



Linear models with time-varying parameters: a comparison of different approaches

Riccardo “Jack” Lucchetti¹ · Francesco Valentini¹

Received: 26 October 2023 / Accepted: 20 December 2023
© The Author(s) 2024

Abstract

Estimation of linear models with time-varying parameters can be accomplished in a variety of ways, each making different assumptions, with varying degrees of accuracy and computational complexity. In this paper, we compare different `gretl` packages by means of simulated and real data focusing on both statistical and computational aspects. Our findings show that all the estimators provide similar results under ideal conditions, but the practitioner’s choice could be far from obvious.

Keywords Time-varying parameters · Linear model · Gretl

1 The case for time-varying parameters in econometrics

In a discipline such as econometrics, where empirical regularities are the outcome of social phenomena and institutions, rather than physical quantities, it is perfectly natural to imagine that the main characteristics of a statistical model may be subject to some change over time, either smooth or abrupt. Therefore, econometric models able to capture potential variability over time of the data generating process have been of considerable interest for both theoretical and practical applications.

Moreover, in economics the theoretical models are typically subject to a *ceteris-paribus* clause: all influences not explicitly modeled are, for theoretical purposes, assumed as being constant, although these other influences are conceived to actually change over time. Such changes, if not too pronounced, may not destroy the conclusions reached in the theoretical *ceteris-paribus* model, but will go along

Riccardo “Jack” Lucchetti and Francesco Valentini contributed equally to this work.

✉ Francesco Valentini
f.valentini@staff.univpm.it

Riccardo “Jack” Lucchetti
r.lucchetti@staff.univpm.it

¹ Dipartimento di Scienze Economiche e Sociali, Università Politecnica delle Marche, Piazzale R. Martelli 8, 60121 Ancona, Italy

with parameter changes. In other words, quantitative economic relationships typically connect indices involving prices and quantities with ever changing weighting schemes, all affected by technological and social changes. This makes it highly implausible to exclude the possibility of time-varying coefficients a priori. As Keynes put it: unlike in the case of a the typical natural science, the subject of economics “is, in too many respects, not homogeneous through time.” (see Johnson and Moggridge 1987).

1.1 Some approaches to time-varying parameters estimation

The econometric literature on time-varying parameters is too large to be effectively surveyed here, and may well be the object of an autonomous paper (see *eg* Fan and Zhang 1999). In this section, we will only briefly survey the main contributions that are instrumental in motivating the evidence we present in this paper.

One line of research has concerned situations in which one assumes that several regimes exist, and during each regime parameters are stable. This idea provides the foundation not only for the classic work by Chow (1960), but also for more recent contributions such as Bai and Perron (2003) or Markov-switching models à la Hamilton (1989). Alternatively, one could imagine that parameter change is something that happens gradually over time. In some models, such as Smooth Transition Models (Chan and Tong 1986; Teräsvirta 1994), the change in the Data Generating Process is modeled via a small set of hyper-parameters. In others, the path through time of parameters is described with different, more agnostic modeling choices. Collectively, these models go under the customary name of *time-varying-parameters* models (see *eg* Cooley and Prescott 1976; Harvey 1990); these are the object of our interest.

Although the typical approach relies on the Kalman filter (KF) apparatus (Athans 1974; Harvey 1990), several other methods have been developed. Among others, we focus on the Flexible Least Squares (FLS) proposed by Kalaba and Tesfatsion (1989, 1990) and more recent contributions such as moments estimators (VC) put forward by Schlicht (2021) and semi-parametric estimators relying on kernel-based (KB) inference (Giraitis et al. 2014, 2018) which has been extended to Instrumental Variables (IV) models (Giraitis et al. 2021).

In the following we compare the finite sample performance of four methods, namely KF, VC, KB and FLS, on simulated and real data. The former experiment is reported in Sect. 3. As for their usage on real data, Sect. 4 illustrates a simple application to Okun’s law.

We find that the four methods considered behave (from a qualitative point of view) very similarly on simulated data, but real-life applications pose several challenges for practitioners, such as implementation difficulties, numerical instability and computational complexity that may hamper the viability of the aforementioned methods. In this work, we aim at investigating these issues so as to provide a set of useful guidelines for practitioners.

