RESEARCH ARTICLE

VC: a method for estimating time-varying coefficients in linear models

Ekkehart Schlicht[1]

Abstract

This paper describes a moments estimator for a standard state-space model with coefficients generated by a random walk. The method calculates the conditional expectations of the coefficients, given the observations. A penalized least squares estimation is linked to the GLS (Aitken) estimates of the corresponding linear model with time-invariant parameters. The estimates are moments estimates. They do not require the disturbances to be Gaussian, but if they are, the estimates are asymptotically equivalent to maximum likelihood estimates. In contrast to Kalman filtering, no specification of an initial state or an initial covariance matrix is required. While the Kalman filter is one sided, the filter proposed here is two sided and therefore uses more of the available information for estimating intermediate states. Further, the proposed filter has a clear descriptive interpretation.

Keywords Time-series analysis · Linear model · State-space estimation · Time-varying coefficients · Moments estimation. Kalman filtering · Penalized least squares · HP-Filter

JEL Classification C2 · C22 · C32 · C51 · C52

1 Introduction

This paper describes and discusses an estimator for a linear time series model with time-varying coefficients. Such a model, the varying 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.

✉ Ekkehart Schlicht
   ekkehart.schlicht@econ.lmu.de

1   Department of Economics, Ludwig-Maximilians-Universität Munich, Geschwister-Scholl-Platz 1, 80539 Munich, Germany

Springer

The VC model poses the statistical problem of determining the smoothing parameters, variances, or splines, that are needed to model the movements of the coefficients over time. Schlicht (1989) and Schlicht and Ludsteck (2006) have proposed an estimation method—the VC method—that is specifically tailored to the case that the time-varying coefficients are generated by a random walk with normal disturbances. The present paper generalizes this approach to the non-Gaussian case.[1]

The paper is organized as follows. In Sect. 1, the model is described and some motivation is provided by drawing on a discussion in economics. The "criteria" or "penalty" approach is explained that permits to estimate the time-paths of the coefficients in a purely descriptive way. A stochastic formulation is given in Sect. 2 which presupposes knowledge of the variances of the disturbances. It is shown that the descriptive approach outlined in Sect. 1 can be justified statistically by estimating the time averages of the coefficients in the corresponding linear GLS model with constant coefficients, again for given variances of the disturbances. Section 3 turns to the estimation of these variances. Two closely related estimators are explained: a moments estimator and a likelihood estimator. The likelihood estimator presupposes Gaussian disturbances while the moments estimator does not require this assumption. The relationship between these estimators is discussed. Section 4 provides some illustration of the way the model works and ponders some methodological issues. A conclusion follows.

2 The varying coefficients model in descriptive mode

This section describes the varying coefficients model and discusses the motivation for using it, as it emerged in economics (Sect. 1.1). Some features of the proposed method will be previewed in Sect. 1.2. The notation is introduced in Sect. 1.3 and the "criteria" or "penalty" approach is described that permits to estimate the development of the relations between the independent variable and the dependent variables over time in a purely descriptive way (Sect. 1.4).

2.1 The linear theoretical model and its empirical application

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

$$y = a_1 x_1 + a_2 x_2 + \cdots + a_n x_n. \tag{1.1}$$

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

---

[1] The calculations in this paper follow largely the calculations given in two unrefereed discussion papers (Schlicht 1989; Schlicht and Ludsteck 2006), for which this author was responsible—including some errors. In this paper the calculations have been rearranged, corrected, improved and adapted to cover the case that all disturbances are non-Gaussian.

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

$$y_t = a_1 x_{1,t} + a_2 x_{2,t} + \cdots + a_n x_{n,t} + u_t, \quad t = 1, 2, \ldots, T. \tag{1.2}$$

In many cases it appears improbable, however, that outside influences not captured in the theoretical model affect only the disturbance term, and not the coefficients themselves. In the case of economics, we may think of changes in technology, preferences, market structure, and the composition of aggregates. All change over time and may affect the coefficients themselves.

In economics, the problem of possibly time-varying coefficients was the subject of the famous Keynes–Tinbergen controversy around 1940.[2] While Tinbergen (1940, 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, 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 *slowly* over time: They are highly auto-correlated. This is formalized by a random walk (Athans 1974; Cooley and Prescott 1973; Schlicht 1973). If $a_{i,t}$ denotes the state of coefficient a_i at time t, it is assumed that

$$a_{i,t+1} = a_{i,t} + v_{i,t} \tag{1.3}$$

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 $\sigma_i^2 = 0$ for some i, the case of a constant (time-invariant) coefficients is covered as well. As a consequence, the standard linear model is replaced by

$$y_t = a_{1,t} x_{1,t} + a_{2,t} x_{2,t} + \cdots + a_{n,t} x_{n,t} + u_t, \quad E\{u_t\} = 0, \quad E\{u_t^2\} = \sigma^2, \tag{1.4}$$

$$a_{i,t+1} = a_{i,t} + v_{i,t}, \quad E\{v_{i,t}\} = 0, \quad E\{v_{i,t}^2\} = \sigma_i^2, \quad i = 1, 2, \ldots, n, \quad t = 1, 2, \ldots, T. \tag{1.5}$$

This is the VC model that is presupposed in the following.

2.2 Properties of the VC method

The VC method that will be developed in this paper estimates the expected time-paths of the coefficients $a_{i,t}$ for given observations $x_{i,t}$ and y_t with $i = 1, 2, \ldots, n$ and

---

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

$T = 1, 2, \ldots, T$. 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, i.e. $\sum_{t=1}^{T} u_t^2 + \gamma_1 \sum_{t=2}^{T} v_{1,t}^2 + \gamma_2 \sum_{t=2}^{T} v_{2,t}^2 + \cdots + \gamma_n \sum_{t=2}^{T} v_{n,t}^2$, 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 $\frac{1}{T} \sum_t a_t$ are GLS estimates of the corresponding regression with fixed coefficients.
- The VC method does not require initial values for the initial state and the initial variances. Rather all states and variances are estimated in an integrated unified procedure. This is an advantage over Kalman filtering which is typically quite sensitive to the choice of initial values, especially when dealing with shorter time series.
- The VC method links the purely descriptive method of employing non-parametric splines through penalized least squares with an explicit statistical model with random-walk coefficients. This offers the possibility of model-based estimation.
- All estimates are moments estimates. It is not necessary to presuppose Gaussian disturbances.
- For increasing sample sizes T and under the assumption that all disturbances are normally distributed, the moments estimates approach the maximum likelihood estimates.

2.3 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 $x_t' = (x_{1,t}, x_{2,t}, \ldots, x_{n,t})$ and y_t for $t = 1, 2, \ldots, T$. We write

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

order T $T \times n$ $T \times Tn$

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

order n Tn n $(T-1)n$

We write further

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

order $n \times n$

and define

$$
p = \begin{pmatrix} -1 & 1 & 0 & & 0 \\ 0 & -1 & 1 & 0 & \\ & & . & . & 0 \\ 0 & & & -1 & 1 \end{pmatrix}, \quad
P = p \otimes I_n = \begin{pmatrix} -I_n & I_n & & 0 \\ & -I_n & I_n & \\ & & . & . \\ 0 & & & -I_n & I_n \end{pmatrix} \tag{1.6}
$$

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

with I_n denoting the identity matrix of order n and $\otimes$ indicating the Kronecker product operator. Note that p and P are of full rank.

