of the eigenvalues of $\left(N'N + I_{(T-1)n}\right)$ and of $\left(NN' + I_T\right)$ and the corresponding determinants are identical. In view of (36) we obtain from (37)

$$\frac{\det M}{\det (PP')} = \frac{\det W \det \left(Z'X'XZ \right)}{(\det \Sigma)^{T-1}} \left(\sigma^2 \right)^{Tn-T-n}$$

and hence

$$\begin{aligned}\log\det W &= \log\det M + (T-1)\log\det\Sigma - (Tn-T-n)\log\sigma^2 \\ &\quad -\log\det\left(PP'\right) - \log\det\left(Z'X'XZ\right).\end{aligned}$$

For any given problem, the last two terms are constant and can be dropped. Hence the first term in the likelihood function (35) is established.

Consider the second term next. The sum of squares at $a = \hat{a}$ is

$$\begin{aligned}Q: &= \frac{1}{\sigma^2}\hat{u}'\hat{u} + \hat{v}'V^{-1}\hat{v} \\ &= \frac{1}{\sigma^2}\left(y' - \hat{a}'X'\right)\left(y - X\hat{a}\right) + \hat{a}'P'V^{-1}P\hat{a} \\ &= \frac{1}{\sigma^2}y'y - \frac{2}{\sigma^2}y'X\hat{a} + \frac{1}{\sigma^2}\hat{a}'\left(X'X + \sigma^2P'V^{-1}P\right)\hat{a} \\ &= \frac{1}{\sigma^2}y'y - \frac{1}{\sigma^2}y'X\hat{a}\end{aligned}$$

From (19) and (30) we find $y'X'\hat{a} = y'y - y'\hat{w} + y'XBX'W^{-1}\hat{w}$. Inserting this into (35) yields

$$Q = \frac{1}{\sigma^2}\left(\left(\hat{w}' + \hat{\lambda}'Z'X'\right)\hat{w} - \left(\hat{w}' + \hat{\lambda}'Z'X'\right)XBX'W^{-1}\hat{w}\right). \tag{38}$$

From the definition (13) we find

$$XBX'W^{-1} = I_T - \sigma^2W^{-1}.$$

and (38) reduces to

$$Q = \hat{w}'W^{-1}\hat{w} - \hat{\lambda}'Z'X'W^{-1}\hat{w}. \tag{39}$$

As $w = y - XZ\hat{\lambda}$, (15) implies that the last term in (39) vanishes and we obtain

$$Q = \frac{1}{\sigma^2}\hat{u}'\hat{u} + \hat{v}'V^{-1}\hat{v} = \hat{w}'W^{-1}\hat{w}$$

which completes the proof.

6 Moments Estimation of the Variances

The maximum likelihood estimates of the variances lack intuitive appeal. The moments estimator that will be developed in this section has, for any sample size, a straightforward interpretation: It is defined by the property that estimated variances are equalized to their expectations. It will turn out that the estimators coincide for

large samples, and do not differ much even in small samples, the intuitive appeal of the moments estimator carries over to the likelihood estimator, and the attractive large-sample properties of the likelihood estimator carry over to the moments estimator.

The estimated coefficients $\hat{a}$ along with the estimated disturbances $\hat{u}$ and $\hat{v}$ are random variables brought about by realizations of the random variables u and v. Consider $\hat{u} = y - X\hat{a} = X(a - \hat{a}) + u$ first. With (32) we obtain

$$\hat{u} = \left(I_T - XM^{-1}X'\right)u + XM^{-1}P'S^{-1}v.$$

Regarding $\hat{v}$, pre-multiplying (32) by F_i from (6) yields

$$\hat{v}_i = F_i\left(I_{(T-1)n} - PM^{-1}P'S^{-1}\right)v + F_iPM^{-1}X'u$$

Thus $\hat{u}$ and $\hat{v}_i$ are linear functions of the normal random variables u and v, and their expected squared errors can be calculated.

$$\begin{aligned}
E\{\hat{u}'\hat{u}\} &= E\{\left(u'\left(I_T + XM^{-1}X'\right) + v'S^{-1}PM^{-1}X'\right)\cdot \\
&\quad \left(\left(I_T - XM^{-1}X'\right)u + XM^{-1}P'S^{-1}v\right)\} \\
&= E\{u'\left(I_T - XM^{-1}X'\right)\left(I_T - XM^{-1}X'\right)u\} + \\
&\quad + E\{v'S^{-1}PM^{-1}X'XM^{-1}P'S^{-1}v\} \\
&= \mathrm{tr}E\{u'\left(I_T - XM^{-1}X'\right)\left(I_T - XM^{-1}X'\right)u\} + \\
&\quad + \mathrm{tr}E\{v'S^{-1}PM^{-1}X'XM^{-1}P'S^{-1}v\}S \\
&= \mathrm{tr}E\{\left(I_T - XM^{-1}X'\right)uu'\left(I_T - XM^{-1}X'\right)\} + \\
&\quad + \mathrm{tr}E\{XM^{-1}P'S^{-1}vv'S^{-1}PM^{-1}X'\} \\
&= \mathrm{tr}\sigma^2\left(I_T - XM^{-1}X'\right)\left(I_T - XM^{-1}X'\right) + \mathrm{tr}\sigma^2XM^{-1}P'S^{-1}PM^{-1}X' \\
&= \sigma^2\mathrm{tr}\left(\left(I_T - XM^{-1}X'\right)\left(I_T - XM^{-1}X'\right) + XM^{-1}P'S^{-1}PM^{-1}X'\right) \\
&= \sigma^2\mathrm{tr}\left(I - 2XM^{-1}X' + XM^{-1}X'XM^{-1}X' + XM^{-1}P'S^{-1}PM^{-1}X'\right) \\
&= \sigma^2\mathrm{tr}\left(I_T - 2XM^{-1}X' + XM^{-1}\left(X'X + P'S^{-1}P\right)M^{-1}X'\right) \\
&= \sigma^2\mathrm{tr}\left(I_T - XM^{-1}X'\right) \\
&= \sigma^2\left(T - \mathrm{tr}XM^{-1}X'\right)
\end{aligned}$$