We perform our experiments in `gretl` since most of the techniques at issue are available as built-in functions or user-written packages. State space techniques can be implemented in `gretl` by means of the native support for such models. The VC estimator (Schlicht 2021) is implemented in the TVC package which, along with the moment estimator, also allow users to estimate parameters by Maximum Likelihood under the assumption of Gaussian innovations. Further, the package `ketvals` provides users a set of functions aimed at performing non-parametric least squares estimation for linear models with time varying parameters with or without (KB) instrumental variables, as proposed by Giraitis et al. (2021). Along with parameter estimates and their standard errors, `ketvals` provides users with several functions aimed at supporting the visualization of results. The FLS package provides Kalaba and Tesfatsion (1989)'s estimator with a few extensions considered in Lütkepohl and Herwartz (1996).

2 Methods for estimating time-varying linear models

The standard linear model with time-varying parameters is given by the following equation,

$$y_t = \mathbf{x}'_t \boldsymbol{\beta}_t + \varepsilon_t \quad \text{with } t = 1, \dots, T, \tag{1}$$

where y_t is the dependent variable, \mathbf{x}_t is a $k \times 1$ vector of regressors, the conformable vector of parameters $\boldsymbol{\beta}_t$ is assumed to vary over time and where ε_t denote shocks, which are assumed to be homoskedastic, uncorrelated both serially and with the regressors.

2.1 Flexible least squares

The FLS (Flexible Least Squares) method was proposed by Kalaba and Tesfatsion (1989) and can be described as follows: given the equations

$$e_t = y_t - \mathbf{x}'_t \boldsymbol{\beta}_t \tag{2}$$

$$\mathbf{u}_t = \Delta \boldsymbol{\beta}_t = \boldsymbol{\beta}_t - \boldsymbol{\beta}_{t-1} \tag{3}$$

the FLS estimator is based on minimizing the criterion function

$$C(\mu, T) = \sum_{t=1}^T e_t^2 + \mu \sum_{t=2}^T \mathbf{u}'_t \mathbf{u}_t \tag{4}$$

where μ is a user specified parameter such that:

- $\mu \rightarrow 0$: large $V(\hat{\boldsymbol{\beta}}_t)$, small $V(\hat{\varepsilon}_t)$
- $\mu \rightarrow \infty$: $V(\hat{\boldsymbol{\beta}}_t) \rightarrow 0$ and $\hat{\boldsymbol{\beta}}_t \rightarrow \hat{\boldsymbol{\beta}}_{OLS}$.

Kalaba and Tesfatsion (1990) refer to the parameter μ as a “Lagrange multiplier”; it can be thought of a “smoothing parameter”, in that higher values produce smoother series of coefficients. Estimation is performed via an algorithm described in detail in Kalaba and Tesfatsion (1989), Sect. 5, which is usually very fast and efficient.

Note that this approach is purely descriptive, in that no distributional assumptions are made on any of the magnitudes in Eqs. (2)–(3), and therefore standard errors for the estimated coefficients are not available. The model above was later extended by Lütkepohl and Herwartz (1996) in various directions. Of all the extensions proposed there, we only consider a slight generalization of (4) given by

$$C(\mu, T) = \sum_{t=1}^T e_t^2 + \mu \sum_{t=2}^T \mathbf{u}_t' D \mathbf{u}_t \quad (5)$$

where D is a symmetric, positive definite metric matrix. For normalization purposes, we require the trace of D to be equal to k . Therefore, the original FLS model corresponds to $D = I$. A rather natural alternative would be a scalar multiple of $X'X$, which would take care of different unit of measurement of the regressors and correlation between them; we call this the “normalized” metric and it will be denoted as FLSN. In our implementation, the scalar equals $k/\text{tr}(X'X)$ to satisfy the normalization requirement above.

The reader will no doubt notice that the FLS estimator is very close in spirit to the Hodrick–Prescott filter, with the exception that the penalty term considers first differences instead of second differences. Experimentation suggests that the μ parameter in the FLS filter is roughly equivalent to the square root of that for the HP filter.