The model is obtained by writing Eqs. (1.4) and (1.5) in matrix form:
The model

$$
y = Xa + u, \quad E\{u\} = 0, \quad E\{uu'\} = \sigma^2 I_T, \tag{1.7}
$$

$$
Pa = v, \quad\quad E\{v\} = 0, \quad E\{vv'\} = V = I_{T-1} \otimes \Sigma. \tag{1.8}
$$

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 1 $Pa = 0$ implies $y \neq Xa$.

This assumption rules out the (trivial) case that the standard linear model (1.2) 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:

$$T > n. \tag{1.9}$$

2.4 Least squares

In a descriptive spirit, the time-paths of the coefficients can be determined by following the penalized least squares approach, where some criteria are employed that formalize some descriptive desiderata.[3] 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 Eq. (1.7), the weighted sum of the squared changes of the coefficients $v_i'v_i$ over time give criteria for the stability of the coefficients over time. The combination of all these criteria gives an overall criterion that combines the desiderata of a good fit and stability of coefficients over time. The weights $(\gamma_1, \gamma_2, \ldots, \gamma_n)$ 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

$$\Gamma = \mathrm{diag}\left(\gamma_1, \ \gamma_2, \ ., \ ., \ \gamma_n \right) \tag{1.10}$$

and

$$G = I_{T-1} \otimes \Gamma. \tag{1.11}$$

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

$$Q = u'u + v'Gv. \tag{1.12}$$

This expression is to be minimized under the constraints given by the model (1.7), (1.8) with the observations X and y :

---

[3] For the penalized least squares approach, see Green and Silverman (2000). The approach was introduced by Whittaker (1923), Henderson (1924) and Leser (1961). It has been used also by Hodrick and Prescott (1997), and has been further developed by Leser (1963), Schlicht (1981), Schlicht and Pauly (1983), Schlicht (1984) and Schlicht (2005a). Other approaches such as Hastie and Tibshirani (1993) and Fan and Zhang (1999) make use of splines. This line of argument will not be pursued here.

$$u = y - Xa, \tag{1.13}$$

$$v = Pa. \tag{1.14}$$

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

$$Q = (y - Xa)'(y - Xa) + a'P'GPa. \tag{1.15}$$

The weighted sum of squares Q is the sum of two positive semi-definite quadratic forms. Assumption 1 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

$$\frac{\partial Q}{\partial a} = -2Xy + 2(X'X + P'GP)a = 0 \tag{1.16}$$

and the second order condition is that the Jacobian

$$\frac{\partial^2 Q}{\partial a \partial a'} = 2(X'X + P'GP) \tag{1.17}$$

is positive definite, which is the case. Solving (1.16) for a and plugging this into (1.13) and (1.14) gives the estimates

$$a_{LS} = (X'X + P'GP)^{-1}X'y \tag{1.18}$$

$$u_{LS} = \left(I_T - X(X'X + P'GP)^{-1}X'\right)y \tag{1.19}$$

$$v_{LS} = P(X'X + P'GP)^{-1}X'y \tag{1.20}$$

where the subscript LS stands for "least squares".

3 The varying coefficients model in stochastic mode

This section considers the statistical treatment of the VC model under the assumption that the variances of the disturbances are known. With the parametrization outlined in Sect. 2.1, the VC model gives rise to a GLS (Aitken) model that permits to estimate the time-averages of the coefficients. With these estimates, the conditional expectations for the coefficients $a_{i,t}$ for given observations X and y can be determined (Sect. 2.2). If the weights chosen for the descriptive estimation outlined in Sect. 1.4 are equal to the inverse variance ratios, the descriptive estimation and the conditional expectation coincide (Sect. 2.3).

3.1 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 a but rather the set of solutions

$$A := \left\{ a = \tilde{P}v + Z\beta \middle| \beta \in \mathbb{R}^n \right\} \tag{2.1}$$

with β as a shift parameter,

$$\tilde{P} := P'\left(PP'\right)^{-1} \tag{2.2}$$

as the right-hand pseudo-inverse of P given in (1.6) of order $Tn \times (T-1)n$, and the matrix

$$Z := \begin{pmatrix} I_n \\ I_n \\ \cdot \\ I_n \end{pmatrix} \tag{2.3}$$

of order $Tn \times n$. It is orthogonal to P:

$$PZ = 0$$

with the square matrix (P', Z) of full rank. For any v we have $a \in A \Leftrightarrow Pa = v$. Hence Eq. (1.7) and the set (2.1) give equivalent descriptions of the relationship between a and v.

Note that

$$Z'Z = T \cdot I_n, \quad P'\left(PP'\right)^{-1}P + ZZ' = I_{Tn}. \tag{2.4}$$

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

$$\begin{aligned} P\tilde{P} &= \tilde{P}'P' = I_{(T-1)n} \\ \tilde{P}P &= P'\tilde{P}' = I_{Tn} - ZZ' \\ Z'\tilde{P} &= \tilde{P}'Z = 0. \end{aligned} \tag{2.5}$$

In view of (2.1), any solution a to $Pa = v$ can be written as

$$a = \tilde{P}v + Z\beta \tag{2.6}$$

for some $\beta \in \mathbb{R}^n$. Equation (1.7) can be re-written as

$$y = u + X\tilde{P}v + XZ\beta. \tag{2.7}$$

The model (2.6), (2.7) will be referred to as the *equivalent orthogonally parameterized model*. It implies the *true model* (1.7), (1.8). It implies, in particular, that a_t is a random walk even though a_t depends, according to (2.6), on past *and* future realizations of v_t.

The formal parameter β has a straightforward interpretation. Pre-multiplying (2.6) by Z' gives

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

and therefore

$$\beta = \frac{1}{T}\sum_{t=1}^{T} a_t. \tag{2.8}$$

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

Equation (2.7) 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 Sect. 3.1.

The orthogonal parametrization (proposed in Schlicht 1985, Sec. 4.3.3 in another context) 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 (2.7) can be interpreted as a standard regression for this parameter vector with the matrix $x = XZ$ giving the explanatory variables:

$$y = x\beta + w \tag{2.9}$$

and the disturbance

$$w = X\tilde{P}v + u. \tag{2.10}$$

It has expectation zero

$$E\{w\} = 0 \tag{2.11}$$

and covariance

$$W := E\{ww'\} = X\tilde{P}V\tilde{P}'X' + \sigma^2 I_T. \tag{2.12}$$

The Aitken estimate β_A satisfies

$$x'W^{-1}(y - x\beta_A) = 0 \tag{2.13}$$

or

$$\beta_A = \left(x'W^{-1}x\right)^{-1}x'W^{-1}y. \tag{2.14}$$

where the subscript A stands for "Aitken". As $x = XZ$ and $W = X\tilde{P}V\tilde{P}'X' + \sigma^2 I_T$, Eqs. (2.13) and (2.14) can be written as

$$Z'X'\left(X\tilde{P}V\tilde{P}'X' + \sigma^2 I_T\right)^{-1}\left(y - XZ\beta_A\right) = 0 \tag{2.15}$$

and Eq. (2.14) gives rise to

$$\beta_A = \left(Z'X'\left(X\tilde{P}V\tilde{P}'X' + \sigma^2 I_T\right)^{-1}XZ\right)^{-1}Z'X'\left(X\tilde{P}V\tilde{P}'X' + \sigma^2 I_T\right)^{-1}y. \tag{2.16}$$

3.2 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 (2.6), (2.9) 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

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

The marginal distribution of y is as given by (2.9) and (2.12). On this basis, we take our estimate of a as

$$a_A = Z\beta_A + \tilde{P}V\tilde{P}'X'\left(X\tilde{P}V\tilde{P}'X' + \sigma^2 I_T\right)^{-1}\left(y - XZ\beta_A\right) \tag{2.17}$$

which is the expectation of a for the case that u and v are Gaussian and y, β, σ^2, and Σ are given. (It will turn out later on that a_A is the expectation of a for non-Gaussian disturbances as well, see Eq. (2.27) below.)