Next, consider the variance of $\hat{v}_i$.

$$
\begin{aligned}
E\left\{\hat{v}_i'\hat{v}_i\right\} &= E\left\{\left(u'XM^{-1}P'F_i' + v'\left(I_{(T-1)n} - S^{-1}PM^{-1}P'\right)F_i'\right)\cdot\right. \\
&\quad \left.\left(F_iPM^{-1}X'u + F_i\left(I_{(T-1)n} - PM^{-1}P'S^{-1}\right)v\right)\right\} \\
&= E\left\{u'XM^{-1}P'F_i'F_iPM^{-1}X'u\right\} + \\
&\quad + E\left\{v'\left(I_{(T-1)n} - S^{-1}PM^{-1}P'\right)F_i'F_i\left(I_{(T-1)n} - PM^{-1}P'S^{-1}\right)v\right\} \\
&= E\left\{\mathrm{tr}u'XM^{-1}P'F_i'F_iPM^{-1}X'u\right\} + \\
&\quad + E\left\{\mathrm{tr}v'\left(I_{(T-1)n} - S^{-1}PM^{-1}P'\right)F_i'F_i\left(I_{(T-1)n} - PM^{-1}P'S^{-1}\right)v\right\} \\
&= \mathrm{tr}E\left\{F_iPM^{-1}X'uu'XM^{-1}P'F_i'\right\} + \\
&\quad + \mathrm{tr}E\left\{F_i\left(I_{(T-1)n} - PM^{-1}P'S^{-1}\right)vv'\left(I_{(T-1)n} - S^{-1}PM^{-1}P'\right)F_i'\right\} \\
&= \sigma^2\mathrm{tr}F_iPM^{-1}X'XM^{-1}P'F_i' + \\
&\quad + \sigma^2\mathrm{tr}F_i\left(I_{(T-1)n} - PM^{-1}P'S^{-1}\right)S\left(I_{(T-1)n} - S^{-1}PM^{-1}P'\right)F_i' \\
&= \sigma^2\mathrm{tr}F_iPM^{-1}X'XM^{-1}P'F_i' + \\
&\quad + \sigma^2\mathrm{tr}F_i\left(I_{(T-1)n} - PM^{-1}P'S^{-1}\right)S\left(I_{(T-1)n} - S^{-1}PM^{-1}P'\right)F_i' \\
&= \sigma^2\mathrm{tr}F_iSF_i' + \sigma^2\mathrm{tr}F_iPM^{-1}X'XM^{-1}P'F_i' + \\
&\quad + \sigma^2\mathrm{tr}F_i\left(-2PM^{-1}P' + PM^{-1}P'S^{-1}PM^{-1}P'\right)F_i' \\
&= (T-1)\sigma_i^2 + \\
&\quad + \sigma^2\mathrm{tr}F_i\left(PM^{-1}\left(X'X + P'S^{-1}P\right)M^{-1}P' - 2PM^{-1}P'\right)F_i' \\
&= (T-1)\sigma_i^2 - \sigma^2\mathrm{tr}F_iPM^{-1}P'F_i'
\end{aligned}
$$

The moments estimators are defined by selecting variances σ^2 and σ_i^2, $i = 1,2,...,n$ such that the expected moments $E\left\{\hat{u}'\hat{u}\right\}$ and $E\left\{\hat{v}_i'\hat{v}_i\right\}$, $i = 1,2,...,n$ are equalized to the estimated moments $\hat{u}'\hat{u}$ and $\hat{v}_i'\hat{v}_i$, $i = 1,2,...,n$. As both the expected moments and the estimated moments are functions of the variances, the moments estimators, denoted by $\check{\sigma}^2$ and $\check{\sigma}_i^2$, $i = 1,2,...,n$, respectively, are defined as a fix point of the system

$$
\frac{\hat{u}'\hat{u}}{\check{\sigma}^2} = T - \mathrm{tr}XM^{-1}X' \tag{40}
$$

$$
\frac{\hat{v}_i'\hat{v}_i}{\check{\sigma}_i^2} = (T-1) - \frac{\check{\sigma}^2}{\check{\sigma}_i^2}\mathrm{tr}F_iPM^{-1}P'F_i', \quad i = 1,2,...n \tag{41}
$$

Summing (40) and (41) gives the corresponding calculated sum of squares (23):

$$Q = (n+1)\,T - n - \mathrm{tr} M^{-1} X'X - \mathrm{tr}\left(\sum_{i=1}^{n} \frac{\check{\sigma}^2}{\check{\sigma}_i^2} M^{-1} P' F_i' F_i P \right) \tag{42}$$

By definitions (3), (6), and (29) of P , F_i , and S we have

$$\sum_{i=1}^{n} \frac{\check{\sigma}^2}{\check{\sigma}_i^2} P' F_i' F_i P = P' S^{-1} P \tag{43}$$

and (42) reduces to

$$Q = (n+1)\,T - n - \mathrm{tr}\left(M^{-1} \left(X'X + P' S^{-1} P \right) \right) = T - n$$

which implies

$$\check{\sigma}^2 = \frac{1}{T-n} \left(\hat{u}' \hat{u} + \hat{v}' S^{-1} \hat{v} \right).$$

7 Another Representation of the Moments Estimator

The relationship between the likelihood estimator and the moments estimator can be elucidated with the aid of the criterion function

$$\mathscr{K}\left(\sigma^2, \quad \Sigma \right) = -\left(\log\det M + (T-1) \log\det \Sigma - T\,(n-1) \log \sigma^2 \right) - \frac{1}{\sigma^2} \hat{u}' \hat{u} - \hat{v}' V^{-1} \hat{v} \tag{44}$$

defined in analogy to the concentrated likelihood function (35).

