Errata: pages 5, 15, 20, 21, 25, 27

VC - A METHOD FOR ESTIMATING TIME-VARYING COEFFICIENTS IN LINEAR MODELS

EKKEHART SCHLICHT

ABSTRACT. This papers describes a moments estimator for a standard state-space model with coefficients generated by a random walk. This estimator does not require that disturbances are normally distributed, but if they are, the proposed estimator is asymptotically equivalent to the maximum likelihood estimator.

Keywords: Time-series analysis, linear model, state-space estimation, time-varying coefficients, moments estimation

Journal of Economic Literature Classifications: C2, C22, C32, C51, C52

This paper relies heavily on two earlier discussion papers (Schlicht 1989 and Schlicht and Ludsteck 2006) I thank Daniela Diekmann, Theo Dijkstra. Walter Krämer, José Alberto Bravo Lopez, Johannes Ludsteck, and Ralf Pauly for comments on earlier attempts. Not the least I wish to thank the authors of the works mentioned in Appendix A. Their use of VC eventually convinced me that it may be worthwhile to provide this documentation.

1. Introduction

This paper describes and discusses an estimator for a linear time series model with time-varying coefficients. Such a model, the variable coefficients model, or "VC model" for short, generalizes the standard linear model. The standard model assumes that the coefficients giving the influence of the independent variables on the dependent variable remain constant. In the VC model, these coefficients are permitted to change over time.

The VC model has been initially proposed for dealing empirically with economic theories that are subject to a *ceteris paribus* clause (Schlicht, 1977, ch.4). Schlicht (1989) has proposed an estimation method – the VC method – which has been embodied in some freely available software packages (Schlicht 2005; 2005, Ludsteck 2004; 2018). Some simulations in Schlicht and Ludsteck (2006) have shown that the VC is preferable for studying the specific class of models for which it was designed.

In the meanwhile, VC has found a number of applications in various settings, mainly dealing with structural change, such as the recent decoupling of growth and pollution in the wake of global warming, the changes occurring in financial markets after the financial crisis of 2008, drifts in Okun's Law over time, and more. Appendix A lists some of these studies.

The paper is divided in two parts. In the first part the VC method is described, and in the second part, some points regarding the application of the VC method and some methodological issues are discussed.

Part I

The VC Method

The following sections introduce the model and describe the "criteria" approach that permits to estimate the time-paths of the coefficients in a purely descriptive way. Based on that, a moments estimator will be proposed. If it is assumed additionally that the disturbances are normally distributed, a maximum likelihood estimator can be given. It is shown that this estimator coincides with the moments estimator for sufficiently long time series.

2. THE LINEAR THEORETICAL MODEL AND ITS EMPIRICAL APPLICATION

Consider a theory stating that the dependent variable y as a linear function of some independent variables $x_1, x_2, \ldots, x_n$:

$$y = a_1 x_1 + a_2 x_2 + \ldots + a_n x_n \tag{2.1}$$

The coefficients $a_1, a_2, \ldots, a_n$ give the influence of the independent variables.

If we have T observations y_t, $x_{1,t}$, $x_{2,t}, \ldots x_{n,t}$ with $t = 1, 2, \ldots T$ denoting the time of an observation, we can try to estimate the theoretical coefficients $a_1, a_2, \ldots, a_n$ by a standard linear regression. In order to do that, we have to add an error term u_t to capture discrepancies of the empirical from the theoretical regularity due to measurement errors *etc.* and obtain

$$y_t = a_1 x_{1,t} + a_2 x_{2,t} + \ldots + a_n x_{n,t} + u_t, \;\; t = 1, 2, \ldots T. \tag{2.2}$$

It appears, however, improbable, that outside influences not captured in the theoretical model (and theoretically held constant under a *ceteris paribus* clause) affect only the disturbance term, and not the coefficients themselves. If these outside influences affect the coefficients themselves, the coefficients might change over time.

The problem of possibly time-varying coefficients was the subject of the famous Keynes-Tinbergen controversy around 1940.[1] While Tinbergen (1940, p. 153) defended the use of regression analysis with the argument that in “many cases only small changes in structure will occur in the near future”, Keynes (1973, p. 294) objected 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.”

It appears that both arguments are correct. The VC model takes care of both by assuming that the coefficients change only *slowly* over time: They are highly auto-correlated. This is formalized by a random walk (Cooley and Prescott 1973, Schlicht 1973, Athans 1974). If $a_{i,t}$ denotes the state of

[1] See Tinbergen (1940), Keynes (1939), Keynes (1973, pp. 285–321).

coefficient a_i at time t, it is assumed that

$$a_{i,t+1} = a_{i,t} + v_{i,t} \tag{2.3}$$

with the disturbance term $v_{i,t}$ of expectation zero and with variance σ_i^2. The assumption of expectation zero formalizes the idea that "the coefficients will be fairly constant" in the short run, while the variance σ_i^2 is a measure of the stability of coefficient i and is to be estimated. For $\sigma_i^2 = 0$, the case of a constant (time-invariant) coefficient is covered as well. As a consequence, the standard linear model is replaced by

$$\begin{aligned} y_t &= a_{1,t}x_{1,t} + a_{2,t}x_{2,t} + \ldots + a_{n,t}x_{n,t} + u_t \\ &\quad E\{u_t\} = 0,\ \ E\left\{u_t^2\right\} = \sigma^2 \qquad (2.4) \\ a_{i,t+1} &= a_{i,t} + v_{i,t}, \\ &\quad E\left\{v_{i,t}\right\} = 0,\ \ E\left\{v_{i,t}^2\right\} = \sigma_i^2 \qquad (2.5) \end{aligned}$$

The VC method estimates the expected time-paths of the coefficients. It can be viewed as a straightforward generalization of the method of least squares:

- While the method of ordinary least squares selects estimates that minimize the sum of squared disturbances $\sum_{t=1}^T u_t^2$ in the equation, VC selects estimates that minimize the sum of squared disturbances in the equation and a weighted sum of squared disturbances in the coefficients $\sum_{t=1}^T u_t^2 + \gamma_1 \sum_{t=2}^T v_{1,t}^2 + \gamma_2 \sum_{t=2}^T v_{2,t}^2 + \ldots + \gamma_n \sum_{t=2}^T v_{n,t}^2$, where the weights for the changes in the coefficients $\gamma_1, \gamma_2, \ldots, \gamma_n$ 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 over time.
- Estimation can proceed by focusing on some selected coefficients and keeping the remaining coefficients constant over time. This is done by keeping the corresponding variances σ_i^2 close to zero, rather than estimating them. (If all coefficients are frozen in this manner, the OLS result is obtained.)
- The time-averages of the regression coefficients are GLS estimates of the corresponding regression with fixed coefficients, *i.e.* $\frac{1}{T}\sum_t a_t = a_{GLS}$.

- As all estimates are moments estimates, it is not necessary to presuppose normally distributed disturbances.
- For increasing sample sizes T and under the assumption that all disturbances are normally distributed, the moments estimates approach the maximum likelihood estimates.

3. NOTATION AND BASIC ASSUMPTIONS

All vectors are conceived as column vectors, and their transposes are indicated by an apostrophe. The observations at time t are $x_t' = (x_{1,t}, x_{2,t}, ..., x_{n,t})$ and y_t for $t = 1, 2, .., T$. We write

$$y = \begin{pmatrix} y_1 \\ y_2 \\ . \\ . \\ y_T \end{pmatrix}, \quad x = \begin{pmatrix} x_1' \\ x_2' \\ . \\ . \\ x_T' \end{pmatrix}, \quad X = \begin{pmatrix} x_1' & & & & 0 \\ & x_2' & & & \\ & & . & & \\ & & & . & \\ 0 & & & & x_T' \end{pmatrix}$$

$$\text{order } T \qquad T \times n \qquad T \times Tn$$

$$a_t = \begin{pmatrix} a_{1,t} \\ a_{2,t} \\ . \\ . \\ a_{n,t} \end{pmatrix}, \quad a = \begin{pmatrix} a_1 \\ a_2 \\ . \\ . \\ a_T \end{pmatrix}, \quad v_t = \begin{pmatrix} v_{1,t} \\ v_{2,t} \\ . \\ . \\ v_{n,t} \end{pmatrix}, \quad v = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ . \\ . \\ v_{T-1} \end{pmatrix}$$

$$\text{order } n \qquad Tn \qquad T-1 \qquad (T-1)\,n$$

We write further

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

$$\text{order} \quad n \qquad\qquad\qquad n \times n$$

and define

$$p = \begin{pmatrix} -1 & 1 & 0 & & & 0 \\ 0 & -1 & 1 & 0 & & \\ & & . & . & & \\ & & & . & . & 0 \\ 0 & & & 0 & -1 & 1 \end{pmatrix}, \quad P = p \otimes I_n = \begin{pmatrix} -I_n & I_n & & & & 0 \\ & -I_n & I_n & & & \\ & & . & . & & \\ & & & . & . & \\ 0 & & & & -I_n & I_n \end{pmatrix}$$

$$\text{order} \quad (T-1) \times T \qquad\qquad\qquad (T-1)\,n \times T n$$

with I_n denoting the identity matrix of order n.

The model is obtained by writing equations (2.4) and (2.5) in matrix form:

The model

$$y = Xa + u, \quad E\{u\} = 0, \quad E\{uu'\} = \sigma^2 I_T \tag{3.1}$$

$$Pa = v, \quad E\{v\} = 0, \quad E\{vv'\} = V = I_{T-1} \otimes \Sigma \tag{3.2}$$

Note that the explanatory variables X are taken as predetermined, rather than stochastic.