What we call the FLS statistic $\hat{\beta}_t$ is in fact the one described in Kalaba and Tesfatsion (1989), Sect. 5.4 and described as the “smoothed” estimator, as opposed to “filtered”. Note that “smoothing”, in this context, has nothing to do with the value of μ , but is a term borrowed from the state-space literature and indicates that the $\hat{\beta}_t$ estimator is based on the whole sample. Conversely, the “filtered” estimate of β_t , that we do not consider in this paper, is based on the data from 1 to t only.

2.2 The VC estimator

The VC estimator (Schlicht 2021) can be thought of as a generalization of the FLS estimator, and is implemented in the TVC package. Starting from Eqs. (2) and (3), the following assumptions on the moments of e_t and \mathbf{u}_t are made:

$$E(e_t) = 0 \quad \text{and} \quad E(e_t^2) = \sigma^2 \quad (6)$$

$$E(\mathbf{u}_t) = \mathbf{0} \quad \text{and} \quad E(\mathbf{u}_t \mathbf{u}_t') = \text{diag}(\phi_1^2, \dots, \phi_K^2) \quad (7)$$

In other words, the VC estimator is derived under the assumption that the regression coefficients evolve through time as orthogonal random walk processes. On the other hand, no distributional assumptions are used, and the VC estimator has a natural

interpretation as a method-of-moments estimator. However, it can be proven that the VC estimator approaches the ML estimator under normality for $T \rightarrow \infty$.

The VC estimator minimizes the objective function (5), where

$$D = \text{diag}(\gamma_1, \dots, \gamma_K)$$

with $\gamma_k = \hat{\sigma}^2 / \hat{\phi}_k^2$. However, a subset of the coefficients can be kept fixed by setting the corresponding ϕ_i^2 to 0. Minimization is performed by numerical optimization on the γ_k parameters.

In the TVC `gretl` package, this is accomplished by minimizing a criterion function which involves the inversion of a square matrix M of dimension $T \cdot k$. As is well known, the complexity of the matrix inversion operation is $O(s^3)$, where s is the size of the matrix; therefore, Schlicht’s algorithm may become rather slow if either T or k are large. We conjecture that this aspect can be improved by taking into account the particular structure of M , and research is currently under way on this aspect.

2.3 State-space modeling

State-space models provide a very natural way of representing models with time-varying parameters, where the law of motion of the vector β_t may be specified in a very flexible way. Moreover, state space modeling is a standard tool for the analysis of time series, so that the theoretical apparatus is well-known and the array of available software is formidable.¹

In general, we may write the state-space representation of the time-varying parameters of a linear model as

$$\begin{aligned} \text{Measurement Equation: } y_t &= \mathbf{x}'_t Z \boldsymbol{\alpha}_t + e_t \\ \text{Transition Equation: } \boldsymbol{\alpha}_t &= \bar{\boldsymbol{\alpha}} + \mathbf{T} \boldsymbol{\alpha}_{t-1} + \mathbf{u}_t; \end{aligned}$$

as general as this formulation is, it remains a fairly standard state-space model.

This formulation nests several special cases that have been considered in the literature: for example, the following setup

$$\begin{aligned} Z &= [I \quad 0] \\ \boldsymbol{\alpha}_t &= \begin{bmatrix} \boldsymbol{\beta}_t \\ \boldsymbol{\mu}_t \end{bmatrix} \\ \bar{\boldsymbol{\alpha}} &= \mathbf{0} \\ \mathbf{T} &= \begin{bmatrix} I & I \\ 0 & I \end{bmatrix} \end{aligned}$$

corresponds to the case when the time path of parameters is assumed to be described by a “local linear trend” model (see Harvey 1990, for a taxonomy of “structural time

¹ The special issue of the *Journal of Statistical Software* (Commandeur et al. 2011) provides a comprehensive review, although a bit dated. For example, the description of the `gretl` implementation given in Lucchetti (2011) is completely out of date.

series models"). In this kind of model, β_t is an $I(2)$ process, and therefore typically exhibits very smooth time paths.

Conversely, one could assume that β_t is a stationary vector process: one of the alternatives arises when

$$\begin{aligned} Z &= I \\ \alpha_t &= \beta_t \\ \bar{\alpha} &= (I - \Phi)\bar{\beta} \\ \mathbf{T} &= \Phi, \end{aligned}$$

with Φ is a matrix whose eigenvalues are smaller than 1 in absolute value. Here, the elements of β_t are stationary $VAR(1)$ processes, fluctuating around a long-term mean equal to $\bar{\beta}$.