Theorem 3. *Minimization of the criterion function (44) with respect to the variances σ^2 and Σ yields the moments estimators as defined in (40), (41).*

Proof. Note that the envelope theorem together with (43) implies

$$\frac{\partial}{\partial \sigma^2} \left(\frac{1}{\sigma^2} \hat{u}' \hat{u} + \hat{v}' V^{-1} \hat{v} \right) = -\frac{1}{\sigma^4} \hat{u}' \hat{u} \tag{45}$$

$$\frac{\partial}{\partial \sigma_i^2} \left(\frac{1}{\sigma^2} \hat{u}' \hat{u} + \hat{v}' V^{-1} \hat{v} \right) = -\frac{1}{\sigma_i^4} \hat{v}_i{}' \hat{v}_i \tag{46}$$

In view of (43) we obtain further

$$\begin{aligned}\frac{\partial \log\det M}{\partial \sigma^2} &= \mathrm{tr} P' S^{-1} P M^{-1} \\ \frac{\partial \log\det M}{\partial \sigma_i^2} &= -\frac{\sigma^2}{\sigma_i^4}\mathrm{tr}\left(M^{-1}P'F_i'F_iP\right).\end{aligned}$$

With these results we find

$$\frac{\partial \mathcal{K}}{\partial \sigma^2} = -\mathrm{tr} P' S^{-1} P M^{-1} + \frac{T(n-1)}{\sigma^2} + \frac{1}{\sigma^4}\hat{u}'\hat{u} = 0 \tag{47}$$

$$\frac{\partial \mathcal{K}}{\partial \sigma_i^2} = -\frac{\sigma^2}{\sigma_i^4}\mathrm{tr} P' F_i' F_i P M^{-1} - (T-1)\frac{1}{\sigma_i^2} + \frac{1}{\sigma_i^4}\hat{v}_i'\hat{v}_i = 0 \tag{48}$$

By definition (25) we have

$$\left(X'X + \sigma^2 P' S^{-1} P\right) M^{-1} = I$$

and therefore

$$\begin{aligned}\sigma^2 \mathrm{tr}\left(P' S^{-1} P M^{-1}\right) &= \mathrm{tr}\left(I_{Tn} - X'X M^{-1}\right) \\ &= Tn - \mathrm{tr}\left(X M^{-1} X'\right)\end{aligned}$$

and (47) can be written as

$$T - \mathrm{tr}\left(X M^{-1} X'\right) = \frac{\hat{u}'\hat{u}}{\sigma^2}$$

which is identical to (40), while (48) reduces directly to (41).

8 The Relationship Between the Likelihood and the Moments Estimator

The likelihood estimates $\hat{\sigma}^2$ and $\hat{\Sigma}$ maximize $\mathcal{L}()$ and the moments estimates $\check{\sigma}^2$ and $\check{\Sigma}$ maximize $\mathcal{K}()$. Hence we have

$$\begin{aligned}\mathcal{L}\left(\hat{\sigma}^2, \hat{\Sigma}\right) &\geq \mathcal{L}\left(\check{\sigma}^2, \check{\Sigma}\right) \\ \mathcal{K}\left(\hat{\sigma}^2, \hat{\Sigma}\right) &\leq \mathcal{K}\left(\check{\sigma}^2, \check{\Sigma}\right)\end{aligned}$$

and

$$\mathscr{K}\left(\begin{array}{cc}\check{\sigma}^2, & \check{\Sigma}\end{array}\right)-\mathscr{L}\left(\begin{array}{cc}\check{\sigma}^2, & \check{\Sigma}\end{array}\right)\geq\mathscr{K}\left(\begin{array}{cc}\hat{\sigma}^2, & \hat{\Sigma}\end{array}\right)-\mathscr{L}\left(\begin{array}{cc}\hat{\sigma}^2, & \hat{\Sigma}\end{array}\right) \tag{49}$$

with strict inequality if the estimators differ.

The difference between the moments criterion (44) and the likelihood criterion (35) is

$$\mathscr{K}\left(\begin{array}{cc}\sigma^2, & \Sigma\end{array}\right)-\mathscr{L}\left(\begin{array}{cc}\sigma^2, & \Sigma\end{array}\right)=n\log\sigma^2$$

and (49) implies

$$\check{\sigma}^2\geq\hat{\sigma}^2 \tag{50}$$

with strict inequality if the estimators differ.

For purposes of comparison and computation it is useful to parametrize the likelihood function and the criterion function by the variance σ^2 and the variance ratios $r=(r_1,r_2,...,r_n)$ instead of the variances σ^2 and Σ. As $\Sigma=\sigma^2\mathrm{diag}\,r$ we have

$$\mathscr{L}\left(\sigma^2,\sigma^2\mathrm{diag}r\right) = -\log\det M-(T-1)\sum_{i=1}^{n}\log r_i+(T-1)\log\sigma^2-\frac{1}{\sigma^2}Q^*(r) \tag{51}$$

$$\mathscr{K}\left(\sigma^2,\sigma^2\mathrm{diag}r\right) = -\log\det M-(T-1)\sum_{i=1}^{n}\log r_i+(T-n)\log\sigma^2-\frac{1}{\sigma^2}Q^*(r) \tag{52}$$

with

$$Q^*(r) := \hat{u}'\hat{u}+\hat{v}'S^{-1}\hat{v}=\sigma^2Q.$$

Maximizing these functions with respect to σ^2 yields the necessary conditions

$$\hat{\sigma}^2 = \frac{1}{T-1}Q^*(\hat{r}) \tag{53}$$

$$\check{\sigma}^2 = \frac{1}{T-n}Q^*(\check{r}) \tag{54}$$

for maxima of the functions $\mathscr{L}$ and $\mathscr{K}$, where $\hat{r}$ and $\check{r}$ denote the variance ratios of the likelihood and the moments estimator, respectively.