Regarding the observations X and y we assume that a perfect fit of the model to the data is not possible:

Assumption ("No Perfect Fit").

$$Pa = 0 \quad \text{implies} \quad y \neq Xa. \tag{3.3}$$

This assumption rules out the (trivial) case that the standard linear model (2.2) fits the empirical data perfectly, a case that cannot reasonably be expected to occur in practical applications. Further, the assumption implies that the number of observations exceeds the number of coefficients to be estimated:

$$T > n. \tag{3.4}$$

4. LEAST SQUARES

In a descriptive spirit, the time-paths of the coefficients can be determined by following the "criteria" approach, where some criteria are employed that formalize some descriptive desiderata.[2] In the case at hand, the desiderata are that the model fits the data well and that the coefficients change only slowly over time – u and v ought to be as small as possible. The sum of the squared errors $u'u$ is taken as a criterion for the goodness of fit of equation (3.1), the weighted sum of the squared changes of the coefficients $v'v$ over time is the criterion for the stability of the coefficients over time, and the combination of both criteria gives an overall criterion that combines the desiderata of a good fit and stability of coefficients over time. The weights $(\gamma_1, \gamma_2, .., \gamma_n)$ give the relative importance of the stability of the coefficients over time, where weight γ_i relates to coefficient a_i. For the time being, these weights are taken as given but will later be estimated, too.

Write

$$\Gamma = \begin{pmatrix} \gamma_1 & 0 & . & & 0 \\ 0 & \gamma_2 & 0 & . & \\ . & 0 & . & . & \\ & & . & . & 0 \\ 0 & & & 0 & \gamma_n \end{pmatrix} \tag{4.1}$$

and

$$G = I_{T-1} \otimes \Gamma. \tag{4.2}$$

[2]The criteria approach was introduced by Leser (1961), used also by Hodrick and Prescott (1997), and has been further developed by Leser (1963), Schlicht (1981), and Schlicht and Pauly (1983).

Adding the sum of squares $u'u$ and the weighted sum of squares $v'Gv$ gives the overall criterion

$$Q = u'u + v'Gv \tag{4.3}$$

This expression is to be minimized under the constraints given by the model (3.1), (3.2) with the observations X and y

$$u = y - Xa \tag{4.4}$$

$$v = Pa. \tag{4.5}$$

This determines the time-paths of the coefficients a that optimize this criterion. Hence we can write

$$Q = \left(y - Xa\right)' \left(y - Xa\right) + a'P'GPa \tag{4.6}$$

The weighted sum of squares Q is the sum of two positive semi-definite quadratic forms. The “no perfect fit” assumption (3.3) rules out the case that Q can be zero. Hence Q is positive definite and of full rank. The first order condition for a minimizing a is

$$\frac{\partial Q}{\partial a} = -2Xy + 2\left(X'X + P'GP\right)a = 0 \tag{4.7}$$

and the second order condition is that the Jacobian

$$\frac{\partial^2 Q}{\partial a \partial a'} = 2\left(X'X + P'GP\right) \tag{4.8}$$

be positive definite, which is the case. Solving (4.7) for a and plugging this into (4.4) and (4.5) gives the estimates

$$a_{LS} = \left(X'X + P'GP\right)^{-1} X'y \tag{4.9}$$

$$u_{LS} = \left(I_T - X\left(X'X + P'GP\right)^{-1} X'\right) y \tag{4.10}$$

$$v_{LS} = P\left(X'X + P'GP\right)^{-1} X'y \tag{4.11}$$

where the subscript LS stands for “least squares”.

5. ORTHOGONAL PARAMETRIZATION

For purposes of estimation we need a model that explains the observation y as a function of the observations X and the random variables u and v.

This would permit calculating the probability distribution of the observations y 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 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

$$A := \left\{ a = \tilde{P}v + Z\beta \,\middle|\, \beta \in \mathbb{R}^n \right\}. \tag{5.1}$$

with β as a shift parameter and

$$\tilde{P} \quad := \quad P'\left(PP'\right)^{-1} \tag{5.2}$$

of order $Tn \times (T-1)\,n$ as the right-hand pseudo-inverse of P. For any v we have $a \in A \Leftrightarrow Pa = v$. Hence equation (3.1) and the set (5.1) give equivalent descriptions of the relationship between a and v.

Define further the $Tn \times n$ matrix

$$Z := \begin{pmatrix} I_n \\ I_n \\ \cdot \\ I_n \end{pmatrix}. \tag{5.3}$$

It is orthogonal to P:

$$PZ \quad = \quad 0$$

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

$$Z'Z = T \cdot I_n,\ P'\left(PP'\right)^{-1} P + ZZ' = I_{Tn}. \tag{5.4}$$

The last equality is implied by the identity

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

Regarding the matrices P, $\tilde{P}$, and Z we have

$$\begin{aligned} P\tilde{P} &= \tilde{P}'P' = I_{(T-1)n} \\ \tilde{P}P &= P'\tilde{P}' = I_{Tn} - ZZ' \\ Z'\tilde{P} &= \tilde{P}'Z = 0. \end{aligned} \tag{5.5}$$

In view of (5.1), any solution a to $Pa = v$ can be written as

$$a = \tilde{P}v + Z\beta \tag{5.6}$$

for some $\beta \in \mathbb{R}^n$. Equation (3.1) can be re-written as

$$y = u + X\tilde{P}v + XZ\beta. \tag{5.7}$$

The model (5.6), (5.7) will be referred to as the *equivalent orthogonally parameterized model.* It implies the *true model* (3.1), (3.2). It implies, in particular, that a_t is a random walk even though a_t depends, according to (5.6), on past *and* future realizations of v_t.

The formal parameter β has a straightforward interpretation. Pre-multiplying (5.6) by Z' gives

$$Z'a \quad = \quad Z'Z\beta = T\beta$$

and therefore

$$\beta \quad = \quad \frac{1}{T}\sum_{t=1}^{T} a_t.$$

Hence β gives the averages of the coefficients $a_{i,t}$ over time.

Equation (5.7) permits calculating the density of y dependent upon the parameters of the distributions of u and v and the formal parameters β. In a second step, all these parameters – σ^2, Σ, and β – can be determined by moments estimators that will be derived in Section 8.

The orthogonal parametrization, proposed in Schlicht (1985, Sec. 4.3.3), entails some advantages with respect to symmetry and mathematical transparency, as compared to more usual procedures, such as parametrization by initial values. It permits to derive our moments estimator that does not require normally distributed disturbances, and to write down an explicit likelihood function for the case of normally distributed disturbances that permits estimation of all relevant parameters in a unified one-shot procedure.

The formal parameter vector β relates directly to the coefficient estimates of a standard generalized least squares (GLS, Aitken) regression. Equation (5.7) can be interpreted as a standard regression for this parameter vector with the matrix $x = XZ$ giving the explanatory variables:

$$y = x\beta + w \tag{5.8}$$

and the disturbance

$$w \quad = \quad X\tilde{P}v + u. \tag{5.9}$$

It has expectation zero

$$E\{w\} \quad = \quad 0 \tag{5.10}$$

and covariance

$$E\{ww'\} = X\tilde{P}V\tilde{P}'X' + \sigma^2 I_T = W. \tag{5.11}$$

The Aitken estimate β_A satisfies

$$x'W^{-1}\left(y - x\beta_A\right) = 0 \tag{5.12}$$

or

$$\beta_A = \left(x'W^{-1}x\right)^{-1} x'W^{-1}y. \tag{5.13}$$

where the subscript A stands for "Aitken".

6. THE FILTER

This section derives the VC filter which gives the expectation of the coefficients a for given observations X and y, a given shift parameter β, and given variances σ^2 and Σ.

For given β and X, the vectors y and a can be viewed as realizations of random variables determined jointly by the system (5.6), (5.8) as brought about by the disturbances u and v:

$$\begin{pmatrix} a \\ y \end{pmatrix} = \begin{pmatrix} Z \\ XZ \end{pmatrix}\beta + \begin{pmatrix} \tilde{P} & 0 \\ X\tilde{P} & I_T \end{pmatrix}\begin{pmatrix} v \\ u \end{pmatrix}$$

The covariance is

$$\begin{aligned} E\left\{\begin{pmatrix} a \\ y \end{pmatrix}\begin{pmatrix} a & y \end{pmatrix}\right\} &= \begin{pmatrix} \tilde{P} & 0 \\ X\tilde{P} & I_T \end{pmatrix}\begin{pmatrix} V & 0 \\ 0 & \sigma^2 I_T \end{pmatrix}\begin{pmatrix} \tilde{P}' & \tilde{P}'X' \\ 0 & I_T \end{pmatrix} \\ &= \begin{pmatrix} \tilde{P}V\tilde{P}' & \tilde{P}V\tilde{P}'X' \\ X\tilde{P}V\tilde{P}' & X\tilde{P}V\tilde{P}'X + \sigma^2 I_T \end{pmatrix}. \end{aligned}$$

The marginal distribution of y is as given by (5.8) and (5.11). The conditional expectation of a and the expected covariance for given y (and β, σ^2,

Σ) are

$$E\{a|y\} = Z\beta + \tilde{P}V\tilde{P}'X'\left(X\tilde{P}V\tilde{P}'X + \sigma^2 I_T\right)^{-1}\left(y - XZ\beta\right) \quad (6.1)$$

$$E\{aa'|y\} = \tilde{P}V\tilde{P}' - \tilde{P}V\tilde{P}'X'\left(X\tilde{P}V\tilde{P}'X + \sigma^2 I_T\right)^{-1} X\tilde{P}V\tilde{P}' \quad (6.2)$$

Hence an estimator for a can be derived by plugging the Aitken estimator β_A from (5.12) into (6.1) and calculating the mean:

$$a_A = Z\beta_A + \frac{1}{\sigma^2}\tilde{P}V\tilde{P}'X'\left(y - XZ\beta_A\right). \quad (6.3)$$

Note that the variance-covariance matrix of w, as given in equation (5.11), tends to $\sigma^2 I_T$ if the the variances σ_i^2 go to zero, and equation (5.7) approaches the standard unweighted linear regression. In this sense, the OLS regression model is covered as a special limiting case by the model discussed here.

7. LEAST SQUARES AND AITKEN

The following theorem states that the least squares estimator a_{LS} and the Aitken estimator a_A coincide if the weights are given by the variance ratios.

Claim 1. $G = \sigma^2 V^{-1}$ implies $a_{LS} = a_A$.