Note that the variance-covariance matrix of w, as given in Eq. (2.12), tends to $\sigma^2 I_T$ if the the variances σ_i^2 go to zero, and Eq. (2.7) 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.

3.3 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 (1.12). The first-order condition (1.16) defines a_{LS} with weights $G = \sigma^2 V^{-1}$ uniquely and can be written as

$$\left(X'X + \sigma^2 P'V^{-1}P\right)a_{LS} = X'y. \tag{2.18}$$

It will be shown that (2.17) implies

$$\left(X'X + \sigma^2 P'V^{-1}P\right)a_A = X'y, \tag{2.19}$$

which will establish the proposition.

Pre-multiplication of (2.17) by $\left(X'X + \sigma^2 P'V^{-1}P\right)$ gives

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

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

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

According to (2.15), the last term is zero and we obtain

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

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

$\square$

As a consequence of Claim 1, the least-squares estimates for u, v, and w and their Aitken counterparts coincide for $G = \sigma^2 V^{-1}$. We need not distinguish them and denote all our estimates by circumflex:

$$a_A = a_{LS} = \hat{a} = Z\hat{\beta} + \tilde{P}V\tilde{P}'X'\left(X\tilde{P}V\tilde{P}'X' + \sigma^2 I_T\right)^{-1}\left(y - XZ\hat{\beta}\right) \qquad (2.20)$$

$$u_A = u_{LS} = \hat{u} = \left(I_T - X\left(X'X + \sigma^2 P'V^{-1}P\right)^{-1}X'\right)y \qquad (2.21)$$

$$v_A = v_{LS} = \hat{v} = P\left(X'X + P'\sigma^2 P'V^{-1}P\right)^{-1}X'y \qquad (2.22)$$

$$w_A = w_{LS} = \hat{w} = X\tilde{P}\hat{v} + \hat{u}. \qquad (2.23)$$

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

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

Proof As $\hat{w} = X\tilde{P}\hat{v} + \hat{u}$, we have

$$\hat{Q} = \hat{u}'\hat{u} + \sigma^2\hat{v}'V^{-1}\hat{v}$$
$$= \hat{u}'\left(\hat{w} - X\tilde{P}\hat{v}\right) + \sigma^2\hat{v}'V^{-1}\hat{v}$$
$$= \hat{u}'\hat{w} - \hat{u}'X\tilde{P}\hat{v} + \sigma^2\hat{v}'V^{-1}\hat{v}$$
$$= \hat{u}'\hat{w} - \left(\hat{u}'X\tilde{P} - \sigma^2\hat{v}'V^{-1}\right)\hat{v}$$
$$= \hat{u}'\hat{w} - \left(\hat{u}'X\tilde{P} - \sigma^2\hat{v}'V^{-1}\right)P\hat{a}.$$

With (2.5), (2.9), (2.12), and (2.20) this gives

$$\hat{Q} = \hat{u}'\hat{w} - (\hat{u}'X\tilde{P} - \sigma^2\hat{v}'V^{-1})P(Z\hat{\beta} + \tilde{P}V\tilde{P}'X'W^{-1}\hat{w})$$

$$= \hat{u}'\hat{w} - (\hat{u}'X\tilde{P} - \sigma^2\hat{v}'V^{-1})P\tilde{P}V\tilde{P}'X'W^{-1}\hat{w}$$

$$= \hat{u}'\hat{w} - (\hat{u}'X\tilde{P} - \sigma^2\hat{v}'V^{-1})V\tilde{P}'X'W^{-1}\hat{w}$$

$$= \hat{u}'\hat{w} - (\hat{u}'X\tilde{P}V\tilde{P}'X' - \sigma^2\hat{v}'\tilde{P}'X')W^{-1}\hat{w}$$

$$= \hat{u}'\hat{w} - \hat{u}'X\tilde{P}V\tilde{P}'X'W^{-1}\hat{w} + \sigma^2\hat{v}'\tilde{P}'X'W^{-1}\hat{w}$$

$$= \hat{u}'\hat{w} - \hat{u}'(X\tilde{P}V\tilde{P}'X' + \sigma^2 I_T - \sigma^2 I_T)W^{-1}\hat{w} + \sigma^2\hat{v}'\tilde{P}'X'W^{-1}\hat{w}$$

$$= \hat{u}'\hat{w} - \hat{u}'(X\tilde{P}V\tilde{P}'X' + \sigma^2 I_T)W^{-1}\hat{w} + \sigma^2\hat{u}'W^{-1}\hat{w} + \sigma^2\hat{v}'\tilde{P}'X'W^{-1}\hat{w}$$

$$= \hat{u}'\hat{w} - \hat{u}'\hat{w} + \sigma^2\hat{u}'W^{-1}\hat{w} + \sigma^2\hat{v}'\tilde{P}'X'W^{-1}\hat{w}$$

$$= \sigma^2(\hat{u}' + \hat{v}'\tilde{P}'X')W^{-1}\hat{w}$$

and finally

$$\hat{Q} = \sigma^2\hat{w}'W^{-1}\hat{w}.$$

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

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

$$M = (X'X + \sigma^2 P'V^{-1}P). \tag{2.24}$$

With this, the normal equation (2.19), that defines the solution for the vector of the coefficients $\hat{a}$ can be written as

$$M\hat{a} = X'y. \tag{2.25}$$

With (1.7) and (2.24) we obtain

$$\hat{a} = M^{-1}X'(Xa + u)$$

$$= M^{-1}(X'Xa + X'u + \sigma^2 P'V^{-1}Pa - \sigma^2 P'V^{-1}Pa) \tag{2.26}$$

$$= a + M^{-1}(X'u - \sigma^2 P'V^{-1}v).$$

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

$$E\{\hat{a}|a\} = a \tag{2.27}$$

and covariance

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

which reduces to

$$E\{(a - \hat{a})(a - \hat{a})'\} = M^{-1}(\sigma^2 X'X + \sigma^4 P'V^{-1}P)M^{-1}$$

and finally to

$$E\big\{(a - \hat{a})(a - \hat{a})'\big\} = \sigma^2 M^{-1}. \tag{2.28}$$

The system matrix (2.24) is determined by the observations X, the variance σ^2 and the variances Σ. Equation (2.28) 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.

4 Variance estimation

This section turns to estimating the variances. In Sect. 3.1 the proposed moments estimators will be derived and in Sect. 3.2 a maximum likelihood criterion $\mathcal{C}_L$ will be given that is based on the parameterized model described in Sect. 2. In Sect. 3.4 a moments criterion $\mathcal{C}_M$ will be given that generates, upon minimization, the moments estimators and it will be argued that, for large T, both criteria approach each other. As a consequence, the theoretical appeal of the likelihood estimator for large samples carries over to the moments estimator in the Gaussian case.