Inserting (53) into (51) and (54) into (52), and disregarding constants, yields the concentrated likelihood function $\mathscr{L}^*$ and the concentrated criterion function $\mathscr{K}^*$:

$$\mathscr{L}^*(r) = -\log\det M - (T-1)\log Q^*(r) - (T-1)\sum_{i=1}^{n}\log r_i \tag{55}$$

$$\mathscr{K}^*(r) = -\log\det M - (T-n)\log Q^*(r) - (T-1)\sum_{i=1}^{n}\log r_i \tag{56}$$

We note that

$$\begin{aligned}\mathscr{L}^*\left(\ \hat{r}\ \right) &\geq \mathscr{L}^*\left(\ \check{r}\ \right)\\ \mathscr{K}^*\left(\ \hat{r}\ \right) &\leq \mathscr{K}^*\left(\ \check{r}\ \right)\end{aligned}$$

and hence

$$\mathscr{K}^*\left(\ \check{r}\ \right) - \mathscr{L}^*\left(\ \check{r}\ \right) \geq \mathscr{K}^*\left(\ \hat{r}\ \right) - \mathscr{L}^*(\hat{r})$$

or equivalently

$$Q^*(\check{r}) \geq Q^*(\hat{r})\,.$$

As Q^* is decreasing in the r_i's (compare (46), we will expect by and large smaller variance ratios for the moments estimator as compared to the likelihood estimator, and hence slightly greater stability in the estimated coefficients over time, and, in view of (50) a larger variance in the disturbance u.

For long time series, the likelihood criterion (55) and the moments criterion (56) are practically identical and the estimates will coincide.

9 Computation

Note that it is only necessary to estimate the variance ratios in a first step. Once these ratios are determined, the variances follow directly through (53) or (54) and (28).

Regarding computation of the variance ratios, there are basically two strategies available. One is to start from (55) or (56) and determine the variance ratios r by a maximization routine. This strategy is implemented in LUDSTECK (2004). The other is to use a gradient method. The VC-package by SCHLICHT (2005a) uses this approach for the moments estimator. It proceeds from (40), (41) and (54). These equations imply

$$\frac{\hat{v}_i'\hat{v}_i}{Q^*} = \frac{\check{r}_i(T-1)+\mathrm{tr}F_iPM^{-1}P'F_i'}{(T-n)}, \quad i=1,2,...n. \tag{57}$$

and therefore

$$\check{r}_i = \frac{1}{T-1}\left(\frac{(T-n)\,\hat{v}_i'\hat{v}_i}{Q^*} - \mathrm{tr}F_iPM^{-1}P'F_i'\right)$$

or

$$\hat{r}_i = \frac{1}{T-1}\left(\frac{(T-1)\,\hat{v}_i'\hat{v}_i}{Q^*} - \mathrm{tr}F_iPM^{-1}P'F_i'\right)$$

The iteration starts with some initial values for the variance ratios r^0 and computes new ratios of estimated variances on the left-hand side of (57) and the corresponding ratio of expected variances on the right-hand side. If the ratio of estimated variances is larger than the ratio of expected variances, the corresponding variance ratio r_i is increased, or, in the opposite case, reduced. This adjustment continues until (57) is satisfied for all variance ratios.

Both approaches – the direct and the gradient approach – do not require matrix inversions, which may pose problems for large systems as the the inverse of the sparse $Tn\times Tn$ band matrix M is full. The solution to the normal equation can be computed in both cases conveniently by a Cholesky decomposition of the system matrix M. The determinant of system matrix M – which is needed in case of direct maximization – or the traces in (57) – which are needed in the case of the gradient method – can be determined in the course of the Cholesky decomposition as by-products.

10 Practical Performance

In order to convey an impression about how the estimation procedure works in practice, we discuss in the following a small example and some simulations. Estimation involves in all cases minimizing (23), which reduces to minimizing the weighted sum of squares

$$u'u + \sum_i \gamma_i v_i'v_i$$

with weights $(\gamma_1,\gamma_2,...,\gamma_n) = \left(\frac{\sigma^2}{\sigma_1^2},\frac{\sigma^2}{\sigma_2^2},...,\frac{\sigma^2}{\sigma_n^2}\right)$. These weights, rather than the variances, determine the estimated time-paths of the coefficients for any given set of observations. Hence we concentrate in the following on the estimation of these weights, rather than the variances.

Consider first the simplest case of two explanatory variables – an intercept term

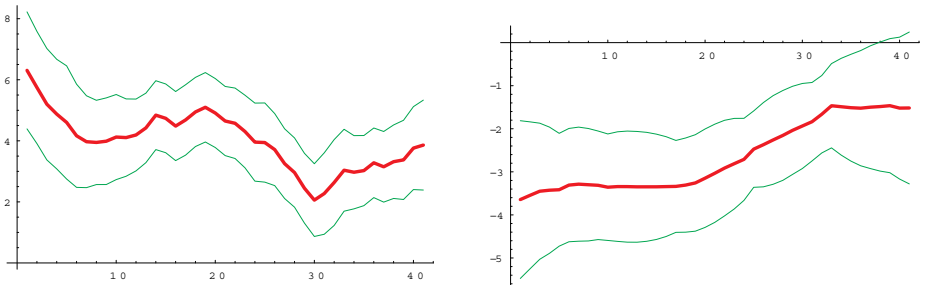


Figure 1: Time-varying intercept and slope for equation (58) with 95 percent confidence bands with data for Germany 1951-91.