Proof. Consider first the necessary conditions for a minimum of (4.3). The first-order condition (4.7) defines a_{LS} with weights $G = \sigma^2 V^{-1}$ uniquely and can be written as

$$\left(X'X + \sigma^2 P'V^{-1}P\right)a_{LS} = X'y \quad (7.1)$$

It will be shown that (6.3) implies

$$\left(X'X + \sigma^2 P'V^{-1}P\right)a_A = X'y \quad (7.2)$$

which will establish the proposition.

Pre-multiplication of (6.3) by $\left(X'X + \sigma^2 P'V^{-1}P\right)$ gives

$$\begin{aligned}\left(X'X + \sigma^2 P'V^{-1}P\right)a_A &= \left(X'X + \sigma^2 P'V^{-1}P\right)Z\beta_A + \\ &+ \left(X'X + \sigma^2 P'V^{-1}P\right)\tilde{P}V\tilde{P}'X'\left(X\tilde{P}V\tilde{P}'X + \sigma^2 I_T\right)^{-1} \cdot \\ &\cdot \left(y - XZ\beta_A\right).\end{aligned}$$

Because of $PZ = 0$ this can be written as

$$\begin{aligned}\left(X'X+\sigma^2 P'V^{-1}P\right)a_A &= X'XZ\beta_A+\\ &+X'X\tilde{P}V\tilde{P}'X'\left(X\tilde{P}V\tilde{P}'X+\sigma^2 I_T\right)^{-1}\left(y-XZ\beta_A\right)\\ &+\sigma^2 P'\tilde{P}'X'\left(X\tilde{P}V\tilde{P}'X+\sigma^2 I_T\right)^{-1}\left(y-XZ\beta_A\right).\end{aligned}$$

Adding and subtracting $\sigma^2 X'\left(X\tilde{P}V\tilde{P}'X+\sigma^2 I_T\right)^{-1}\left(y-XZ\beta_A\right)$ and using $P'\tilde{P}'=\left(I_{Tn}-ZZ'\right)$ results in

$$\begin{aligned}\left(X'X+\sigma^2 P'V^{-1}P\right)a_A &= X'XZ\beta_A+\\ &+X'\left(X\tilde{P}V\tilde{P}'X'+\sigma^2 I_T\right)\left(X\tilde{P}V\tilde{P}'X+\sigma^2 I_T\right)^{-1}\left(y-XZ\beta_A\right)\\ &-\sigma^2 X'\left(X\tilde{P}V\tilde{P}'X+\sigma^2 I_T\right)^{-1}\left(y-XZ\beta_A\right)\\ &+\sigma^2\left(I_{Tn}-ZZ'\right)X'\left(X\tilde{P}V\tilde{P}'X+\sigma^2 I_T\right)^{-1}\left(y-XZ\beta_A\right)\end{aligned}$$

which reduces to

$$\begin{aligned}\left(X'X+\sigma^2 P'V^{-1}P\right)a_A &= X'XZ\beta_A+\\ &+X'\left(y-XZ\beta_A\right)\\ &-\sigma^2 X'\left(X\tilde{P}V\tilde{P}'X+\sigma^2 I_T\right)^{-1}\left(y-XZ\beta_A\right)\\ &+\sigma^2 X'\left(X\tilde{P}V\tilde{P}'X+\sigma^2 I_T\right)^{-1}\left(y-XZ\beta_A\right),\end{aligned}$$

hence to

$$\left(X'X+\sigma^2 P'V^{-1}P\right)a_A = X'XZ\beta_A+X'\left(y-XZ\beta_A\right)$$

and finally to

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

This shows that the least squares estimator a_{LS} and the Aitken estimator a_A coincide. □

For the sake of completeness and later use, the following observation is added:

Claim 2. $G=\sigma^2 V^{-1}$ implies $Q=\sigma^2 w'W^{-1}w$. In other words: the sum of squared deviations weighted by the variance ratios $\frac{\sigma^2}{\sigma_1^2},\frac{\sigma^2}{\sigma_2^2},\ldots,\frac{\sigma^2}{\sigma_n^2}$ equals the weighted sum of squares (the squared Mahalanobis distance) in the Aitken regression.

Proof.

$$\begin{aligned}
Q &= u'u + \sigma^2 v'V^{-1}v \\
&= u'\left(w - X\tilde{P}v\right) + \sigma^2 v'V^{-1}v \\
&= u'w - \left(\left(w - X\tilde{P}v\right)' X\tilde{P} - \sigma^2 v'V^{-1}\right) v \\
&= u'w - \left(\left(w - X\tilde{P}v\right)' X\tilde{P} - \sigma^2 v'V^{-1}\right) Pa \\
&= u'w - \left(\left(w - X\tilde{P}v\right)' X\tilde{P} - \sigma^2 v'V^{-1}\right) P\left(Z\beta + \tilde{P}V\tilde{P}'X'W^{-1}w\right) \\
&= u'w - \left(\left(w - X\tilde{P}v\right)' X\tilde{P} - \sigma^2 v'V^{-1}\right) P\tilde{P}V\tilde{P}'X'W^{-1}w \\
&= u'w - \left(\left(w - X\tilde{P}v\right)' X\tilde{P} - \sigma^2 v'V^{-1}\right) V\tilde{P}'X'W^{-1}w \\
&= u'w - \left(u'X\tilde{P}V\tilde{P}'X' - \sigma^2 v'\tilde{P}'X'\right) W^{-1}w \\
&= u'w - u'X\tilde{P}V\tilde{P}'X'W^{-1}w + \sigma^2 v'\tilde{P}'X'W^{-1}w \\
&= u'w - u'\left(X\tilde{P}V\tilde{P}'X' + \sigma^2 I_T - \sigma^2 I_T\right) W^{-1}w + \sigma^2 v'\tilde{P}'X'W^{-1}w \\
&= u'w - u'\left(X\tilde{P}V\tilde{P}'X' + \sigma^2 I_T\right) W^{-1}w + \sigma^2 u'W^{-1}w + \sigma^2 v'\tilde{P}'X'W^{-1}w \\
&= u'w - u'w + \sigma^2 u'W^{-1}w + \sigma^2 v'\tilde{P}'X'W^{-1}w \\
&= \sigma^2\left(u' + v'\tilde{P}'X'\right) W^{-1}w \\
&= \sigma^2 w'W^{-1}w
\end{aligned}$$

Hence the weighted sum of squares Q equals the squared Mahalanobis distance. □

Consider now the distribution of a_A. The matrix $\left(X'X + \sigma^2 P'V^{-1}P\right)$, henceforth referred to as the "system matrix", will be denoted by M:

$$M = \left(X'X + \sigma^2 P'V^{-1}P\right). \tag{7.3}$$

With this, the normal equation (7.2), which defines the solution for the vector of the coefficients a_A can be written as

$$M a_A = X'y. \tag{7.4}$$

With (3.1) and (7.3) we obtain

$$\begin{aligned}
a_A &= M^{-1}X'(Xa + u) \\
&= M^{-1}\left(X'Xa + X'u + \sigma^2 P'V^{-1}Pa - \sigma^2 P'V^{-1}Pa\right) \\
&= a + M^{-1}\left(X'u - \sigma^2 P'V^{-1}v\right). \qquad (7.5)
\end{aligned}$$

Given a realization of the time-path of the coefficients a, the estimator a_A is distributed with mean a and covariance

$$E\left\{(a-a_A)'(a-a_A)\right\} = M^{-1}\begin{pmatrix} X' & -\sigma^2 P'V^{-1} \end{pmatrix}\begin{pmatrix} \sigma^2 I_T & 0 \\ 0 & V \end{pmatrix}\begin{pmatrix} X \\ -\sigma^2 V^{-1}P \end{pmatrix}M^{-1}$$

which reduces to

$$E\left\{(a-a_A)'(a-a_A)\right\} = M^{-1}\begin{pmatrix} \sigma^2 X'X & +\sigma^4 P'V^{-1}P \end{pmatrix}M^{-1}$$

and finally to

$$E\left\{(a-a_A)'(a-a_A)\right\} = \sigma^2 M^{-1}. \tag{7.6}$$

The system matrix (7.3) is determined by the observations X, the variance σ^2 and the variances Σ. Equation (7.6) gives the precision of our estimate which is directly related to the system matrix M. The next step is to determine the variance σ^2 and the variances Σ.

8. MOMENTS ESTIMATION OF THE VARIANCES

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 the variances of the disturbances in the estimated coefficients equal their expectations. It has, thus, a straightforward connotation even in shorter time series and does not presuppose that the perturbations u and v are normally distributed. It will be shown later that the moments estimators approach the respective maximum likelihood estimators in large samples if the disturbances are normally distributed. Hence 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.

In the following we denote the estimated coefficients by $\hat{a}$ and the estimated perturbations by $\hat{u}$ and $\hat{v}$. For some variances σ^2 and $\sum = \text{diag}\begin{pmatrix} \sigma_1^2, & \sigma_2^2, & . & . & . & , & \sigma_n^2 \end{pmatrix}$, 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 (7.5) we obtain

$$\begin{aligned}\hat{u} &= -X\left(M^{-1}\left(X'u-\sigma^2 P'V^{-1}v\right)\right)+u\\ \hat{u} &= \left(I_T - XM^{-1}X'\right)u+\sigma^2 XM^{-1}P'V^{-1}v.\end{aligned}$$

Regarding $\hat{v}$, consider the vectors $\hat{v}_i' = \begin{pmatrix} \hat{v}_{i,2}^2, & \hat{v}_{i,3}^2, & . & . & , & \hat{v}_{i,T}^2 \end{pmatrix}$ for $i = 1,2,..,n$, that is, the disturbances in the coefficients $\hat{a}_i$ for each coefficient separately. These are obtained as follows.