4.1 Moments estimation

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.

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 $\sum = \text{diag}\big(\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2\big)$, 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 (2.26) we obtain

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

Regarding $\hat{v}$, consider the vectors $\hat{v}_i' = \big(\hat{v}_{i,1}^2, \hat{v}_{i,3}^2, \ldots, \hat{v}_{i,T-1}^2\big)$ for $i = 1, 2, \ldots, n$, 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

$$E_i := I_{T-1} \otimes e_i' \tag{3.1}$$

that picks the time-path of the i–th disturbance $v_i = \left(v_{i,1}, v_{i,3}, \dots, v_{i,T-1}\right)'$ from the disturbance vector v:

$$v_i := E_i v.$$

Note that, from (1.8),

$$\sum_{i=1}^{n} \sigma_i^2 E_i' E_i = V. \tag{3.2}$$

Pre-multiplying (2.26) 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 σ^2 and Σ, the expected squared deviations of $\hat{u}$ and $\hat{v}_i$, $i = 1, 2, \dots, n$ are*

$$E\{\hat{u}'\hat{u}\} = \sigma^2\left(T - \mathrm{tr}\left(XM^{-1}X'\right)\right). \tag{3.3}$$

$$E\{\hat{v}_i'\hat{v}_i\} = (T-1)\sigma_i^2 - \sigma^2 \mathrm{tr}\left(E_i PM^{-1}P'E_i'\right). \tag{3.4}$$

This implies that the expected sum of squares is

$$E\{\hat{Q}\} = \sigma^2(T - n). \tag{3.5}$$

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

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

and hence

$$
E\{\hat{u}'\hat{u}\}=\sigma^2\big(T-\mathrm{tr}\big(XM^{-1}X'\big)\big). \tag{3.6}
$$

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\{\hat{v}_i'\hat{v}_i\} =& E\big\{\big(u'XM^{-1}P'E_i' + v'\big(I_{(T-1)n} - \sigma^2 V^{-1}PM^{-1}P'\big)E_i'\big) \\
& \cdot \big(E_i PM^{-1}X'u + E_i\big(I_{(T-1)n} - \sigma^2 PM^{-1}P'V^{-1}\big)v\big)\big\} \\
=& E\big\{u'XM^{-1}P'E_i'E_i PM^{-1}X'u + v'\big(I_{(T-1)n} - \sigma^2 V^{-1}PM^{-1}P'\big)E_i'E_i \\
& \times \big(I_{(T-1)n} - \sigma^2 PM^{-1}P'V^{-1}\big)v\big\} = E\big\{\mathrm{tr}\big(u'XM^{-1}P'E_i'E_i PM^{-1}X'u \\
& + v'\big(I_{(T-1)n} - \sigma^2 V^{-1}PM^{-1}P'\big)E_i'E_i\big(I_{(T-1)n} - \sigma^2 PM^{-1}P'V^{-1}\big)v\big)\big\} \\
=& E\big\{\mathrm{tr}\big(E_i PM^{-1}X'uu'XM^{-1}P'E_i' + E_i\big(I_{(T-1)n} - \sigma^2 PM^{-1}P'V^{-1}\big)vv' \\
& \times\big(I_{(T-1)n} - \sigma^2 V^{-1}PM^{-1}P'\big)E_i'\big)\big\} = \mathrm{tr}\big(\sigma^2 E_i PM^{-1}X'XM^{-1}P'E_i'\big) \\
& + \mathrm{tr}\big(E_i\big(I_{(T-1)n} - \sigma^2 PM^{-1}P'V^{-1}\big)V\big(I_{(T-1)n} - \sigma^2 V^{-1}PM^{-1}P'\big)E_i'\big) \\
=& \mathrm{tr}\big(\sigma^2 E_i PM^{-1}X'XM^{-1}P'E_i'\big) + \mathrm{tr}\big(E_i\big(V - \sigma^2 PM^{-1}P'\big) \\
& \times\big(I_{(T-1)n} - \sigma^2 V^{-1}PM^{-1}P'\big)E_i'\big) = \mathrm{tr}\big(\sigma^2 E_i PM^{-1}X'XM^{-1}P'E_i'\big) \\
& + \mathrm{tr}\big(E_i\big(V - \sigma^2 PM^{-1}P'\big)E_i' - \sigma^2 E_i\big(V - \sigma^2 PM^{-1}P'\big)V^{-1}PM^{-1}P'E_i'\big) \\
=& \mathrm{tr}\big(\sigma^2 E_i PM^{-1}X'XM^{-1}P'E_i'\big) + \mathrm{tr}\big(E_i\big(V - \sigma^2 PM^{-1}P' - \sigma^2 PM^{-1}P' \\
& + \sigma^4 PM^{-1}P'V^{-1}PM^{-1}P'\big)E_i'\big) = \mathrm{tr}\big(\sigma^2 E_i PM^{-1}X'XM^{-1}P'E_i' \\
& + E_i\big(V - \sigma^2 PM^{-1}P' - \sigma^2 PM^{-1}P' + \sigma^4 PM^{-1}P'V^{-1}PM^{-1}P'\big)E_i'\big) \\
=& \mathrm{tr}\big(E_i\big(\big(\sigma^2 PM^{-1}\big(X'X + \sigma^2 P'V^{-1}P\big)M^{-1}P'\big) + V - 2\sigma^2 PM^{-1}P'\big)E_i'\big) \\
=& \mathrm{tr}\big(E_i\big(V - \sigma^2 PM^{-1}P'\big)E_i'\big) = \mathrm{tr}\big(E_i VE_i' - \sigma^2 E_i PM^{-1}P'E_i'\big) \\
=& \mathrm{tr}\big(\big(I_{T-1} \otimes e_i'\big)\big(I_{T-1} \otimes \Sigma\big)\big(I_{T-1} \otimes e_i\big) - \sigma^2 E_i PM^{-1}P'E_i'\big) \\
=& \mathrm{tr}\big(I_{T-1} \otimes e_i' \Sigma e_i\big) - \sigma^2 \mathrm{tr}\big(E_i PM^{-1}P'E_i'\big)
\end{aligned}
$$

and hence

$$
E\{\hat{v}_i'\hat{v}_i\} = (T-1)\sigma_i^2 - \sigma^2 \mathrm{tr}\big(E_i PM^{-1}P'E_i'\big). \tag{3.7}
$$

Regarding $\hat{Q}$ we note that

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

and obtain

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

and hence

$$
E\{\hat{Q}\} = \sigma^2(T - n).
$$

$\square$

The moments estimators are obtained by selecting variances σ^2 and $\sigma_i^2, i = 1, 2, \ldots, n$ such that the expected moments $E\{\hat{u}'\hat{u}\}$ and $E\{\hat{v}_i'\hat{v}_i\}, i = 1, 2, \ldots, n$ given in (3.6) and (3.7) are equalized to the estimated moments $\hat{u}'\hat{u}$ and $\hat{v}_i'\hat{v}_i, i = 1, 2, \ldots, n$. As both the expected moments and the estimated moments are functions of the variances, the moments estimators, denoted by $\hat{\sigma}^2$ and $\hat{\sigma}_i^2, i = 1, 2, \ldots, n$, respectively, are defined as a fix point of the system