In this paper, however, we will concentrate on the choice that is by far the most popular among practitioners,² which is also compatible with the VC model described earlier, in which the β_t vector is assumed to be a random walk process:

$$\begin{aligned} \text{Measurement Equation: } y_t &= \mathbf{x}'_t \beta_t + e_t \\ \text{Transition Equation: } \beta_t &= \beta_{t-1} + \mathbf{u}_t \end{aligned}$$

ML estimation is fairly standard under the customary normality assumptions and independence assumptions:

$$\begin{aligned} e_t &\sim N(0, \sigma_e^2) \\ E(\mathbf{u}_t) &= \mathbf{0} \\ E(\mathbf{u}_t \mathbf{u}'_t) &= \Sigma_u \\ E(e_t \mathbf{u}'_t) &= \mathbf{0}'. \end{aligned}$$

by using standard numerical optimization methods as explained, for example, in Pelagatti (2015), Sect. 5.4. In this article, we maximize the log-likelihood by means of the limited-memory variant of the BFGS method (LBFGS) by Byrd et al. (1995), which proved to be very effective.

2.4 Kernel-based method

Kernel-based inference provides a powerful generalization of the traditional rolling-window technique: the approach we use here is the one developed in Giraitis et al. (2014) and later developed in a number of derived papers such as Giraitis et al. (2021), where an extension to instrumental-variable estimation is developed.

The estimator we use here is the time-varying OLS estimator (TV-OLS, hereafter) defined as

² Among the many examples, a few recent ones are Cimadomo and D'Agostino (2016), Cai and Wu (2021) or Bjørnland and Thorsrud (2019).

$$\hat{\beta}_t = \left(\sum_{j=1}^T b_{H,|j-t|} \mathbf{x}_j \mathbf{x}'_j \right)^{-1} \left(\sum_{j=1}^T b_{H,|j-t|} \mathbf{x}_j y'_j \right), \tag{8}$$

where $b_{H,|j-t|}$ denotes a kernel weight and $H = T^h$ is the bandwidth given by a user-specified parameter h . The kernel weighting function could be any of the many ones considered in ordinary nonparametric estimation (Gaussian, Epanechnikov, etc.). The rolling-window approach is but a special case, where the kernel function is rectangular.

In a way, this estimator is similar in spirit to the FLS estimator, in that no assumption is made on the law of motion for the β_t coefficients, which are only assumed to vary “smoothly” through time, either as smoothly varying deterministic functions or smoothly varying persistent stochastic processes (see Assumptions 2 and 3 in Giraitis et al. (2021), respectively). On the other hand, suitable regularity conditions on the distribution of the disturbance term ε_t in Eq. (1) make it possible to compute standard errors for the estimated coefficients and perform various inference procedures, such as hypothesis tests.

The choice for the kernel function and the associated bandwidth is up to the user and is, in many cases, highly subjective: the main elements to consider are whether to choose a finite- or infinite-support kernel function and the choice of the smoothing parameter h . The latter choice can be semi-automated by employing a modified version of the AIC criterion, due to Cai (2007). We will denote the KB estimator obtained with the automatic bandwidth selection as KBA. As shown in Lucchetti and Valentini (2023), however, results can be in some cases highly sensitive to these choices, and in some cases it is advisable to complement statistical procedures with qualitative economic reasoning.

Our implementation of choice for this technique is the gretl `ketvals` package, in which the estimator is computed by means of kernel-weighted least squares technique and poses no particular numerical challenge (no numerical optimization is needed), so estimation is usually very fast and stable.

3 Simulation experiments

This section analyses the properties and the computational costs³ of the estimators at issue by using simulated data. The estimators we consider are:

- KF: State space model with random walk coefficients (see Sect. 2.3).
- KB: Kernel-based (see Sect. 2.4) with a-priori fixed bandwidth ($h = 0.66$) and Gaussian kernel.

³ The experiments in this section are run on a Debian Linux system with n.1 Intel(R) Core(TM) i3-4150 CPU 3.50 GHz and 8 GB of RAM. The Gretl version is 2023b.