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

$$E_i := I_{T-1} \otimes e_i' \tag{8.1}$$

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:

$$v_i := E_i v.$$

Note that

$$\sum_{i=1}^{n} \sigma_i^2 E_i' E_i \quad = \quad V. \tag{8.2}$$

Pre-multiplying (7.5) with the matrices E_i yields

$$\hat{v}_i = E_i\left(I_{(T-1)n} - \sigma^2 PM^{-1}P'V^{-1}\right)v + E_i PM^{-1}X'u$$

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

Claim 3. For given observations X and y and given variances $\sigma^2$and Σ, the expected squared deviations of $\hat{u}$ and $\hat{v}_i$, $i = 1,2,\dots,n$ are

$$E\left\{\hat{u}'\hat{u}\right\} \quad = \quad \sigma^2\left(T - \mathrm{tr}XM^{-1}X'\right) \tag{8.3}$$

$$E\left\{\hat{v}_i'\hat{v}_i\right\} \quad = \quad (T-1)\,\sigma_i^2 - \sigma^2 \mathrm{tr}E_i PM^{-1}P'E_i'. \tag{8.4}$$

This implies that the expected sum of squares is

$$E\left\{\hat{Q}\right\} \quad = \quad \sigma^2\,(T-n)\,. \tag{8.5}$$

Proof. The expectation of the squared estimated error $\hat{u}$ is

$$
\begin{aligned}
E\left\{\hat{u}'\hat{u}\right\} &= E\left\{\left(u'\left(I_T + XM^{-1}X'\right) + \sigma^2 v'V^{-1}PM^{-1}X'\right)\cdot\right.\\
&\quad \left.\left(\left(I_T - XM^{-1}X'\right)u + \sigma^2 XM^{-1}P'V^{-1}v\right)\right\}\\
&= E\left\{u'\left(I_T - XM^{-1}X'\right)\left(I_T - XM^{-1}X'\right)u\right\} +\\
&\quad +\sigma^4 E\left\{v'V^{-1}PM^{-1}X'XM^{-1}P'V^{-1}v\right\}\\
&= \mathrm{tr}E\left\{u'\left(I_T - XM^{-1}X'\right)\left(I_T - XM^{-1}X'\right)u\right\} +\\
&\quad +\sigma^4 \mathrm{tr}E\left\{v'V^{-1}PM^{-1}X'XM^{-1}P'V^{-1}v\right\} S\\
&= \mathrm{tr}E\left\{\left(I_T - XM^{-1}X'\right)uu'\left(I_T - XM^{-1}X'\right)\right\} +\\
&\quad +\sigma^4 \mathrm{tr}E\left\{XM^{-1}P'V^{-1}vv'V^{-1}PM^{-1}X'\right\}\\
&= \mathrm{tr}\sigma^2\left(I_T - XM^{-1}X'\right)\left(I_T - XM^{-1}X'\right) + \mathrm{tr}\sigma^4 XM^{-1}P'V^{-1}PM^{-1}X'\\
&= \sigma^2\mathrm{tr}\left(\left(I_T - XM^{-1}X'\right)\left(I_T - XM^{-1}X'\right) + \sigma^2 XM^{-1}P'V^{-1}PM^{-1}X'\right)\\
&= \sigma^2\mathrm{tr}\left(I - 2XM^{-1}X' + XM^{-1}X'XM^{-1}X' + \sigma^2 XM^{-1}P'V^{-1}PM^{-1}X'\right)\\
&= \sigma^2\mathrm{tr}\left(I_T - 2XM^{-1}X' + XM^{-1}\left(X'X + \sigma^2 P'V^{-1}P\right)M^{-1}X'\right)\\
&= \sigma^2\mathrm{tr}\left(I_T - XM^{-1}X'\right)\\
&= \sigma^2\left(T - \mathrm{tr}XM^{-1}X'\right).
\end{aligned}
$$

In a similar way, the expectation of the squared estimated disturbance in the i-th coefficient $\hat{v}_i$ is evaluated as

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

$$
\begin{aligned}
E\left\{\hat{v}_i'\hat{v}_i\right\} &= \operatorname{tr}\left(\sigma^2 E_i P M^{-1} X' X M^{-1} P' E_i'\right) + \\
&\quad \operatorname{tr}\left(E_i\left(V - \sigma^2 P M^{-1} P'\right)\left(I_{(T-1)n} - \sigma^2 V^{-1} P M^{-1} P'\right) E_i'\right) \\
&= \operatorname{tr}\left(\sigma^2 E_i P M^{-1} X' X M^{-1} P' E_i'\right) + \\
&\quad \operatorname{tr}\left(E_i\left(V - \sigma^2 P M^{-1} P'\right)\left(I_{(T-1)n} - \sigma^2 V^{-1} P M^{-1} P'\right) E_i'\right) \\
&= \operatorname{tr}\left(\sigma^2 E_i P M^{-1} X' X M^{-1} P' E_i'\right) + \\
&\quad \operatorname{tr}\left(E_i\left(V - \sigma^2 P M^{-1} P'\right) E_i' - \sigma^2 E_i\left(V - \sigma^2 P M^{-1} P'\right) V^{-1} P M^{-1} P' E_i'\right) \\
&= \operatorname{tr}\left(\sigma^2 E_i P M^{-1} X' X M^{-1} P' E_i'\right) + \\
&\quad \operatorname{tr}\left(E_i\left(V - \sigma^2 P M^{-1} P' - \sigma^2 P M^{-1} P' + \sigma^4 P M^{-1} P' V^{-1} P M^{-1} P'\right) E_i'\right) \\
&= \operatorname{tr}\left(\sigma^2 E_i P M^{-1} X' X M^{-1} P' E_i' + \right. \\
&\quad \left. E_i\left(V - \sigma^2 P M^{-1} P' - \sigma^2 P M^{-1} P' + \sigma^4 P M^{-1} P' V^{-1} P M^{-1} P'\right) E_i'\right) \\
&= \operatorname{tr}\left(E_i\left(\left(\sigma^2 P M^{-1}\left(X' X + \sigma^2 P' V^{-1} P\right) M^{-1} P'\right) + V - 2\sigma^2 P M^{-1} P'\right) E_i'\right) \\
&= \operatorname{tr}\left(E_i\left(V - \sigma^2 P M^{-1} P'\right) E_i'\right) \\
&= \operatorname{tr}\left(E_i V E_i' - \sigma^2 E_i P M^{-1} P' E_i'\right) \\
&= \operatorname{tr}\left(\left(I_{T-1} \otimes e_i'\right)\left(I_{T-1} \otimes \Sigma\right)\left(I_{T-1} \otimes e_i\right) - \sigma^2 E_i P M^{-1} P' E_i'\right) \\
&= \operatorname{tr}\left(I_{T-1} \otimes e_i' \Sigma e_i\right) - \sigma^2 \operatorname{tr}\left(E_i P M^{-1} P' E_i'\right) \\
&= (T-1)\,\sigma_i^2 - \sigma^2 \operatorname{tr}\left(E_i P M^{-1} P' E_i'\right).
\end{aligned}
$$

Regarding $\hat{Q}$ we note that

$$
X'X + \sigma^2 P' V^{-1} P \quad = \quad X'X + \sigma^2 \sum_{i=1}^{n} \frac{1}{\sigma_i^2} P_i' P_i = M
$$

and obtain

$$\begin{aligned}
E\{\hat{Q}\} &= \sigma^2\left(T - \mathrm{tr}XM^{-1}X'\right) + \sum_{i=1}^{n}\frac{\sigma^2}{\sigma_i^2}\left((T-1)\,\sigma_i^2 - \sigma^2\mathrm{tr}E_iPM^{-1}P'E_i'\right) \\
&= \sigma^2\left(T - \mathrm{tr}XM^{-1}X' + \sum_{i=1}^{n}(T-1) - \sum_{i=1}^{n}\frac{\sigma^2}{\sigma_i^2}\mathrm{tr}E_iPM^{-1}P'E_i'\right) \\
&= \sigma^2\left(T + n(T-1) - \mathrm{tr}XM^{-1}X' - \mathrm{tr}\left(\sum_{i=1}^{n}\frac{\sigma^2}{\sigma_i^2}E_iPM^{-1}P'E_i'\right)\right) \\
&= \sigma^2\left(Tn - T - n - \mathrm{tr}M^{-1}X'X - \mathrm{tr}\left(M^{-1}\sum_{i=1}^{n}\frac{\sigma^2}{\sigma_i^2}P'E_i'E_iP\right)\right) \\
&= \sigma^2\left(Tn - T - n - \mathrm{tr}M^{-1}X'X - \mathrm{tr}\left(M^{-1}\sum_{i=1}^{n}\sigma^2P'V^{-1}P\right)\right) \\
&= \sigma^2\left(Tn - T - n - \mathrm{tr}M^{-1}\left(X'X - \sigma^2P'V^{-1}P\right)\right) \\
&= \sigma^2\left(Tn - T - n - \mathrm{tr}I_{nT}\right) \\
&= \sigma^2(T-n).
\end{aligned}$$

□

The moments estimators are obtained by selecting variances σ^2 and $\sigma_i^2, i = 1,2,...,n$ such that the expected moments $E\{\hat{u}'\hat{u}\}$ and $E\{\hat{v}_i'\hat{v}_i\}, i = 1,2,...,n$ 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 $\hat{\sigma}^2$ and $\hat{\sigma}_i^2$, $i = 1,2,...,n$, respectively, are defined as a fix point of the system

$$\begin{aligned}
E\{\hat{u}'\hat{u}\} &= \hat{u}'\hat{u} \\
E\{\hat{v}_i'\hat{v}_i\} &= \hat{v}_i'\hat{v}_i
\end{aligned}$$

Alternatively, the moments estimators can be equivalently defined as a fix point of the system:

$$\begin{aligned}
E\{\hat{v}_i'\hat{v}_i\} &= \hat{v}_i'\hat{v}_i \\
E\{\hat{Q}\} &= \hat{Q}.
\end{aligned}$$

The implementations Schlicht (2005a, 2005b) use the latter alternative and proceed as follows. The equation system to be solved is

$$\begin{aligned} \hat{v}_i'\hat{v}_i &= (T-1)\,\hat{\sigma}_i^2 - \hat{\sigma}^2 \mathrm{tr} E_i P \hat{M}^{-1} P' E_i' \\ \frac{1}{T-n}\hat{Q} &= \hat{\sigma}^2. \end{aligned}$$

which can be written as

$$\frac{\hat{\sigma}_i^2}{\hat{\sigma}^2} = \left(\frac{\hat{v}_i'\hat{v}_i}{\hat{Q}}(T-n) - \mathrm{tr} E_i P \hat{M}^{-1} P' E_i'\right)\frac{1}{T-1} \tag{8.6}$$

$$\hat{\sigma}^2 = \frac{1}{T-n}\hat{Q}. \tag{8.7}$$

Iteration starts with some variance ratios $\gamma_i = \frac{\sigma^2}{\sigma_i^2}$. This permits to determine the right-hand sides of equations (8.6) and (8.7). The variance ratios at the left-hand side of (8.6) and the variance at the left hand side of (8.7) are used for a new iteration, and this continues until convergence is reached, delivering the fix-point values $\hat{\gamma}_i = \frac{\hat{\sigma}^2}{\hat{\sigma}_i{}^2}$ and $\hat{\sigma}^2$ and the corresponding variances $\hat{\sigma}_i^2 = \frac{\hat{\sigma}^2}{\hat{\gamma}_i}$.