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

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

$$
\begin{aligned}
\hat{v}_i'\hat{v}_i &= (T-1)\sigma_i^2 - \sigma^2\operatorname{tr}\left(E_iPM^{-1}P'E_i'\right) \\
\hat{Q} &= \sigma^2(T - n).
\end{aligned}
$$

The implementations by Schlicht (2005b, 2021) use the latter alternative and employ a gradient process to find the solution of the equation system

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

This can be written as

$$\frac{\hat{\sigma}_i^2}{\hat{\sigma}^2} = \left(\frac{\hat{v}_i'\hat{v}_i}{\hat{Q}}(T-n) - \text{tr}\left(E_i P \hat{M}^{-1} P' E_i'\right) \right) \frac{1}{T-1} \tag{3.8}$$

$$\hat{\sigma}^2 = \frac{1}{T-n}\hat{Q}. \tag{3.9}$$

Iteration starts with some variance ratios $\gamma_i = \frac{\sigma^2}{\sigma_i^2}$. This permits to determine the right-hand sides of Eqs. (3.8) and (3.9). The variance ratios at the left-hand side of (3.8) and the variance at the left hand side of (3.9) 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 $\hat{\sigma}_i^2 = \frac{\hat{\sigma}^2}{\hat{\gamma}_i}$. (If this process does not converge, another solution procedure is available that will be discussed in Sect. 3.3 below.)

4.2 Likelihood estimation

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

Using Eqs. (1.8) and (2.10)–(2.14) together with the identity $x = XZ$, the concentrated log-likelihood function for the Aitken regression (2.9) can be written as

$$\mathscr{L}\left(\sigma^2, \Sigma\right) = -\frac{1}{2}(T(\log 2 + \log \pi) + \log \det W) - \frac{1}{2}(y - XZ\beta)'W^{-1}(y - XZ\beta) \tag{3.10}$$

with

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

By maximizing (3.10) with respect to β, σ^2 and Σ, the maximum likelihood estimates for the variances are obtained and the corresponding expectation for the parameter a is given in analogy to (2.20) as

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

with a caron denoting the maximum likelihood estimates and $\check{V} = \left(I_{T-1} \otimes \check{\Sigma}\right)$.

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.

$$\log \det W = \log \det\left(PMP'\right) + (T-1)\sum_{i=1}^{n}\log \sigma_i^2$$
$$- ((T-1)n - T)\log \sigma^2 - 2\log \det\left(PP'\right). \tag{3.11}$$

Claim 4

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

Proof

Hence the result

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

is obtained. $\qquad\square$

Claim 5 *Minimizing the criterion*

$$
\begin{aligned}
\mathcal{C}_L = {}& \log \det \left(PMP' \right) + (T-1) \sum_{i=1}^n \log \sigma_i^2 - ((T-1)n - T) \log \sigma^2 \\
& + \frac{1}{\sigma^2} u'u + v'V^{-1}v
\end{aligned} \tag{3.12}
$$

is equivalent to maximizing the likelihood function (3.10).

Proof With (3.11) we have

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

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

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

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

4.3 Another representation of the moments estimators

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 (3.12). This criterion function is

$$
\mathcal{C}_M\left(\sigma^2,\ \Sigma\ \right) = \log \det M + \log \det PP' + (T - 1) \sum_{i=1}^{n} \log \sigma_i^2 - T(n - 1) \log \sigma^2
$$

$$
+ \frac{1}{\sigma^2} u'u + v'V^{-1}v.
$$

$$(3.13)$$

Claim 6 *Minimization of the criterion function* (3.13) *with respect to the disturbances u and v and the variances σ^2 and Σ yields the moments estimators as defined in* (3.3) *and* (3.4).

Proof Note that the envelope theorem together with (3.2) implies

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

$$(3.14)$$

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

$$(3.15)$$

In view of (3.2) we obtain further

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

$$(3.16)$$

By definition (2.24) we have

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

and hence

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

With this, Eq. (3.16) can be written as

$$\frac{\partial \log \det M}{\partial \sigma^2} = \mathrm{tr}\Big(\frac{1}{\sigma^2}\big(I_{Tn} - M^{-1}X'X\big)\Big)$$

$$= \frac{1}{\sigma^2}\big(\mathrm{tr}(I_{Tn}) - \mathrm{tr}(M^{-1}X'X)\big)$$

$$= \frac{Tn}{\sigma^2} - \frac{1}{\sigma^2}\mathrm{tr}(XM^{-1}X').$$

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

and we find

$$\frac{\partial \mathcal{C}_M}{\partial \sigma^2} = \frac{Tn}{\sigma^2} - \frac{1}{\sigma^2}\mathrm{tr}(XM^{-1}X') - \frac{T(n-1)}{\sigma^2} - \frac{1}{\sigma^4}\hat{u}'\hat{u} = 0 \qquad (3.17)$$

$$\frac{\partial \mathcal{C}_M}{\partial \sigma_i^2} = -\frac{\sigma^2}{\sigma_i^4}\mathrm{tr}(P'E_i'E_iPM^{-1}) + (T-1)\frac{1}{\sigma_i^2} - \frac{\sigma^2}{\sigma_i^4}\hat{v}_i'\hat{v}_i = 0 \qquad (3.18)$$

which gives

$$\hat{u}'\hat{u} = \sigma^2\big(T - \sigma^2\mathrm{tr}(XM^{-1}X')\big) \qquad (3.19)$$

$$\hat{v}_i'\hat{v}_i = (T-1)\sigma_i^2 - \sigma^2\mathrm{tr}(P'E_i'E_iPM^{-1}). \qquad (3.20)$$

These first-order conditions are equivalent to Eqs. (3.3), (3.4) that define the moments estimator. □

Johannes Ludsteck's (2004, 2018) Mathematica packages for VC proceed by minimizing the criterion function (3.13). 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.

4.4 The relationship between the likelihood and the moments estimator

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

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

$$\det PMP' \approx \det M \det (PP').$$

Proof Define the $Tn \times Tn$ matrix

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

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

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

$$= \begin{pmatrix} PMP' & T^{-\frac{1}{2}}PMZ \\ T^{-\frac{1}{2}}Z'MP' & T^{-1}Z'Z \end{pmatrix}$$

$$= \begin{pmatrix} PMP' & T^{-\frac{1}{2}}PX'XZ \\ T^{-\frac{1}{2}}Z'X'XP' & I_n \end{pmatrix}.$$

This implies

$$\det \mathbb{P}M\mathbb{P}' = \det I_n \det \left(PMP' - \frac{1}{T}PX'XZZ'X'XP' \right)$$

$$= \det \left(PMP' - \frac{1}{T}PX'xx'XP' \right)$$

$$= \det \left(P\left(M - \frac{1}{T}X'xx'X \right)P' \right)$$

$$= \det \left(P\left(X'\left(I_T - \frac{1}{T}xx' \right)X + \sigma^2 P'V^{-1}P \right)P' \right).$$

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

$$\det \mathbb{P}M\mathbb{P}' \approx \det PMP' \qquad (3.21)$$

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

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

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

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

As

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

$$\det \mathbb{P}M\mathbb{P}' = \det M \det \left(PP' \right) \qquad (3.22)$$

is obtained. Combining (3.21) and (3.22) gives the result. $\square$

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

Proof For increasing T and in view Claim 7, $\mathcal{C}_M$ tends to $\mathcal{C}_L$. □

Hence the minimization of both criteria with respect to the variances will generate in the limit the same result.[4] 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.