Table 1 Sources of arbitrariness

Estimator	Initial conditions	Hyper-parameters
KF	$\alpha_t = \hat{\beta}_{KB}$ $Z = I$ $T = I$ $V(u_t) = I/100$	–
KB	–	Gaussian kernel, $h = 0.66$ (bandwidth)
KBA	–	Gaussian kernel
VC	–	–
FLS	–	$\mu = 100$ (smoothing)

KBA: Kernel-based with automatic (modified AIC) bandwidth, as proposed by Lucchetti and Valentini (2023) and Gaussian kernel.

VC: Schlicht's VC estimator (see Sect. 2.2).

FLS: Plain FLS estimator with $\mu = 100$ (see Sect. 2.1).

FLSN: Normalized version of the FLS estimator.

First, we propose a simple exercise where we generate one realization only and show how estimated coefficients mimic the path of the true parameters in a single sample. The purpose of this experiment is to give the reader an intuitive way to visualize the differences between the methods we compare.

Next, we perform a Monte Carlo experiment in order to assess the finite sample performance and the computational burden of the aforementioned techniques. Finally, we test the sensitivity of the CPU-time taken by the different routines to different sample sizes and number of regressors.

We want to stress that the estimators above requires the user to make specific choices. Table 1 summarizes the potential sources of arbitrariness induced by user specified options. Hence, the following results have to be considered conditional on such discretion.

3.1 Three simple experiments

We now show three simple simulation experiments, with different data generating processes for the time-varying coefficients. In all the cases, we consider, for $t = 1, \dots, T = 128$, the model

$$y_t = x_{1,t}\beta_{1,t} + x_{2,t}\beta_{2,t} + \varepsilon_t \quad (9)$$

where we include $J = 2$ covariates $x_{j,t}$, with $j = \{1, 2\}$, following a standard Gaussian distribution whose correlation is governed by a parameter $\rho = \{0, 0.5\}$.

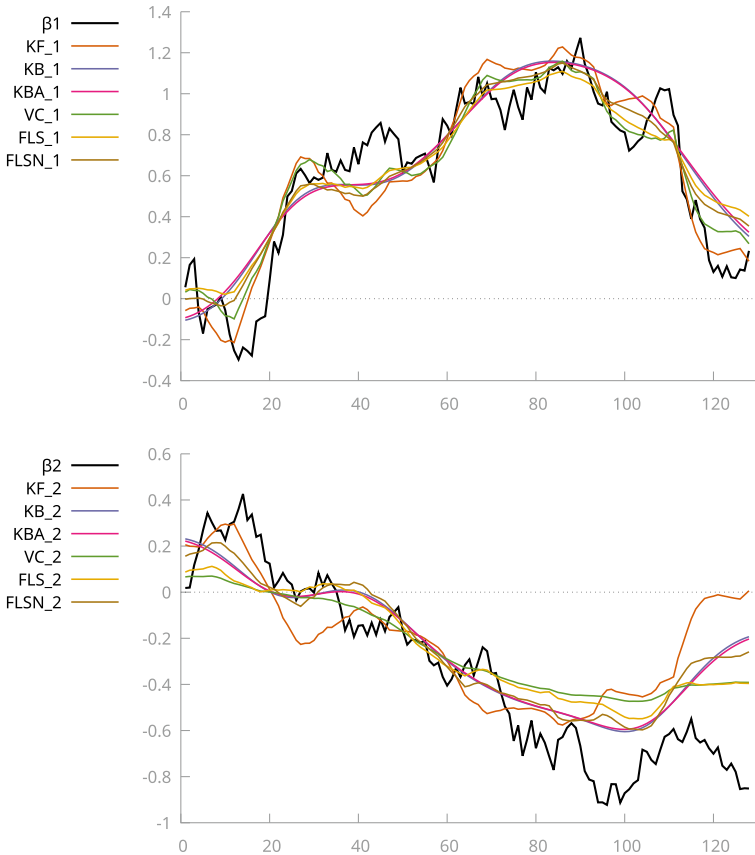


Fig. 1 Simulation with random walk parameters. Note: β_1 and β_2 denote the true values of the model parameters. Covariates are generated with $\rho = 0.5$

As a first experiment, we generate time-varying parameters which conform to the assumptions of the VC and KF estimators. Therefore we initially generate model parameters $\beta_{j,t}$ as cumulated normally distributed draws with variance 0.01 so as to produce a time path of coefficient in line the assumption, common to all the estimator at issue, of “smooth” or “well behaved” evolution over time. The error term ε_t is also drawn from a zero mean normal distribution with standard deviation equal to 0.7.