9. MAXIMUM LIKELIHOOD ESTIMATION OF THE VARIANCES

This section derives a maximum-likelihood estimator for the variances under the additional assumption that the disturbances u and v are normally distributed.

Using equations (3.2) and (5.9) – (5.13) together with the identity $x = XZ$, the concentrated log-likelihood function for the Aitken regression (5.8) can be written as

$$\mathscr{L}\left(\ \sigma^2, \Sigma\ \right) = -\frac{1}{2}\left(T\left(\log 2 + \log \pi\right) + \log\det W\right) - \frac{1}{2}\hat{w}' W^{-1}\hat{w} \tag{9.1}$$

with

$$W = X\tilde{P}\left(I_{T-1}\otimes\Sigma\right)\tilde{P}'X' + \sigma^2 I_T$$

and

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

By maximizing (9.1) with respect to σ^2 and Σ, the maximum likelihood estimates for the variances are obtained and the corresponding expectation for the parameter a is given by (6.3) and its covariance matrix is given by (7.6).

The maximum likelihood estimator can be characterized in another way. This will be explained in the following. In order to do so, the following lemma is needed.

Claim 4.

$$\begin{aligned}\log\det W &= \log\det\left(PMP'\right) + (T-1)\sum_{i=1}^{n}\log\sigma_i^2 - \\ &\quad ((T-1)\,n - T)\log\sigma^2 - 2\log\det\left(PP'\right). \end{aligned} \tag{9.2}$$

Proof. Note that $V = (I_{T-1}\otimes\Sigma)$ and that $\tilde{P} = P'\left(PP'\right)^{-1}$ and write

$$\begin{aligned}\det W &= \det\left(X\tilde{P}V\tilde{P}'X' + \sigma^2 I_T\right) \\ &= \left(\sigma^2\right)^T \det\left(\frac{1}{\sigma^2}X\tilde{P}V^{\frac{1}{2}}V^{\frac{1}{2}}\tilde{P}'X' + I_T\right) \\ &= \left(\sigma^2\right)^T \det\left(\frac{1}{\sigma^2}V^{\frac{1}{2}}\tilde{P}'X'X\tilde{P}V^{\frac{1}{2}} + I_{(T-1)n}\right) \\ &= \left(\sigma^2\right)^T \det\left(V^{\frac{1}{2}}\left(\frac{1}{\sigma^2}\tilde{P}'X'X\tilde{P} + V^{-1}\right)V^{\frac{1}{2}}\right) \\ &= \left(\sigma^2\right)^T \det\left(V\left(\frac{1}{\sigma^2}\left(PP'\right)^{-1}PX'XP'\left(PP'\right)^{-1} + V^{-1}\right)\right) \\ &= \left(\sigma^2\right)^T \det\left(\frac{1}{\sigma^2}V\left(PP'\right)^{-1}P\left(X'X + \sigma^2P'V^{-1}P\right)P'\left(PP'\right)^{-1}\right) \\ &= \left(\sigma^2\right)^T \det\left(\frac{1}{\sigma^2}V\right)\det\left(PP'\right)^{-1}\det\left(PMP'\right)\det\left(PP'\right)^{-1} \\ &= \left(\sigma^2\right)^T \left(\prod_{i=1}^{n}\frac{\sigma_i^2}{\sigma^2}\right)^{(T-1)}\det\left(PP'\right)^{-2}\det\left(PMP'\right).\end{aligned}$$

Hence the result

$$\begin{aligned}\log\det W &= \log\det\left(PMP'\right) + (T-1)\sum_{i=1}^{n}\log\sigma_i^2 - \\ &\quad (T-1)n - T)\log\sigma^2 - 2\log\det\left(PP'\right)\end{aligned}$$

is obtained. □

Claim 5. Minimizing the criterion

$$\mathscr{C}_L = \log\det\left(PMP'\right) + (T-1)\sum_{i=1}^{n}\log\sigma_i^2 - (T-1)n - T)\log\sigma^2 + \\ + \frac{1}{\sigma^2}\hat{u}'\hat{u} + \hat{v}'V^{-1}\hat{v} \qquad (9.3)$$

is equivalent to maximizing the likelihood function (9.1).

Proof. With (9.2) we have

$$\mathscr{C}_L + 2\mathscr{L}\left(\sigma^2, \Sigma\right) = \frac{1}{\sigma^2}\hat{u}'\hat{u} + \hat{v}'V^{-1}\hat{v} - \hat{w}'W^{-1}\hat{w} + 2\log\det\left(PP'\right) - T\left(\log 2 + \log\pi\right).$$

As, according to Claim 2, $\hat{w}'W^{-1}\hat{w}$ equals $\frac{1}{\sigma^2}\hat{u}'\hat{u} + \hat{v}'V^{-1}\hat{v}$ and $\log\det\left(PP'\right)$ and $T\left(\log 2 + \log\pi\right)$ are independent of the variances, we can write

$$\mathscr{C}_L = -2\mathscr{L}\left(\sigma^2, \Sigma\right) + \text{constant}$$

where "constant" is independent of the variances and maximization of $\mathscr{L}$ with regard to the variances is equivalent to minimization of $\mathscr{C}_L$. □

10. ANOTHER REPRESENTATION OF THE MOMENTS ESTIMATOR

The relationship between the likelihood estimator and the moments estimator can be elucidated with the aid of a criterion that is very similar to the likelihood criterion (9.3). This criterion function is

$$\mathscr{C}_M\left(\sigma^2, \Sigma\right) = \log\det M + (T-1)\sum_{i=1}^{n}\log\sigma_i^2 - T(n-1)\log\sigma^2 + \\ + \frac{1}{\sigma^2}\hat{u}'\hat{u} + \hat{v}'V^{-1}\hat{v}. \qquad (10.1)$$

Claim 6. Minimization of the criterion function (10.1) with respect to the variances σ^2 and Σ yields the moments estimators as defined in (8.3) and (8.4).

Proof. Note that the envelope theorem together with (8.2) implies

$$\frac{\partial}{\partial\sigma^2}\left(\frac{1}{\sigma^2}\hat{u}'\hat{u} + \hat{v}'V^{-1}\hat{v}\right) = -\frac{1}{\sigma^4}\hat{u}'\hat{u} \qquad (10.2)$$

$$\frac{\partial}{\partial\sigma_i^2}\left(\frac{1}{\sigma^2}\hat{u}'\hat{u} + \hat{v}'V^{-1}\hat{v}\right) = -\frac{\sigma^2}{\sigma_i^4}\hat{v}_i'\hat{v}_i. \qquad (10.3)$$

In view of (8.2) we obtain further

$$\frac{\partial \log\det M}{\partial \sigma^2} \quad = \quad \mathrm{tr}\left(M^{-1}P'V^{-1}P\right). \tag{10.4}$$

By definition (7.3) we have

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

and hence

$$M^{-1}P'V^{-1}P \quad = \quad \frac{1}{\sigma^2}\left(I-M^{-1}X'X\right).$$

With this, equation (10.4). can be written as

$$\begin{aligned}
\frac{\partial \log\det M}{\partial \sigma^2} &= \mathrm{tr}\left(\frac{1}{\sigma^2}\left(I_{Tn}-M^{-1}X'X\right)\right)\\
&= \frac{1}{\sigma^2}\left(\mathrm{tr} I_{Tn}-\mathrm{tr} M^{-1}X'X\right)\\
&= \frac{Tn}{\sigma^2}-\frac{1}{\sigma^2}\mathrm{tr} XM^{-1}X'.
\end{aligned}$$

$$\frac{\partial \log\det M}{\partial \sigma_i^2} \quad = \quad -\frac{\sigma^2}{\sigma_i^4}\mathrm{tr}\left(M^{-1}P'E_i'E_iP\right)$$

and we find

$$\frac{\partial \mathscr{C}_M}{\partial \sigma^2} \quad = \quad \frac{Tn}{\sigma^2}-\frac{1}{\sigma^2}\mathrm{tr} XM^{-1}X'.-\frac{T(n-1)}{\sigma^2}-\frac{1}{\sigma^4}\hat{u}'\hat{u}=0 \tag{10.5}$$

$$\frac{\partial \mathscr{C}_M}{\partial \sigma_i^2} \quad = \quad -\frac{\sigma^2}{\sigma_i^4}\mathrm{tr} P'F_i'F_iPM^{-1}+(T-1)\frac{1}{\sigma_i^2}-\frac{\sigma^2}{\sigma_i^4}\hat{v}_i{}'\hat{v}_i=0 \tag{10.6}$$

which gives

$$\begin{aligned}
\hat{u}'\hat{u} &= \sigma^2\left(T-\sigma^2\mathrm{tr} XM^{-1}X'\right)\\
\hat{v}_i{}'\hat{v}_i &= (T-1)\sigma_i^2-\sigma^2\mathrm{tr} P'F_i'F_iPM^{-1}.
\end{aligned}$$

These first-order conditions are equivalent to equations (8.3), (8.4) that define the moments estimator. □

Johannes Ludsteck's (2004, 2018) Mathematica packages for VC proceed by minimizing the criterion function (10.1). This permits very clean and transparent programming. As Claim 6 is confined to moments and does not require any assumption about the normality of the disturbances, Ludsteck's estimators are moments estimators as well.