5 Miscellaneous notes

The following offers remarks on computation (Sect. 4.1), comments on some applications of the VC method in economics that illustrate aspects of the VC method of potential interest in other fields (Sect. 4.2). Some illustration provided by simulation studies is given (Sect. 4.3). Section 4.4 discusses the problem of artifacts. Some methodological concerns are raised in Sect. 4.5.

5.1 Notes on computation

The VC method has been embodied in some freely available software packages (Ludsteck 2004, 2018; Schlicht 2005b, 2021). Although these have been developed under the assumption that all disturbances are Gaussian, the numerical routines, briefly sketched at the end of Sects. 3.1 and 3.3, remain appropriate for the non-Gaussian case.

Schlicht and Ludsteck (2006, Sec. 11) have compared the performance of the moments estimator with that of the Kalman filter in the EViews (2005) implementation for the Gaussian case and conclude that "both estimators perform very similar—with the *caveat* that the Eviews estimates have been calculated by using the theoretical values as starting values. ...The distributions of the estimates for the weights are practically indistinguishable." Given that true variances would be unavailable in practical applications and that the the Kalman results appear to be quite sensitive to the choice of initial values, that speaks for the VC method in the case that the coefficients follow a random walk. Further, the VC method dispenses of necessity to specify initial values and offers additional descriptive features, as indicated by Claim 1 and Eq. (2.8).

5.2 Notes on applications

In spite of its so far insufficient documentation, VC has found a quite a number of applications in various settings, mainly dealing with structural change. As any of the

---

[4] A referee rightly pointed out that, in general, the convergence of functions does not necessarily imply the convergence of their maximizers. In this case, this criticism does not seem to apply, because both $\mathcal{C}_M$ and $\mathcal{C}_L$ are smoothly differentiable functions. The minima are characterized by the gradient equations (3.17), (3.18), i.e. $\frac{\partial \mathcal{C}_M}{\partial \sigma^2} = 0$, $\frac{\partial \mathcal{C}_M}{\partial \sigma_i^2} = 0$ and the corresponding equations for the likelihood criterion $\frac{\partial \mathcal{C}_L}{\partial \sigma^2} = 0$, $\frac{\partial \mathcal{C}_L}{\partial \sigma_i^2} = 0$. These conditions converge, too.

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. Yet it may be appropriate to illustrate possible uses of the VC method by means of some examples taken from my field, economics.

In the wake of the financial crisis of 2008, it has been observed that "monetary policy rules change gradually, pointing to the importance of applying a time-varying estimation framework" (Baxa et al. 2014) and that, "by applying the time-varying coefficients method ...it was clear that the past financial crisis caused the central bank to be more expansionary in its policy than usual towards financial stress" (Madsen 2012). Further, analyses of inflation targeting (IT) in "a time-varying coefficients methodology ...show a clear picture of credibility gains from the adoption of IT" (Nogueira 2009). Another application dealt with the recent decoupling of greenhouse gas emissions and gross domestic product in the wake of global warming where it has been found that "the evidence for decoupling among the richer countries gets weaker." (Cohen et al. 2017). Regarding the relationship between unemployment and economic growth, known in economics as "Okun's Law", it has been contested that the relationship has been static over time (Jalles 2018) and that, actually, "deregulation in labor and product markets and recessions have strengthened the response of unemployment to the business cycle" (Furceri et al. 2019).

Such applications suggest to me that the VC method may offer an additional useful way for dealing with linear models with coefficients that follow a random walk, and I hope that similar applications will be found in other fields.

5.3 Some illustration

To illustrate the practical workings of VC, assume a model with an intercept term a_t and a single explanatory variable x_t with coefficient b_t[5]:

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

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 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.

---

[5] The following is taken from Schlicht and Ludsteck (2006, Sec. 10).
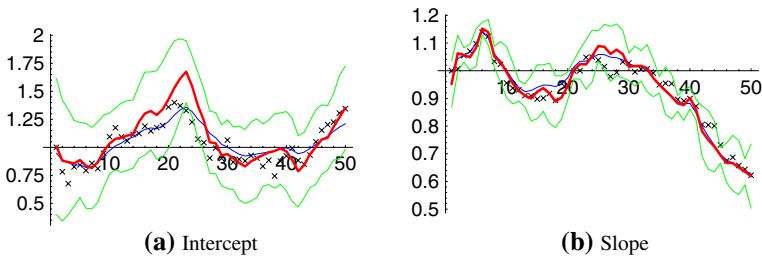
(a) Intercept   **(b)** Slope

Fig. 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 $\hat{\sigma}_b^2 = 0.0029$)

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 random vector) is optimal and cannot be improved upon, and the standard of comparison must be the estimates obtained with optimal weights, as in Fig. 1.

The distribution of the weights in the above setting is illustrated in Fig. 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 Fig. 2 illustrates that the estimates cluster around their theoretical values.

5.4 Artifacts

Suppose that the data of a particular problem have been generated by the standard linear model (1.2). If this is the case, the VC model is 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 ratios—are infinite, whereas VC implicitly assumes that the weights are finite. As the VC estimates with sufficiently large weights γ_i are indistinguishable from the OLS estimates, the VC estimation would nevertheless be approximately correct if the estimated weights are sufficiently large.[6]

As VC estimates involve 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.

---

[6] The option "keep selected coefficients constant" in Schlicht (2005b, 2021) is implemented with $\sigma_i^2 = 10^{-10}$ for those coefficients that are kept constant.
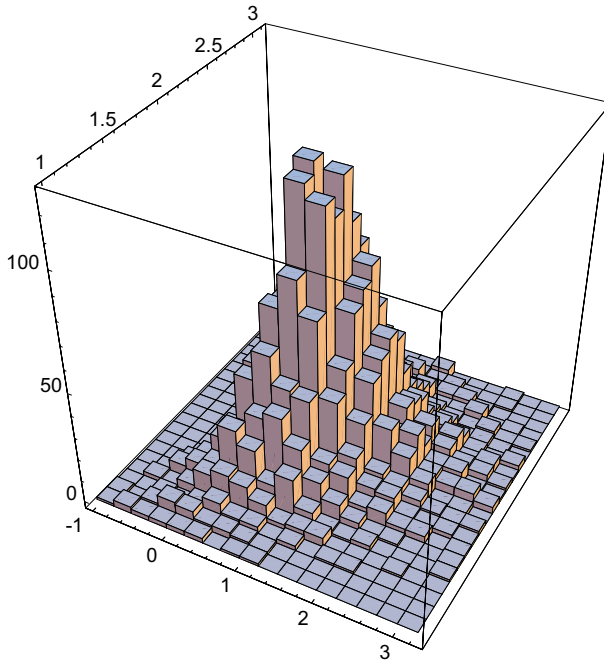
Fig. 2 Histogram of estimates for the $\log_{10}$ weights. The theoretical values are $\log_{10}\gamma_a = 1$ and $\log_{10}\gamma_b = 2$. The distribution of estimates clusters around this peak. ($T = 50$, 5000 trials)
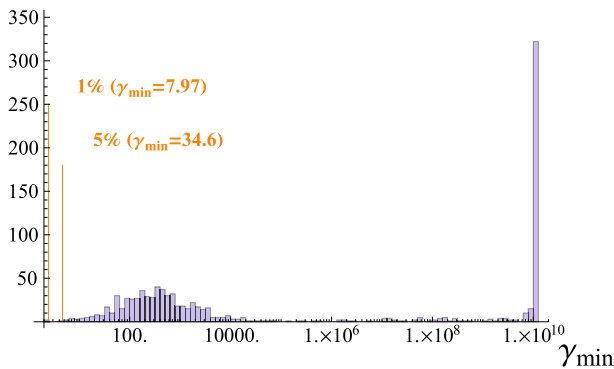