Figure 1 reports the time paths of estimated coefficients along with the true values of the model parameters in a typical run. In general, all the estimators perform similarly and float around the true $\beta_{j,t}$, being able to reproduce the path of the time-varying coefficients for most of the sample. All the methods also exhibit the problem of overestimating $\beta_{2,t}$ in the final part of the sample.

It is worth stressing that the KB and FLS estimators are quite sensitive to the choice of the smoothing parameters which are fixed at $h = 0.66$ and $\mu = 100$,

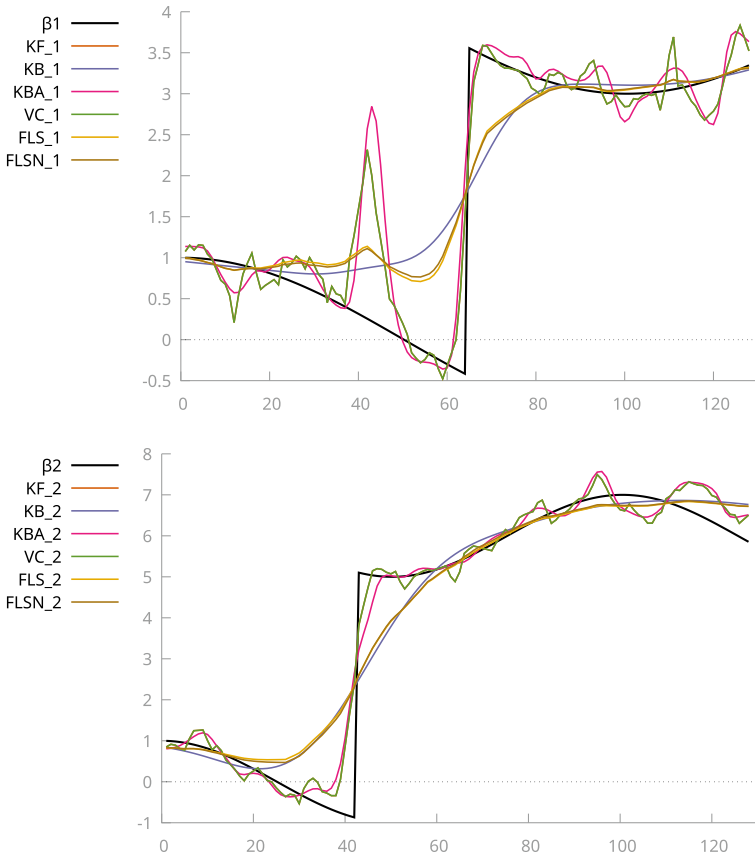


Fig. 2 Simulation with jumps in the coefficients. Notes: β_1 and β_2 denote the true values of the model parameters generated according to Eq. (10). Covariates are generated with $\rho = 0$

respectively. In particular, KB and KBA seem to produce very smooth time paths for the coefficients, while FLS and FLSN are more erratic and close to the ones obtained with methods relying on parametric assumptions such as KF and VC.

In our second experiment, we consider a data generating process that violates the “smoothness” of the coefficients time path, and is similar in spirit to the numerical experiments carried out in Kalaba and Tesfatsion (1990) and Giraitis et al. (2014).

In this case, we generate the coefficients as

$$\beta_{j,t} = \cos((t \cdot j)/32) + (2 + 2 \cdot j) \cdot \mathbf{1} \left\{ t > \frac{T}{j+1} \right\} \tag{10}$$

so that the first term gives coefficients with a sinusoidal path and the second term is a “jump” that occurs at a certain point in time. Again, we produce a one-sample experiment, whose results are reported in Fig. 2. Note that the results for KF and VC are so similar that they are visually indistinguishable.