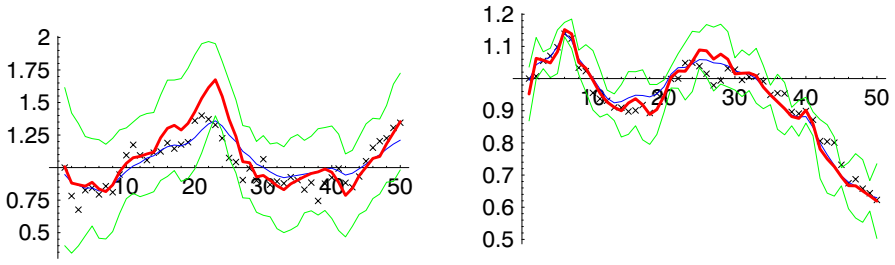
and a single exogenous variable. The following model

$$dY_t = a_t + b_t du_t + \varepsilon_t \tag{58}$$

with dY_t as the percentage change in the GDP, du_t the change in the unemployment rate, ε_t as the error term, and a_t, b_t the time-varying coefficients of intercept and slope has been estimated. Taking data from Germany (provided in SCHLICHT 2005a), the moments estimator yields $\check{\gamma}_a = 3.48$ and $\check{\gamma}_b = 11.8$, $\check{\sigma}^2_\varepsilon = 1.22$, $\check{\sigma}^2_a = 0.35$, and $\check{\sigma}^2_b = 0.10$. The corresponding likelihood estimates are $\hat{\gamma}_a = 2.83$, $\hat{\gamma}_b = 10.3$, $\hat{\sigma}^2_\varepsilon = 1.15$, $\hat{\sigma}^2_a = 0.4$, and $\hat{\sigma}^2_b = 0.11$. The estimates obtained by the Kalman-Filter, as implemented in the program EVIEWS (2005), are $\gamma^{EV}_a = 2.25$, $\gamma^{EV}_b = 8.88$, $\sigma^{2\,EV}_\varepsilon = 1.01$, $\sigma^{2\,EV}_a = 0.45$, and $\sigma^{2\,EV}_b = 0.11$. The estimates for the time-paths of the coefficients obtained from these three sets of estimators are practically indistinguishable. The estimated time paths for intercept a_t and slope b_t together with their 95 percent confidence bands (here for the moments estimator) are given in Figure 1. For comparison we note the OLS estimates $\bar{a}^{OLS} = 4.22$ and $\bar{b}^{OLS} = -2.98$ with standard errors $SE(a)^{OLS} = 0.27$ and $SE(b)^{OLS} = 0.30$ whereas the corresponding averages of the time-varying specification are $\bar{a}^{VC} = 4.0$ and $\bar{b}^{VC} = -2.65$ with average standard errors $SE\left(a^{VC}\right) = 0.66$ and $SE\left(b^{VC}\right) = 0.62$. Hence the averages of the results of the VC method are similar to the results obtained by ordinary least squares – see the discussion around equation (20) above.

All this does not tell very much, however, about how well the method recovers the variances and time-paths of the coefficients. In order to obtain an impression about this, some simulations were conducted.

The first exploration was to assume a model with an intercept term a_t and a single explanatory variable x_t with coefficient b_t. A time series for the explanatory



(a) Intercept (b) Slope

Figure 2: Optimally calculated expectations (thin lines) and VC estimates (thick lines) for intercept (left) and slope (right), together with the realizations of the coefficients (x) and the VC confidence bands. The example has been selected to visually exhibit differences between the true expectations and the VC estimates; usually the weights are estimated better and the curves lie quite close together. As the estimated smoothing weights are considerably smaller than the true weights, the time-paths of the VC estimates are less smooth than the true expectations (True weights are $\gamma_a = 10$ and $\gamma_b = 100$, while the estimated weights are $\check{\gamma}_a = 1.60$ and $\check{\gamma}_b = 14.76$ here. The true variances are $\sigma_u^2 = 0.1$, $\sigma_a^2 = 0.01$, and $\sigma_b^2 = 0.001$, the estimated variances are $\check{\sigma}_u^2 = 0.040$, $\check{\sigma}_a^2 = 0.025$, and $\check{\sigma}_b^2 = 0.0029$.)

variable was generated with $x_t \sim \mathcal{N}(0,100)$, $t = 1,2,\ldots,50$. Further it was assumed that $u_t \sim \mathcal{N}(0,0.1)$, $(a_t - a_{t-1}) \sim \mathcal{N}(0,0.01)$, and $(b_t - b_{t-1}) \sim \mathcal{N}(0,0.001)$. Typically the optimally computed expectations of the time paths (calculated by using the true variances) and the VC estimates lie very close together. Figure 2 illustrates a somewhat atypical run with estimated smoothing weights that deviate from the true smoothing weights by the order of five. The optimally estimated time-paths of the coefficients (based on the true variances) and the estimated time-paths (based on the estimated coefficients) move together. This illustrates the general impression that the filtering results, especially the qualitative time-patterns, are not extremely sensitive with regard to the weights used for filtering.

It is, obviously, never possible to extract the movement of the true coefficients from the data, irrespective how long the time series is. (Only the estimation of the weights will improve with the length of the time series.) The best that can be done is to estimate the expectations of the coefficients. Given the variances, the VC estimate (which is the mean of a normally distributed vector) is optimal and cannot be improved upon, and the standard of comparison must be the estimates obtained with optimal weights, as in Figure 2.