Fig. 3 Histogram of lowest weights $\gamma_{\min} = \min\{\gamma_1, \gamma_2\}$ of VC estimates for a linear model with time-invariant coefficients. ($T = 50$, 1000 trials)

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 $\sigma_u^2 = 0.1$. The histogram of the lowest estimated weights is given in Fig. 3. In 99% of the cases, the minimum weight is above 7.97, and in 95% of the cases, the minimum weight is above 34.6. The corresponding VC
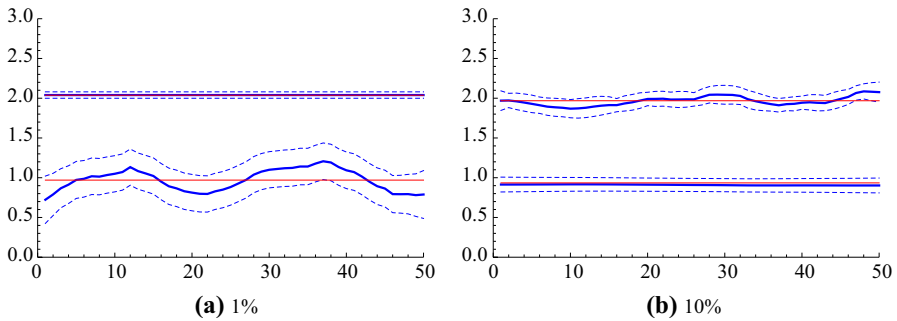
Fig. 4 VC estimates at the 1% quantile (**a**) and the 5% quantile (**b**) of Fig. 3. The dashed lines give the confidence band for ± two standard deviations. The red lines indicate the OLS estimates of the coefficients. The true coefficients are 1 and 2. As the OLS estimates fit into the confidence bands, VC does not reject the case of constant coefficients

estimates are given in Fig. 4. In the 1% case, the estimate of the time paths involve severe artifacts. In the 5% case, artifacts are still there, but in the majority of cases, VC estimates conform to OLS estimates. Further, VC does not reject the hypothesis of time-invariant parameters in 99 per cent. of the cases. This observation suggests that VC may be used to check the linear specification of a time-series model.

With higher/lower noise, the problem of artifacts becomes more/less severe.[7] Still the problem has to be kept in mind when interpreting VC results.

5.5 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

---

[7] Even in the rather ill-conditioned case of $\sigma_u^2 = 1$ VC does not reject the hypothesis that the coefficients may be time-invariant in 90 per cent. of the cases. The interested reader may explore the programming underlying Figs. 3 and 4 as well as further cases by consulting the Mathematica Notebook given in the accompanying material. Other cases of interest may be explored by running the notebook with alternative parameter settings.

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 some-what 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, 24). I think that this is, in the context of aggregate models dealing with long-run time series, a red herring, distracting from consider-ing seriously what aggregate models represent.[8]

Another reason, I submit, was the reductionist bent of economists. If a structure changes over time, this warrants explanation. Hence there was a tendency to add additional explanatory variables as 'controls' in order to explain the change. While this may be sensible in certain cases, it is unnecessary and even obfuscating if the changes brought about by such outside forces are slow and independent of the rela-tionships under study.[9] Further, the introduction of such controls seems, statistically speaking, problematic because of the following theorem that has been provided by Theo Dijkstra (1995, 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 vari-ables $x_1, x_2, \ldots, x_n$ in the standard linear model (1.2), will deliver arbitrarily prede-termined 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, Pyr-rho's problem.

Let me add another remark. The VC model (1.4), (1.5) 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 $a_{i,t+1} = \theta_i (a_{i,t} - \bar{a}_i) + v_{i,t}$. 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 descrip-tive transparency of the VC method which is, to me, a major concern—trumping more technical statistical considerations.

---

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

[9] This is formally stated by the "Moving Equilibrium Theorem" (Schlicht 1978, 1997).

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 conform to the desiderata underlying the descriptive account (Sect. 1.4). These are that the coefficients remain as time-invariant as possible and that a good fit is obtained. This makes sure that all estimated variability over time is driven by the data, rather than by another 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.

6 Conclusion

The VC method outlined in this paper addresses linear models with coefficients that are generated by a random walk. This is a rather special case, as it does not cover models with coefficients that presuppose more general stochastic processes. Yet it is a case that has received special attention in the literature, at least in economics. Focussing on this somewhat narrow class of models offers some benefits, however. Non-Gaussian disturbances can be admitted. The researcher is not required to postulate initial values that are typically unknown. Further, the time-averages of the estimated coefficients can be linked to the coefficients of an associated linear GLS (Aitken) model with constant coefficients. Further, the VC method permits an easy treatment of the case that only a subset of the coefficients varies over time while the rest of them remains constant. In all this, it allows for a more satisfactory treatment than would be possible within more general approaches, such as Kalman filtering.

Regarding further developments, and being confined to the rather narrow perspective of an economist, I would find it appropriate to conceive a way for dealing with missing observations in an appropriate way.[10] Further, moments estimators for simultaneous equations, again with a clear descriptive interpretation, appear desirable. Yet any attempt in this direction definitely exceeds my capabilities.

---

[10] Regarding the HP filter, I have proposed a method along with a software package to cope with missing data and structural breaks (Schlicht 2008, 2011). As the HP filter covers just a special case of VC—VC applied to a time series with a single explanatory variable of unity and a time-varying coefficient for this explanatory variable—a generalization of this method to VC may be feasible.

the VC method. Without the encouragement that this conveyed I would not have been motivated to write this documentation.

References

Athans, M. (1974). The importance of Kalman filtering methods for economic systems. In *Annals of economic and social measurement* (Vol. 3, No. 1, pp. 49–64). NBER Chapters. National Bureau of Economic Research, Inc. https://ideas.repec.org/h/nbr/nberch/9994.html

Baxa, J., Horváth, R., & Vašíček, B. (2014). How does monetary policy change? Evidence on inflation-targeting countries. *Macroeconomic Dynamics*, *18*(3), 593–630. https://doi.org/10.1017/S1365100512000545

Cohen, G., Jalles, J., Loungani, P., & Marto, R. (2017). Emissions and growth: Trends and cycles in a globalized world. IMF Working Paper (WP17/191). https://www.imf.org/~/media/Files/Publications/WP/2017/wp17191.ashx

Cooley, T. F. & Prescott, E. C. (1973). An adaptive regression model. *International Economic Review, 14*(2), 364–371. https://ideas.repec.org/a/ier/iecrev/v14y1973i2p364-71.html

Dijkstra, T. (1995). Pyrrho's lemma, or have it your way. *Metrika: International Journal for Theoretical and Applied Statistics 42*(1), 119–125. https://ideas.repec.org/a/spr/metrik/v42y1995i1p119-125.html