11. THE RELATIONSHIP BETWEEN THE LIKELIHOOD AND THE MOMENTS ESTIMATOR

The likelihood estimates minimize, according to Claim 5, the criterion $\mathscr{C}_L$ and the moments estimates minimize, according to Claim 6, the criterion $\mathscr{C}_M$. It is claimed in the following that, for increasing T and bounded X, both estimates tend to coincide.

Claim 7. For sufficiently large T and bounded explanatory variables X, the following holds true approximately:

$$\det PMP' \quad \approx \quad \det M \det\left(PP'\right).$$

Proof. Define the $Tn \times Tn$ matrix

$$\mathbb{P} \quad = \quad \begin{pmatrix} P \\ T^{-\frac{1}{2}} Z' \end{pmatrix}$$

and consider the matrix $\mathbb{P}M\mathbb{P}'$. One way to calculate it is as follows:

$$\begin{aligned}
\mathbb{P}M\mathbb{P}' \quad &= \quad \begin{pmatrix} P \\ T^{-\frac{1}{2}} Z' \end{pmatrix} M \begin{pmatrix} P' & T^{-\frac{1}{2}} Z \end{pmatrix} \\
&= \quad \begin{pmatrix} PMP' & T^{-\frac{1}{2}} PMZ \\ T^{-\frac{1}{2}} Z'MP' & T^{-1} Z'Z \end{pmatrix} \\
&= \quad \begin{pmatrix} PMP' & T^{-\frac{1}{2}} PX'XZ \\ T^{-\frac{1}{2}} Z'X'XP' & I_n \end{pmatrix}.
\end{aligned}$$

This implies

$$\begin{aligned}\det \mathbb{P}M\mathbb{P}' &= \det I_n \det\left(PMP' - \frac{1}{T}PX'XZZ'X'XP'\right)\\ &= \det\left(PMP' - \frac{1}{T}PX'xx'XP'\right)\\ &= \det\left(P\left(M - \frac{1}{T}X'xx'X\right)P'\right)\\ &= \det\left(P\left(X'\left(I_T - \frac{1}{T}xx'\right)X + \sigma^2 P'V^{-1}P\right)P'\right).\end{aligned}$$

For increasing T and bounded x, $\frac{1}{T}xx'$ tends to zero and $\left(I_T - \frac{1}{T}xx'\right)$ tends to I_T. Hence $\det \mathbb{P}M\mathbb{P}'$ tends to $\det PMP'$ and we can write

$$\det \mathbb{P}M\mathbb{P}' \approx \det PMP' \tag{11.1}$$

for large T. Another way to evaluate $\det(\mathbb{P}M\mathbb{P})$ is the following:

$$\begin{aligned}\det \mathbb{P}M\mathbb{P}' &= \det\left(M\mathbb{P}'\mathbb{P}\right)\\ &= \det M \det\left(\mathbb{P}'\mathbb{P}\right)\\ &= \det M \det\left(\mathbb{P}\mathbb{P}'\right)\end{aligned}$$

As

$$\det\left(\mathbb{P}\mathbb{P}'\right) = \det\begin{pmatrix} PP' & 0 \\ 0 & I_n\end{pmatrix} = \det\left(PP'\right),$$

$$\det \mathbb{P}M\mathbb{P}' = \det M \det\left(PP'\right) \tag{11.2}$$

is obtained. Combining (11.1) and (11.2) gives the result. □

Claim 8. For increasing T and bounded explanatory variables X, the moments criterion and the likelihood criterion coincide.

Proof. For large T and in view Claim (7), $\mathscr{C}_M$ and $\mathscr{C}_L$ differ by the constant $\log\det\left(PP'\right) + n$. Hence the minimization of both criteria with respect to the variances will generate the same result. □

In consequence, the descriptive appeal of the moments estimator carries over to the likelihood estimator, and the theoretical appeal of the likelihood estimator carries over to the moments estimator.

PART II
NOTES ON THE VC METHOD

The actual workings of the VC method are best illustrated by the applications found in the literature. Some are listed in Appendix A. As any of the authors of these studies will be a better judge regarding the practical performance of the VC method than this author (who is neither an applied economist, nor an econometrican, nor a statistician), any comments in this regard from my side appear unwarranted. Further, Schlicht and Ludsteck (2006) offer Monte-Carlo studies that illustrate the performance of the VC method from a statistical point of view.

12. AN ILLUSTRATION[3]

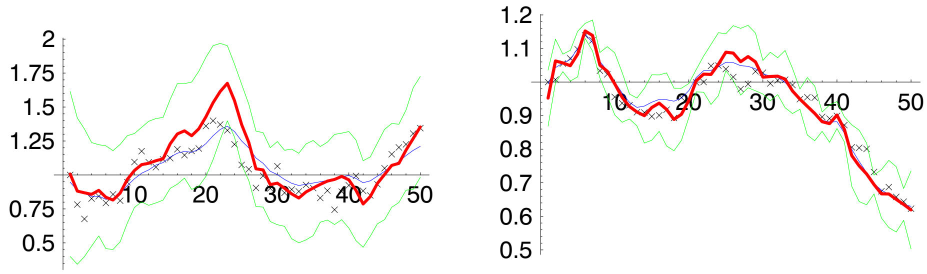
To illustrate the working of VC, assume a model with an intercept term a_t and a single explanatory variable x_t with coefficient b_t:

$$y_t \quad = \quad a_t + b_t x_t + u_t$$

Using the simulation tool from Ludsteck (2004; 2018), a time series for the explanatory 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 12.1 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

[3]This is taken from Schlicht and Ludsteck (2006, Sec. 10)



(A) Intercept

(B) Slope

FIGURE 12.1. 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 $\hat{\gamma}_a = 1.60$ and $\hat{\gamma}_b = 14.76$ 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 $\hat{\sigma}_u^2 = 0.040$, $\hat{\sigma}_a^2 = 0.025$, and $\hat{\sigma}_b^2 = 0.0029$.)

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 12.1.

The distribution of the weights in the above setting is illustrated in Figure 12.2. The time series for x, u, and v have been generated as described above and the VC moments estimation applied 5000 times. The histogram Figure 12.2 illustrates that the estimates cluster around their theoretical values.

13. ARTIFACTS

Suppose that the data have been generated by the standard linear model (2.2). If this is the case, the VC model is slightly misspecified, because a correct estimation would require that the variances $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$ of the coefficients are zero and the weights $\gamma_1, \gamma_2, \ldots, \gamma_n$ – the inverse variance

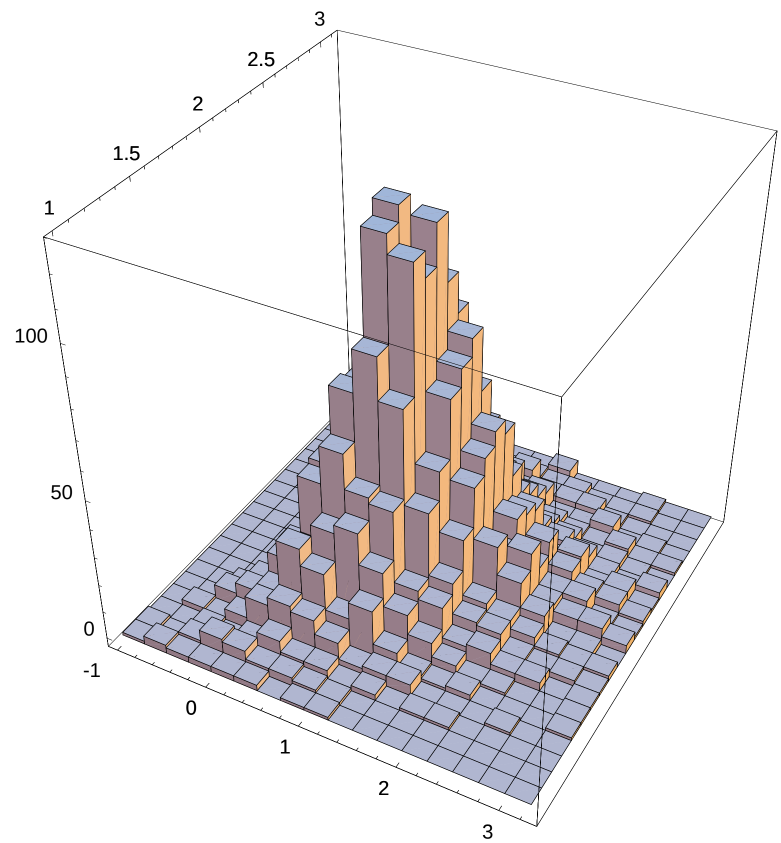


FIGURE 12.2. Histogram of estimates for the $\log_{10}$ weights. The theoretical values are $\log_{10}\gamma_a = 1$ and $\log_{10}\gamma_b = 2$. The distribution of estimates clusters around this peak. ($T = 50$, 5000 trials.)

ratios – are infinite whereas VC implicitly assumes that the weights are finite. As it appears that the VC estimates with sufficiently large weights γ_i are indistinguishable from the OLS estimates, the VC estimation would be approximately correct if the estimated weights are sufficiently large.[4]

As VC estimates nearly twice as many parameters as OLS, there is more room for artifacts in VC. From this point of view, VC ought to be used with caution, especially if all parameters are permitted to vary over time, rather just a selected few. To illustrate, consider a linear model $y_t = a + bx_t + u_t$ with $a = 1$, $b = 2$, x_t drawn from a Normal distribution with mean zero and variance 5, and u_t normally distributed with mean zero and variance $\sigma_u^2 =$

[4]The option "keep selected coefficients constant" in Schlicht (2005a) and Schlicht (2005b) is implemented with $\sigma_i^2 = 10^{-10}$ for those coefficients that are kept constant.