The distribution of the weights in the above setting is illustrated in Figure 3. The time series for x, u, and v have been generated as described above and the VC

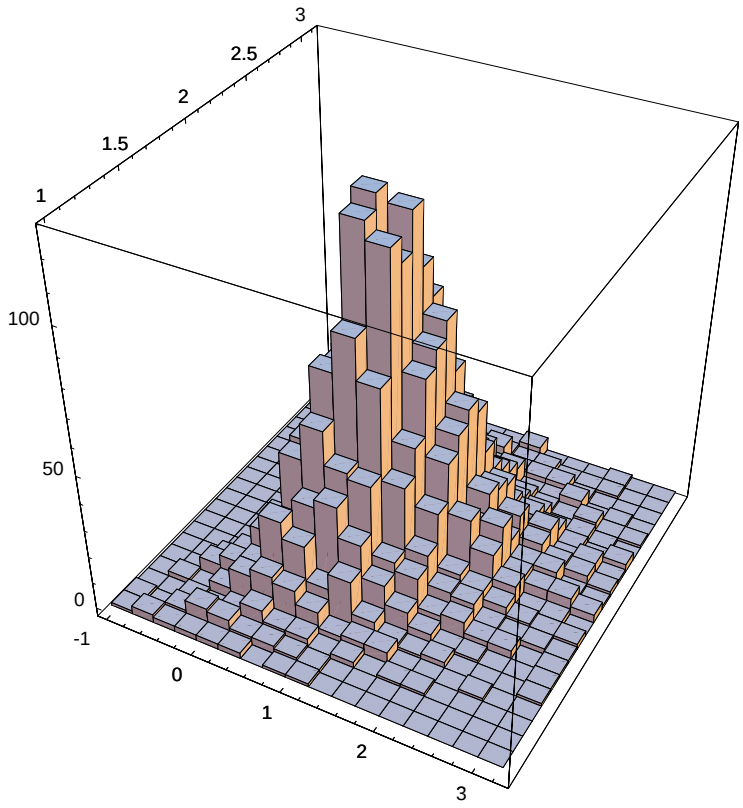
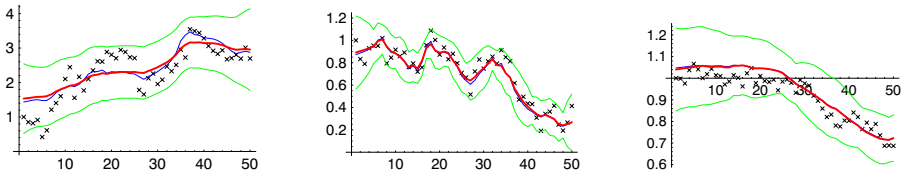


Figure 3: Histogram of estimates for the $\log_{10}$ weights. The theoretical values are $\log_{10}\gamma_a = 1$ and $\log_{10}\gamma_b = 2$. The distribution of estimates clusters around this peak. ($T = 50$, 5000 trials.)

moments estimation applied 5000 times. The histogram illustrates that the estimates cluster around their theoretical values.

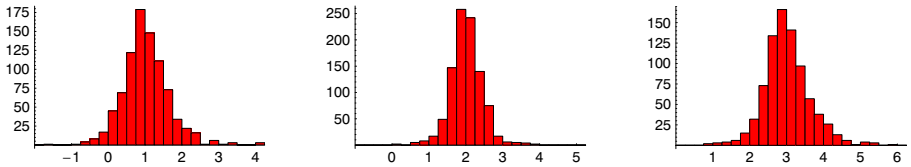
Consider now the case of three exogenous variables: An intercept term $a_{1,t}$, and two explanatory variables $x_{2,t}$ and $x_{3,t}$ with associated coefficients $a_{2,t}$ and $a_{3,t}$. (The first explanatory variable is always taken as unity, *i.e.* $x_{1,t} = 1$.) Thus the model is summarized by

$$\begin{aligned} y_t &= a_{1,t} + a_{2,t}x_{2,t} + a_{3,t}x_{3,t} + u_t\,,\ t = 1,2,\ldots,T \\ a_{i,t} &= a_{i,t-1} + v_i\,,\ i = 1,2,3,\ldots n,\ ,\ t = 2,3,\ldots,T \end{aligned} \tag{59}$$



(a) a_1 (b) a_2 (c) a_3

Figure 4: The time paths of the coefficients a_1, a_2, and a_3: Theoretical expectations (thin lines), VC estimates (thick lines) and values (x) , with estimated confidence bands. The theoretical weights are $\gamma_1 = 10$, $\gamma_2 = 100$, and $\gamma_3 = 1000$, the estimated weights are $\check{\gamma}_1 = 28.2$, $\check{\gamma}_2 = 177.4$, and $\check{\gamma}_3 = 994.9$. The theoretical variances are $\sigma_u^2 = 1$, $\sigma_1^2 = 0.1$, $\sigma_2^2 = 0.01$, and $\sigma_3^2 = 0.001$, the estimated variances are $\check{\sigma}_u^2 = 1.30$, $\check{\sigma}_1^2 = 0.046$, $\check{\sigma}_2^2 = 0.0073$, and $\check{\sigma}_3^2 = 0.0013$.



(a) $\log_{10}\check{\gamma}_1$ (b) $\log_{10}\check{\gamma}_2$ (c) $\log_{10}\check{\gamma}_3$

Figure 5: Histograms for the $\log_{10}$ weight estimates. The theoretical weights are $\log_{10}\gamma_1 = 1$, $\log_{10}\gamma_2 = 2$, and $\log_{10}\gamma_3 = 3$. The estimates cluster around these values. ($T = 50$, 1000 trials.)

with

$$u_t \sim \mathcal{N}\left(0, \sigma_u^2\right), \; v_{i,t} \sim \mathcal{N}\left(0, \sigma_i^2\right).$$

We follow, again, the example given in the Mathematica notebook (LUDSTECK, 2004) to permit easy replication. The explanatory variables $x_{2,t}$ and $x_{3,t}$ are generated as normally distributed random variables, both with variance 100 and contemporaneous covariance 50. The number of periods is taken as $T = 50$. For our example we take $\sigma_u^2 = 1.$, $\sigma_1^2 = 0.1$, $\sigma_2^2 = 0.01$, and $\sigma_3^2 = 0.001$. A typical run is illustrated Figure 4. Taking 1000 replications yields the distribution of estimates for the variance ratios given in Figure 5.