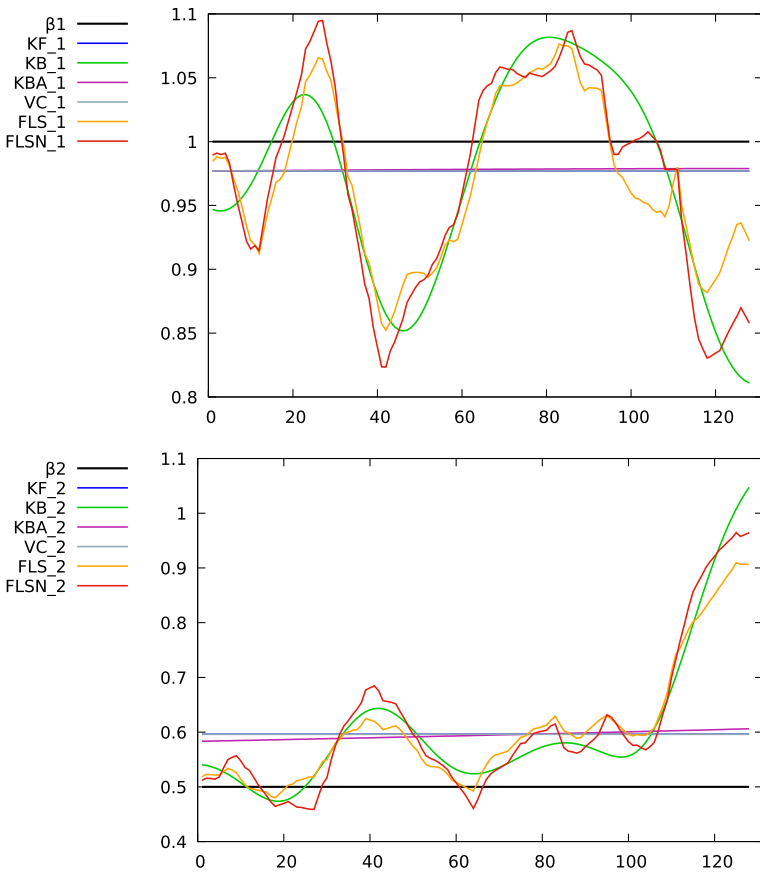


Fig. 3 Simulation with time constant coefficients. Notes: β_1 and β_2 denote the true values of the model parameters. Covariates are generated with $\rho = 0$

As we can see, the response to a violation of “smooth” changes in β is not uniform. KBA, KF and VC are able to rapidly adjust and capture the abrupt shifts in the time path of the coefficients while FLS and KB with a fixed bandwidth display some stickiness and adjust slowly since their volatility is limited by the fixed smoothness parameters μ and h , respectively. Nevertheless, the price that one pays for the greater responsiveness of KBA, KF and VC is that they may pick some spurious movement in the estimated coefficient that is just not there in the true one: consider, for example the spurious peak occurring in $\beta_{1,t}$ around observation 40.

Finally, we consider the special case of a data generating process with time-constant parameters; it is interesting to consider this case because in real-life cases the researcher does not know if parameters are time-changing. Of course, several test for structural stability are available, such as for example the CUSUM test, but that is beyond the point we are making here, that is evaluating the tendency of the various methods we compare to generate parameter volatility as a spurious artifact.

Therefore, we set $\beta_{1,t} = 1$ and $\beta_{2,t} = 0.5$ for $t = 1, \dots, 128$. The estimation results are reported in Fig. 3.

In line with previous results, the response is not homogeneous. Again, KBA, KF and VC produce similar results and the related estimated coefficients are essentially overlapping and display almost no time variability around the corresponding (time-constant) least squares estimates. Conversely, KB, FLS and FLSN show a much larger variability even though they also appear to gravitate towards the simple OLS estimates. In light of the previous exercises, the result is quite obvious. Here, the fixed smoothing parameters for KB and FLS force the estimated coefficient to float excessively so that the time-variability can be reduced by the tuning hyper-parameters, accordingly.

These simple exercises give us some first insights about the behaviour of the estimators under consideration. In the next section we rely on a Monte Carlo experiment in order to properly analyse the finite sample performance of the estimators.