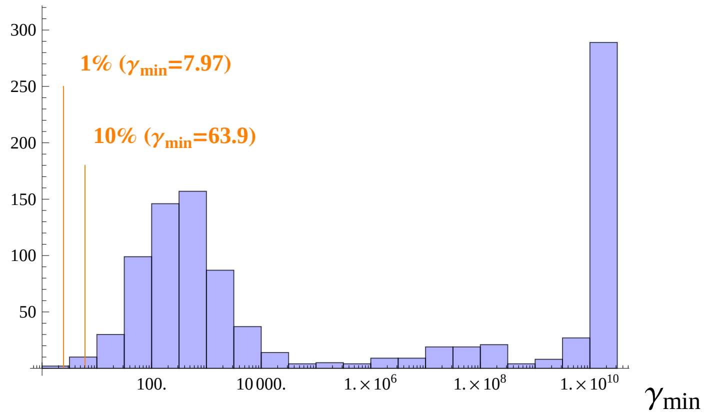
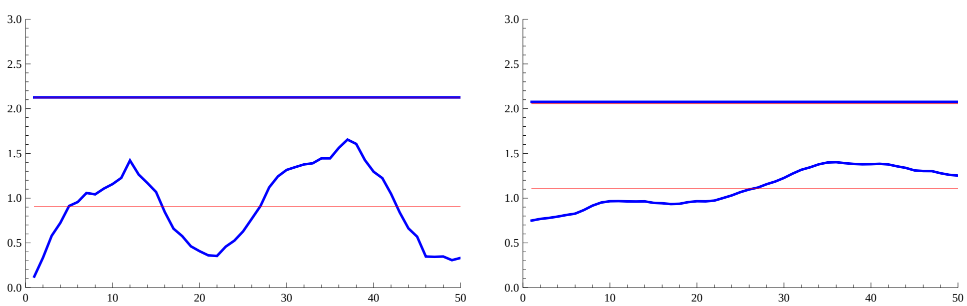


FIGURE 13.1. Histogram of lowest weights $\gamma_{\min} = \min\{\gamma_1, \gamma_2\}$ of VC estimates for a linear model with time-invariant coefficients. ($T = 50$, 1000 trials.)



(A) 1% (B) 10%

FIGURE 13.2. VC estimates at the 1% quantile (A) and the 10% quantile (B) of Figure 13.1. The red lines indicate the OLS estimates of the coefficients. The true coefficients are 1 and 2.

1. The histogram of the lowest estimated weights is given in Figure 13.1 In 99% of the cases, the minimum weight is above 7.97, and in 90% of the cases, the minimum weight is above 63,9. The corresponding VC estimates are given in Figure 13.2. In the 1% case, the estimate of the time paths involve severe artifacts. In the 10% case, artifacts are still pronounced, but in the majority of cases, VC estimates conform to OLS estimates.

With lower noise ($\sigma_u^2 = 0.1$ rather than $\sigma_u^2 = 1$ in the above example) the problem of artifacts is significantly reduced. Still the problem has to be kept in mind when interpreting VC results.

14. AGGREGATE DATA, PYRRHO'S LEMMA, AND THE VC PHILOSOPHY

Almost all economic models deal with aggregate data. Employment comprises women and men, different age groups and various occupations in sundry industries scattered over many regions. The wage level summarizes the earnings of all these people. Similarly, production comprises a multitude of goods and services, and the price level is just an index of thousands of the attached prices. The structures of these aggregates are not rigid but change over time in response to changing technologies, shifting tastes, and volatile business conditions. To assume that time-invariant laws govern the interaction of time series of such aggregates seem preposterous to me. Some researchers tried to cope with the problem by using weighted regression – giving higher weights to more recent observations (Gilchrist 1967, Rouhiainen 1978). This seems to me to be an inferior alternative to VC.

The reason for developing VC was my desire to show that a Marshallian view of economics, that involves time-varying structures does not render quantitative economics impossible. Estimation can be done by using Kalman filtering, or the VC method described in this paper, or perhaps other methods. I advocated estimating time-varying structures with Kalman filtering in Schlicht (1977, Appendix B), but without any resonance. This puzzled me. Was this really such a bad idea?

Maybe it wasn't, but the puzzle remains. What were the reasons for the decade-long resistance to dealing with time-varying coefficients? And why has this somewhat changed over the past fifteen years?

One reason may have been that structures changing over time cannot represent the 'true model' economists were chasing during the heydays of 'dynamic stochastic general equilibrium' macroeconomics. The existence of such a 'true model' was simply postulated (Lucas, 1976, p. 24). I think that this is, in the context of aggregate models dealing with long-run time series, a red herring, distracting from considering seriously what aggregate models represent.[5]

Another reason, I submit, was the reductionist bent of economists. If a structure changes over time, this warrants explanation. Hence there was a

[5] My view if aggregation is outlined in Schlicht (1977), Schlicht (1985), and Schlicht (1990).

tendency to add additional explanatory variables as 'controls' in order to explain the change. While this may be sensible in many cases, it seems, statistically speaking, a problematic way of dealing with time-varying coefficients because of the following theorem that has been provided by Theo Dijkstra (1995, p. 122).

> *Pyrrho's Lemma*: For every collection of vectors, consisting of observations on a regressand and regressors, it is possible to get any set of coefficients as well as any set of predictions with variances as small as one desires, just by adding one additional vector from a continuum of vectors.

In other words: There exists a time series x_{n+1} that, if added to the explanatory variables $x_{1,} x_2, ..., x_n$ in the standard linear model (2.2), will deliver arbitrarily predetermined coefficients and variances as estimates. This should make us reluctant to seek to explain too much by inserting additional controls which, taken together, span an entire set of such additional time series. Further, the procedure can generate the mirage of a 'true model' in cases when such a model actually does not exist. Using VC reduces the necessity for adding further controls and mitigates, therefore, Pyrrho's problem.

Let me add another remark. The VC model (2.4), (2.5) can easily be generalized in many ways. A possibility would be, for instance, to replace $a_{i,t+1} = a_{i,t} + v_{i,t}$ by $a_{i,t+1} = \theta_i \left(a_{i,t} - \bar{a}_i\right) + v_{i,t}$. Such generalizations (and many more) can be handled by Kalman filtering. So why not allow for more general specifications?

My objection would be that such generalizations would impinge on the descriptive transparency of the VC method which is, to me, a major concern – trumping more technical statistical considerations.

An estimation method, such as VC, can be viewed as a filter that seeks to identify certain patterns in clouds of data. In doing so, such a filter gives preference to certain patterns rather than others. The patterns preferred by the VC method are the desiderata underlying the descriptive account (Section 4). These are that that the coefficients remain as time-invariant as possible and that a good fit is obtained. This makes sure that all time-variance estimated is driven by the data, rather than a preference of the model, as would be the case in auto-regressive specifications.

Unfortunately the determination of weights used in VC is descriptively less transparent than the desiderata of stable coefficients and a good fit, but it carries nevertheless some descriptive meaning; in this regard, at least, there is room for improvement.

REFERENCES

Athans, Michael (1974). “The Importance of Kalman Filtering Methods for Economic Systems”. In: *Annals of Economic and Social Measurement, Volume 3, number 1*. NBER Chapters. National Bureau of Economic Research, Inc, pp. 49–64. URL: *https://ideas.repec.org/h/nbr/nberch/9994.html*.

Cooley, Thomas F and Edward C Prescott (1973). “An Adaptive Regression Model”. In: *International Economic Review* 14.2, pp. 364–371. URL: *https://ideas.repec.org/a/ier/iecrev/v14y1973i2p364-71.html*.

Dijkstra, Theo (1995). “Pyrrho’s lemma, or have it your way”. In: *Metrika: International Journal for Theoretical and Applied Statistics* 42.1, pp. 119–125. DOI: *10.1007/s00184-005-0020-0*. URL: *https://ideas.repec.org/a/spr/metrik/v42y1995i1p119-125.html*.

Gilchrist, W. G. (1967). “Methods of Estimation Involving Discounting”. In: *Journal of the Royal Statistical Society. Series B (Methodological)* 29.2, pp. 355–369. ISSN: 00359246. URL: *http://www.jstor.org/stable/2984595*.

Hodrick, Robert J and Edward C Prescott (1997). “Postwar U.S. Business Cycles: An Empirical Investigation”. In: *Journal of Money, Credit and Banking* 29.1, pp. 1–16. URL: *https://ideas.repec.org/a/mcb/jmoncb/v29y1997i1p1-16.html*.

Keynes, John Maynard (1939). “Professor Tinbergen’s Method”. In: *Economic Journal* 49.195, pp. 558–68. URL: *https://www.jstor.org/stable/2224838*.

– (1973). *The General Theory and After, Part II: Defense and Development*. Ed. by Moggridge. Vol. xiv. The Collected Works of John Maynard Keynes. London: Macmillan.

Leser, C. E. V. (1961). "A Simple Method of Trend Construction". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 23, pp. 91–107. URL: *http://www.jstor.org/stable/2983845*.

– (1963). "Estimation of Quasi-Linear Trend and Seasonal Variation". In: *Journal of the American Statistical Association* 58, pp. 1033–43. URL: *http://www.jstor.org/stable/2283329*.

Lucas, Robert (1976). "The Econometric Policy Evaluation: A Critique". In: *Phillips Curve and Labor Markets*. Ed. by K. Brunner and A. H. Meltzer. Amsterdam: North Holland, pp. 19–46. URL: *https://EconPapers.repec.org/RePEc:eee:crcspp:v:1:y:1976:i::p:19-46*.

Ludsteck, Johannes (2004). *VC Package for Mathematica*. URL: *http://library.wolfram.com/infocenter/MathSource/5195/*.

– (2018). *VC Packages for Estimating Time-Varying Coefficients with Mathematica 8 - 11*. URL: *https://epub.ub.uni-muenchen.de/59479/*.

Rouhiainen, Juhani (1978). "The problem of changing parameters in demand analysis and forecasting". In: *European Review of Agricultural Economics* 5.3-4, pp. 349–359. URL: *https://ideas.repec.org/a/oup/erevae/v5y1978i3-4p349-359..html*.

Schlicht, Ekkehart (1973). *Forcasting Markov Chains. A Theoretical Foundation for Exponential Smoothing*. Working paper B 13, Department of Economics, University of Regensburg. URL: *http://www.semverteilung.vwl.uni-muenchen.de/mitarbeiter/es/paper/schlicht-exponential_smoothing.pdf*.