11 The VC Estimator and the Kalman Filter: Some Simulations

The VC estimator has been devised specifically for the model with time-varying coefficients (1), (2) that can also be estimated by means of the Kalman filter (HAMILTON 1994, Sect. 13.8). It is of interest, therefore, to compare the practical performance of the estimators by means of a simulation study. We concentrate on the estimation of weights, again, because the weights are determining the estimation of the time path of the coefficients in both cases. Given the weights, the estimation of the time-paths of the coefficients in VC is theoretically optimal, anyway, and the Kalman filter, being one-sided, can not do better. (We note however that, for given weights, both methods seem to produce nearly identical estimates for the time-paths of the coefficients.)

For purposes of comparison, we have conducted a small Monte-Carlo simulation study for some plausible parameter combinations. The Kalman-filter computations are performed using the popular statistical program package EVIEWS (2005). With regard to the Eviews estimates it is to be noted that the package uses a gradient method that requires initial values for the variances. We have supplied the *true* values (which are, of course, in practical applications, unavailable) as initial values for the Eviews estimates throughout. Hence the practical performance of the program may be overstated in our examples.[2] For the VC estimator, we have used the moments rather than the likelihood version because some preliminary checks suggested to us that the moments estimator is to preferred in short time series and equivalent to the likelihood estimator for longer time series.

As in the previous illustrations, the simulation is based on a simple model with an intercept term a_t and a slope coefficient b_t that describes the effect of a single explanatory variable $x_t \sim \mathcal{N}(0,5)$. The disturbance term is taken as $u_t \sim \mathcal{N}(0,1)$ in all simulations. The starting values of the coefficients are taken as $a_1 = 1$ and $b_1 = 1$. Thus the variance of the disturbance u_t is relatively large, as compared to the variance in the explanatory variable x_t, and given the average levels of the coefficients a_t and b_t around unity. This comparatively ill-conditioned setting has been selected to bring out differences between the methods that would not be visible otherwise.

We have simulated time series of different length and for different combinations of variances. The cases studied are given by all combinations of

- the length of the time series from $T \in \{40, 60, 80, 100\}$
- variances from $\sigma_a^2 \in \{0.001, 0.01\}$ and $\sigma_b^2 \in \{0.001, 0.01\}$

[2] The fact that Eviews performed much worse if good starting values were not supplied indicates a possibly crucial problem for real applications.

For each combination of these parameters we have generated $K = 1000$ realizations of the time series by obtaining pseudorandom draws for the x_t, u_t, $\nu_{a,t}$, and $\nu_{b,t}$. It is, however, problematic, to directly compare the estimates of σ_u, σ_a, and σ_b with their theoretical values because the pseudorandom variates may deviate considerably from their theoretical values in small samples. In order to avoid this possible source of error, we have standardized the pseudorandom series by multiplying them with appropriate scale factors.

The relative performance of the VC moments estimator in comparison to that of the Kalman filter is then evaluated by computing the mean squared error of the logarithms of the estimated weights

$$MSE_{10}(\gamma_i) := \frac{1}{K}\sum_{k=1}^{K}\left(\log_{10}\gamma_{i,k}^{estimated} - \log_{10}\gamma_{i,k}^{true}\right)^2$$

for $i = a, b$ and all combinations of the parameters.

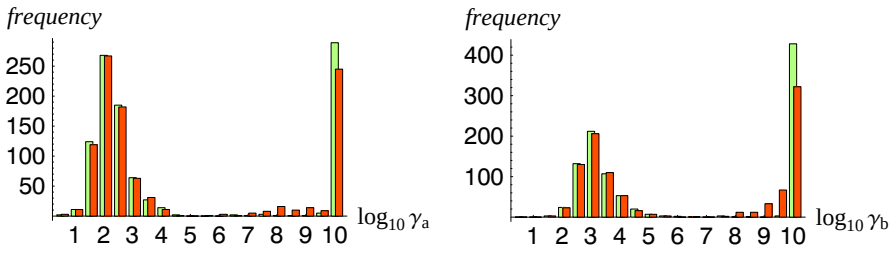
The reason for concentrating on the deviations of the logarithms is that the weights seem to act multiplicatively in the decomposition. Yet the use of logarithms gives rise to the problem that the logarithm of a weight approaches infinity if the the variance of the corresponding coefficient is estimated as being close to zero, *viz.* that the coefficient is estimated as remaining constant over time. This renders the computation of mean squared errors infeasible. In order to avoid artifacts produced in this way, we have computed VC estimates under the restriction $\sigma_u^2/\sigma_i^2 \leq 10^{10}$ and forced the Eviews estimates to conform to this restriction as well. (A weight of 10^{10} enforces, in the setting studied here, constancy of the corresponding parameter anyway. By imposing the restriction, we avoid the problem that differences in weights above this threshold may affect our results.)

Table 1 summarizes our findings. It gives the mean squared errors of the estimated logarithms of the weights for the various parameter constellations, as obtained by the Eviews Kalman filter, and the VC moments estimator, respectively. The performance of the VC estimator is slightly better.

A closer look at the results reconfirms the impression that both estimators perform very similar—with the *caveat* that the Eviews estimates have been calculated by using the theoretical values as starting values. Figure 6 illustrates this for the case by $\sigma_u^2 = 1$, $\sigma_a^2 = 0.01$, $\sigma_b^2 = 0.001$, and $T = 60$. The distributions of the estimates for the weights are practically indistinguishable.