EViews (2005). Eviews 5.1 standard edition. http://eviews.com/home.html

Fan, J., & Zhang, W. (1999). Statistical estimation in varying coefficient models. *Annals of Statistics*, *27*, 1491–1518. https://doi.org/10.1214/aos/1017939139

Furceri, D., Jalles, J., & Loungani, P. (2019). On the determinants of the Okun's Law: New evidence from time-varying estimates. *Comparative Economic Studies*. https://doi.org/10.1057/s41294-019-00111-1

Gilchrist, W. G. (1967). Methods of estimation involving discounting. *Journal of the Royal Statistical Society. Series B (Methodological), 29*(2), 355–369. http://www.jstor.org/stable/2984595

Green, P. & Silverman, B. (2000). Nonparametric regression and generalized linear models. A roughness penalty approach. Boca Raton: Chapman & Hall. https://www.amazon.de/Nonparametric-Regression-Generalized-Linear-Models/dp/0412300400

Hastie, T. & Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society. Series B (Methodological), 55*(4), 757–796. http://www.jstor.org/stable/2345993

Henderson, R. (1924). A new method of graduation. *Transactions of the Actuarial Society of America*, *25*, 29–40.

Hodrick, R. J., & Prescott, E. C. (1997). Postwar U.S. business cycles: An empirical investigation. *Journal of Money, Credit and Banking, 29*(1), 1–16. https://ideas.repec.org/a/mcb/jmoncb/v29y1997i1p1-16.html

Jalles, J. T. (2018). On the time-varying relationship between unemployment and output: What shapes it? *Scottish Journal of Political Economy*, *66*, 605–630.

Keynes, J. M. (1939). Professor Tinbergen's method. *Economic Journal, 49*(195), 558–568. https://www.jstor.org/stable/2224838

Keynes, J. M. (1973). The general theory and after, part II: Defense and development. The collected works of John Maynard Keynes (Vol. XIV). London: Macmillan.

Leser, C. E. V. (1961). A simple method of trend construction. *Journal of the Royal Statistical Society. Series B (Methodological), 23*, 91–107. http://www.jstor.org/stable/2983845

Leser, C. E. V. (1963). Estimation of quasi-linear trend and seasonal variation. *Journal of the American Statistical Association, 58*, 1033–1043. http://www.jstor.org/stable/2283329

Lucas, R. (1976). The econometric policy evaluation: a critique. In K. Brunner & A. H. Meltzer (Eds.), Phillips curve and labor markets (pp. 19–46). Amsterdam: North Holland. https://EconPapers.repec.org/RePEc:eee:crcspp:v:1:y:1976:i::p:19-46

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

Ludsteck, J. (2018). VC packages for estimating time-varying coefficients with mathematica, pp. 8–11. https://epub.ub.uni-muenchen.de/59479/

Madsen, M. W. (2012). Does financial stress have an impact on monetary policy? An econometric analysis using Norwegian data, Ph.D. thesis. Department of Economics, University of Oslo. https://www.duo.uio.no/bitstream/handle/10852/17126/Madsen_Michael_Masteroppgave.pdf

Nogueira, R. P. (2009). Testing credibility with time-varying coefficients. *Applied Economics Letters*, *16*(18), 1813–1817. https://doi.org/10.1080/13504850701719611

Rouhiainen, J. (1978). The problem of changing parameters in demand analysis and forecasting. *European Review of Agricultural Economics, 5*(3–4), 349–359. https://ideas.repec.org/a/oup/erevae/v5y1978i3-4p349-359..html

Schlicht, E. (1973). Forcasting Markov chains. A theoretical foundation for exponential smoothing, Working paper B 13. Department of Economics, University of Regensburg. http://www.semverteilung.vwl.uni-muenchen.de/mitarbeiter/es/paper/schlicht-exponential_smoothing.pdf

Schlicht, E. (1977). Grundlagen der ökonomischen Analyse. Reinbek: Rowohlt. https://epub.ub.uni-muenchen.de/25821/

Schlicht, E. (1978). Die Methode der Gleichgewichtsbewegung als Approximationsverfahren. Chapters in Economics (pp. 293–305). Berlin: Duncker und Humblot. https://ideas.repec.org/h/lmu/muench/3149.html

Schlicht, E. (1981). A seasonal adjustment principle and a seasonal adjustment method derived from this principle. *Journal of the American Statistical Association, 76*(374), 374–378. Paper presented at the Econometric Society European Meeting Helsinki 1976. http://www.semverteilung.vwl.uni-muenchen.de/mitarbeiter/es/paper/schlicht-seasonal_adjustment.pdf

Schlicht, E. (1984). Seasonal adjustment in a stochastic model. *Statistical Papers, 25*, 1–12. http://www.semverteilung.vwl.uni-muenchen.de/mitarbeiter/es/paper/schlicht_seasonal-adjustment-in-a-stochastic-model.pdf

Schlicht, E. (1985). Isolation and aggregation in economics, annotated electronic reprint 2017 Edition. Berlin: Springer. https://ideas.repec.org/p/lmu/muenec/38821.html

Schlicht, E. (1989). Variance estimation in a random coefficients model, Munich Discussion Paper. Paper presented at the Econometric Society European Meeting, Munich. https://epub.ub.uni-muenchen.de/59143/

Schlicht, E. (1990). Local aggregation in a dynamic setting. *Journal of Economics, 51*(3), 287–305. https://ideas.repec.org/a/kap/jeczfn/v51y1990i3p287-305.html

Schlicht, E. (1997). The moving equilibrium theorem again. *Economic Modelling, 14*(2), 271–278. https://ideas.repec.org/a/eee/ecmode/v14y1997i2p271-278.html

Schlicht, E. (2005a). Estimating the smoothing parameter in the so-called Hodrick–Prescott filter. *Journal of the Japan Statistical Society*, *35*, 99–119. https://doi.org/10.14490/jjss.35.99

Schlicht, E. (2005b). VCC—A program for estimating time-varying coefficients. Console Version With Source Code in C. http://epub.ub.uni-muenchen.de/archive/00000719/

Schlicht, E. (2008). Trend extraction with missing observations and structural breaks. *Journal of the Japan Statistical Association*. https://www.jstage.jst.go.jp/article/jjss/38/2/38_2_285/_pdf

Schlicht, E. (2011). Mend-A mathematica package for mending time series with missing observations and structural breaks. https://epub.ub.uni-muenchen.de/12227/

Schlicht, E. (2021). VC—A program for estimating time-varying coefficients. Version, 6. https://doi.org/10.5282/ubm/epub.684

Schlicht, E. & Ludsteck, J. (2006). Variance estimation in a random coefficients model. Munich Discussion Paper (2006–2012). https://ideas.repec.org/p/lmu/muenec/904.html

Schlicht, E. & Pauly, R. (1983). Descriptive seasonal adjustment by minimizing perturbations. *Empirica*. http://www.semverteilung.vwl.uni-muenchen.de/mitarbeiter/es/paper/schlicht-pauly-perturbations.pdf

Tinbergen, J. (1940). On a method of statistical business-cycle research. A reply. *Economic Journal* *50*(197), 141–154. http://www.jstor.org/stable/2225763

Whittaker, E. T. (1923). On a new method of graduation. *Proceedings of the Edinburgh Mathematical Society*, *41*, 63–75.