– (1977). *Grundlagen der ökonomischen Analyse*. Reinbek: Rowohlt. URL: *https://epub.ub.uni-muenchen.de/25821/*.

– (1981). "A Seasonal Adjustment Principle and a Seasonal Adjustment Method Derived from This Principle". In: *Journal of the American Statistical Association* 76(374). Paper presented at the Econometric Society European Meeting Helsinki 1976, pp. 374–78. URL: *http://www.lrz.de/~ekkehart*.

– (1985). *Isolation and Aggregation in Economics*. Annotated electronic reprint. 38821. Springer-Verlag. URL: *https://ideas.repec.org/p/lmu/muenec/38821.html*.

– (1989). “Variance Estimation in a Random Coefficients Model”. In: *Munich Discussion Paper*. mwb 2006-12. Paper presented at the Econometric Society European Meeting Munich 1989. URL: *https://epub.ub.uni-muenchen.de/59143/*.

Schlicht, Ekkehart (1990). “Local aggregation in a dynamic setting”. In: *Journal of Economics* 51.3, pp. 287–305. DOI: *10.1007/s00712-008-0015-1*. URL: *https://ideas.repec.org/a/kap/jeczfn/v51y1990i3p287-305.html*.

– (2005a). *VC - A Program for Estimating Time-Varying Coefficients*. URL: *http://epub.ub.uni-muenchen.de/archive/00000684/*.

– (2005b). *VCC - A Program for Estimating Time-Varying Coefficients. Console Version With Source Code in C*. URL: *http://epub.ub.uni-muenchen.de/archive/00000719/*.

Schlicht, Ekkehart and Johannes Ludsteck (2006). “Variance Estimation in a Random Coefficients Model”. In: *Munich Discussion Paper* 2006-12. URL: *https://ideas.repec.org/p/lmu/muenec/904.html*.

Schlicht, Ekkehart and Ralf Pauly (1983). “Descriptive Seasonal Adjustment by Minimizing Perturbations”. In: *Empirica*. URL: *http://www.semverteilung.vwl.uni-muenchen.de/mitarbeiter/es/paper/schlicht-pauly-perturbations.pdf*.

Tinbergen, Jan (1940). “On a Method of Statistical Business-Cycle Research. A Reply”. In: *Economic Journal* 50.197, pp. 141–54. URL: *http://www.jstor.org/stable/2225763*.

APPENDIX A. CONTRIBUTIONS THAT HAVE EMPLOYED VC

Afonso, Antonio, Pedro Gomes, and Davide Furceri (2011). “Sovereign credit ratings and financial markets linkages: application to European data”. In: *Working Paper Series* 1347. URL: *https://ideas.repec.org/p/ecb/ecbwps/20111347.html*.

Afonso, Antonio and Joao Tovar Jalles (2015). “Economic Volatility and Sovereign Yields’ Determinants: A Time-Varying Approach”. In: *Working Papers Department of Economics*. URL: *https://ssrn.com/abstract=2712221*.

– (2017a). "Euro area time-varying fiscal sustainability". In: *International Journal of Finance & Economics* 22.3, pp. 244–254. URL: *https://ideas.repec.org/a/wly/ijfiec/v22y2017i3p244-254.html*.

Afonso, Antonio and Joao Tovar Jalles (2017b). "Sovereign debt composition and time-varying public finance sustainability". In: *The North American Journal of Economics and Finance* 42.C, pp. 144–155. DOI: *10.1016/j.najef.2017.08.0*. URL: *https://ideas.repec.org/a/eee/ecofin/v42y2017icp144-155.html*.

– (2017c). "Sovereign Debt Effects and Composition: Evidence from Time-Varying Estimates". In: *Working Papers Department of Economics* 2017/03. URL: *https://ideas.repec.org/p/ise/isegwp/wp032017.html*.

– (2019). "Decomposing and Analysing the Determinants of Current Accounts' Cyclicality: Evidence from the Euro Area". In: *Open Economies Review*. DOI: *10.1007/s11079-018-9503-2*. URL: *https://ideas.repec.org/a/kap/openec/v30y2019i1d10.1007_s11079-018-9503-2.html*.

Baxa, Jaromir, Roman Horvath, and Borek Vasicek (2011). "Time Varying Monetary Policy Rules and Financial Stress". In: ed. by S. Eijffinger and Donato Masciandaro. Edward Elgar Publishing. Chap. Time Varying Monetary Policy Rules and Financial Stress, pp. 268–312. URL: *http://www.e-elgar.com/shop/handbook-of-central-banking-financial-regulation-and-supervision*.

– (2014). "How Does Monetary Policy Change? Evidence on Inflation-Targeting Countries". In: *Macroeconomic Dynamics* 18.3, pp. 593–630. DOI: *10.1017/S1365100512000545*.

Cohen, Gail et al. (2017). "Emissions and Growth: Trends and Cycles in a Globalized World". In: *IMF Working Paper* WP17/191. URL: *https://www.imf.org/~/media/Files/Publications/WP/2017/wp17191.ashx*.

Conway, P., R. Herd, and T. Chalaux (2010). "Reforming China's Monetary Policy Framework to Meet Domestic Objectives". In: *6th NIPFP DEA Research Program Conference*. DOI: *doi.org/10.1787/5km32vmsq6f2-en*.

Cournede, Boris, Rudiger Ahrend, and Robert Price (2008). "Have Long-Term Financial Trends Changed the Transmission of Monetary Policy?" In: *OECD Economic Department Working Papers* 634. URL: *http://dx.doi.org/10.1787/238203348082*.

Cournede, Boris and Frederic Gonand (2006). "Restoring Fiscal Sustainability in the Euro Area: Raise Taxes or Curb Spending?" In: *Working Paper*. URL: *http://pcsi.pa.go.kr/files/JT03216787.pdf*.

Flaschel, P. et al. (2008). *Topics in Applied Macrodynamic Theory*. Berlin-Heidelberg: Springer-Verlag. DOI: *10.1007/978-3-540-72542-8*.

Flaschel, Peter (2008). "Linking Goods with Labor Markets: Okun's Law and Beyond". In: *Topics in Applied Macrodynamic Theory*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 213–283. ISBN: 978-3-540-72542-8. DOI: *10.1007/978-3-540-72542-8_5*. URL: *https://doi.org/10.1007/978-3-540-72542-8_5*.

Flaschel, Peter and Ekkehart Schlicht (2006). "The New Keynesian Theory and the New Phillips Curves: A Competing Approach". In: *Quantitative and Empirical Analysis of Nonlinear Dynamic Macromodels*. Ed. by Carl Chiarella et al. Elsevier. Chap. The New Keynesian Theory and the New Phillips Curves: A Competing Approach, pp. 113–145. URL: *https://books.google.de/books?id=5MmwM5vyhUQC*.

Franke, Reiner (2006). *Themes on Okun's Law and Beyond*. Tech. rep. SCEPA.

Furceri, Davide and Joao Tovar Jalles (2017). "Fiscal Stabilization and Productive Investment: Evidence from Advanced Economies". In: *Working Paper*. URL: *https://www.bancaditalia.it/pubblicazioni/altri-atti-convegni/2018-workshop-public-finance/7-Furceri-Jalles.pdf?language_id=1*.

– (2018). "Fiscal counter-cyclicality and productive investment: evidence from advanced economies". In: *The B.E. Journal of Macroeconomics*. DOI: *10.1515/bejm-2017-0222*.

Furceri, Davide, Joao Tovar Jalles, and Aleksandra Zdzienicka (2017). "China Spillovers: New Evidence from Time-Varying Estimates". In: *Open Economies Review* 28.3, pp. 413–429. URL: *https://link.springer.com/article/10.1007/s11079-016-9430-z*.

Haan, Wouter J. Den and Thomas Drechsel (2018). "Agnostic Structural Disturbances (ASDs) : Detecting and Reducing Misspecification in Empirical Macroeconomic Models". In: URL: *https://www.semanticscholar.org/paper/Agnostic-Structural-Disturbances-(-ASDs-)-%3A-and-in-Haan-Drechsel/0c724c80bfa5c50f9eb68fff65d9df711e1e7802*.

He, Qizhi (2012). "Research on the Dynamic Relationship between Prices of Agricultural Futures in China and Japan". In: *Journal of Computers* 7(9). DOI: *10.4304/jcp.7.9.2342-2350*.

Jalles, Joao Tovar (2018). "On the Time-Varying Relationship between Unemployment and Output: What shapes it?" In: *Scottish Journal of Political Economy* 0. DOI: *10.1111/sjpe.12200*.

López-Herrera, Francisco, Humberto Valencia-Herrera, et al. (2016). "Hacia un Modelo de Valuación de Activos de Capital para México: Análisis de Activos Individuales con Coeficientes Variantes en el Tiempo". In: *Panorama Económico* 11.22, pp. 75–103. URL: *http://yuss.me/revistas/panorama/pano2016v11n22a04p074_103.pdf*.

Madsen, Michael Walter (2012). "Does Financial Stress Have an Impact on Monetary Policy?: An Econometric Analysis Using Norwegian Data". PhD thesis. Department of Economics, University of Oslo. URL: *https://www.duo.uio.no/bitstream/handle/10852/17126/Madsen_Michael_Masteroppgave.pdf*.

Nogueira, Reginaldo Pinto (2009). "Testing credibility with time-varying coefficients". In: *Applied Economics Letters* 16.18, pp. 1813–1817. DOI: *10.1080/13504850701719611*.

Rodriguez, Cesar M. (2014). "Financial development, fiscal policy and volatility: Their effects on growth". In: *The Journal of International Trade & Economic Development* 23.2, pp. 223–266. DOI: *10.1080/09638199.2012.711014*.

EKKEHART SCHLICHT IS PROFESSOR EMERITUS OF ECONOMICS AT THE DEPARTMENT OF ECONOMICS, UNIVERSITY OF MUNICH. GERMANY.

Current address: Hurtenstrasse 13, 82346 Andechs, Germany

Email address: schlicht@lmu.de