		EV	*VC*	*EV*	*VC*	*EV*	*VC*	*EV*	*VC*
	(γ_a, γ_b)	(1000, 1000)		(100, 1000)		(1000, 100)		(100, 100)	
$T = 40$	$MSE_{10}(\gamma_a)$	29.9	24.6	26.7	24.8	30.5	23.4	27.2	24.2
	$MSE_{10}(\gamma_b)$	25.1	22.6	25.8	23.4	12.7	11.8	13.2	11.9
$T = 60$	$MSE_{10}(\gamma_a)$	29.2	24.3	19.2	18.3	29.5	23.4	20.0	18.4
	$MSE_{10}(\gamma_b)$	21.6	20.3	21.4	20.0	6.9	6.6	7.1	6.9
$T = 80$	$MSE_{10}(\gamma_a)$	26.4	22.3	12.1	11.7	27.1	21.4	13.0	12.2
	$MSE_{10}(\gamma_b)$	16.8	15.5	17.0	15.7	2.9	3.1	3.3	3.3
$T = 100$	$MSE_{10}(\gamma_a)$	23.3	19.9	8.1	7.6	24.0	19.3	9.2	8.6
	$MSE_{10}(\gamma_b)$	13.2	12.3	13.6	12.5	1.3	1.3	1.6	1.5

Table 1: A comparison between the Kalman filter as implemented in Eviews (*EV*) and the VC moments estimator (*VC*), based on 1000 replications for each parameter constellation. The EV estimator is obtained by using the true variances as initial values while the VC estimator does not use initial values; yet the VC estimator outperforms the Kalman filter slightly.



(a) Intercept (b) Slope

Figure 6: Histograms for the estimated log weights for intercept (left) and slope (right). Each pair of columns gives the frequencies of estimates obtained by the Eviews Kalman filter (left column) and the VC moments estimator (right column). The distributions are nearly identical. The estimated weights cluster around their theoretical values ($\log_{10}\gamma_a = 2$ and $\log_{10}\gamma_b = 3$). (The case depicted here is not well conditioned, being characterised by $\sigma_u^2 = 1$, $\sigma_a^2 = 0.01$, $\sigma_b^2 = 0.001$, and $T = 60$, 1000 replications. In roughtly a third of the cases, one or the other weight is estimated exceeding 10^6 which implies that the corresponding parameter is estimated as invariant over time. Yet even here, both methods produce practically identical results.)

12 Concluding comments

The VC estimator discussed here is specifically designed to deal with linear models where the coefficients are following a random walk. The main advantage over the Kalman filter can be seen in its mathematical and descriptive transparency, its ease of use and its free availability. Our simulations suggest that the VC moment estimator and the Eviews Kalman filter perform very similar, provided the starting values of the Eviews filter are provided correctly. For practical purposes this is not feasible, though, and our experience suggests that the performance of the Eviews filter is quite unsatisfactory if proper initial values are not supplied. In our simulations we have used the true parameter values as starting values, though. Yet even under these circumstances, the VC estimator seems to perform slightly better. Our conclusion is that VC filter is preferable for studying the specific class of models for which it is designed. The Kalman filter remains, however, applicable to a much larger class of models. In this sense the method described here does not provide a substitute but rather a complement to existing approaches.

References

AKAIKE, H. 1989, "Bayesian Modeling for Time Series Analysis," in: R. S. MARIANO, K. S. LIM, and K. D. LING (eds.), *Advances in Statistical Analysis and Statistical Computing*, pp. 59–69, Greenwich (Conn.) and London: JAI Press.

ATHANS, M. 1974, "The Importance of Kalman Filtering Methods for Economic Systems," *Annals of Economic and Social Measurement*, 3, 24–49.

COOLEY, T. F. and E. F. PRESCOTT 1973, "An Adaptive Regression Model," *International Economic Review*, 14, 364–71.

EVIEWS 2005, "EViews 5.1 standard edition," online at *www.eviews.com*.

HAMILTON, J. D. 1994, *Time Series Analysis*, Princeton University Press, Princeton.

KEYNES, J. M. 1939, "Professor Tinbergen's Method," *Economic Journal*, 49, 558–68.

— 1973, *The General Theory and After, Part II: Defense and Development*, vol. xiv of *The Collected Works of John Maynard Keynes*, Macmillan, London.

LUCAS, R. 1976, "The Econometric Policy Evaluation: A Critique," in: K. BRUNNER and A. H. MELTZER (eds.), *Phillips Curve and Labor Markets*, pp. 19–46, North Holland, Amsterdam.

LUCAS, R. and T. J. SARGENT 1979, "After Keynesian Macroeconomics," *Quarterly Review of the Federal Reserve Bank of Minneapolis*, 3, online at *http://minneapolisfed.org/research/qr/qr321.pdf*.

LUDSTECK, J. 2004, "VC Package for Mathematica," online at *http://library.wolfram.com/infocenter/MathSource/5195/*.

MARSHALL, A. 1949, *Principles of Economics*, 8th ed., London: Macmillan, online at *http://www.econlib.org/library/Marshall/marPtoc.html*, reprint of 8th edition 1920, first edition 1890.

SCHLICHT, E. 1973, "Forcasting Markov Chains. A Theoretical Foundation for Exponential Smoothing," Discussion paper 13, University of Regensburg, online at *http://www.semverteilung.vwl.uni-muenchen.de/mitarbeiter/es/paper/schlicht-exponential_smoothing.pdf*.

— 1977, *Grundlagen der ökonomischen Analyse*, Rowohlt, Reinbek, online at *http://epub.ub.uni-muenchen.de/archive/00000891/*.

— 1985, *Isolation and Aggregation in Economics*, Springer-Verlag Berlin-Heidelberg-New York, online at *http://epub.ub.uni-muenchen.de/archive/00000003/*.

— 2005a, "VC - A Program for Estimating Time-Varying Coefficients," online at *http://epub.ub.uni-muenchen.de/archive/00000684/*.

— 2005b, "VCC - A Program for Estimating Time-Varying Coefficients. Console Version With Source Code in C," online at *http://epub.ub.uni-muenchen.de/archive/00000719/*.