3.2 Monte Carlo experiments

We implement our Monte Carlo experiment by generating data via the two cases examined in Sect. 3.1, first with random-walk coefficients and then with jumps. We considered four scenarios by combining $T = \{64, 128\}$ and $\rho = \{0, 0.5\}$ and for each of them we perform $R = 1000$ replications from which we collect the median deviation (MD) and the median absolute deviation (MAD) of the estimators,⁴ defined as

$$\text{MD} = \frac{1}{R} \sum_{r=1}^{1000} \text{med}_{t=1, \dots, T}(\hat{\beta}_t^r - \beta_t^r)$$

$$\text{MAD} = \frac{1}{R} \sum_{r=1}^{1000} \text{med}_{t=1, \dots, T}|\hat{\beta}_t^r - \beta_t^r|$$

where $\hat{\beta}$ denotes a generic estimator and the superscript r refers to the r -th Monte Carlo sample.

Table 2 reports the simulation results for the random-walk case. For each of the six methods, the table includes the MD and MAD indices associated to the two parameters along with the average CPU time required to get estimates.

The results corroborate the visual impressions described in Sect. 3.1, that the six estimators behave similarly. All of them are well-anchored to the true value of the coefficients in the sense that the MD index is always close to zero, meaning that none of the techniques produce systematically divergent estimates. Moreover, the MAD indices (which are almost identical across the estimators) show that the median deviations from the true values are modest in size and tend to shrink as the sample size grows (comparing the $T = 64$ and $T = 128$ cases).

On the other hand, CPU times are markedly different. Of course, estimators such as KF and VC, which require numerical optimization, are the slowest, while

⁴ We also considered mean deviation and mean absolute deviation which are not reported here since the related results are equivalent to those obtained with the median.

Table 2 Simulation results: random walk parameters

T		$\beta_{1,t}$		$\beta_{2,t}$		CPU time (sec)
		MD	MAD	MD	MAD	
$\rho = 0$						
64	KF	- 0.0007	0.1164	- 0.0015	0.1229	0.325
	KB	- 0.0005	0.1134	- 0.0009	0.1166	0.001
	KBA	- 0.0010	0.1143	- 0.0012	0.1203	0.030
	VC	- 0.0010	0.1152	- 0.0025	0.1215	0.276
	FLS	- 0.0008	0.1117	- 0.0003	0.1168	0.001
	FLSN	- 0.0008	0.1121	- 0.0006	0.1172	0.001
128	KF	0.0019	0.1051	0.0004	0.1048	0.644
	KB	0.0009	0.1145	0.0003	0.1116	0.002
	KBA	0.0020	0.1068	0.0000	0.1060	0.132
	VC	- 0.0013	0.1057	0.0004	0.1033	1.642
	FLS	0.0033	0.1062	0.0007	0.1037	0.002
	FLSN	0.0033	0.1064	0.0005	0.1040	0.001
$\rho = 0.5$						
64	KF	0.0023	0.1346	- 0.0049	0.1336	0.286
	KB	0.0017	0.1290	- 0.0042	0.1270	0.001
	KBA	0.0004	0.1298	- 0.0039	0.1311	0.030
	VC	0.0024	0.1293	- 0.0030	0.1292	0.275
	FLS	0.0020	0.1240	- 0.0039	0.1235	0.001
	FLSN	0.0024	0.1282	- 0.0033	0.1282	0.001
128	KF	0.0007	0.1198	- 0.0022	0.1149	0.588
	KB	0.0005	0.1235	- 0.0002	0.1195	0.002
	KBA	0.0004	0.1212	- 0.0004	0.1180	0.136
	VC	0.0008	0.1168	0.0013	0.1126	1.786
	FLS	0.0010	0.1172	- 0.0006	0.1129	0.002
	FLSN	0.0007	0.1212	- 0.0007	0.1174	0.001

the KB and FLS routines are the fastest and require a small fraction of a second. The KBA estimator has intermediate performance, since numerical optimization is limited to one parameter only (the bandwidth).

The KF and VC estimators display comparable CPU time with $T = 64$, about 0.3 s on average. In the second scenario with $T = 128$, the relative performance is stable. As one might have expected, the correlation between covariates plays no role here. In order to provide a comprehensive view of the results, Fig. 4 shows the distribution of the CPU times required by the competing estimators. It is worth stressing that none of the candidates appear to be sensitive to particular data figures since the boxplots do not display high variability nor huge outliers.

We then repeat the experiment above, but with parameters generated according to Eq. (10). Results are reported in Table 3.

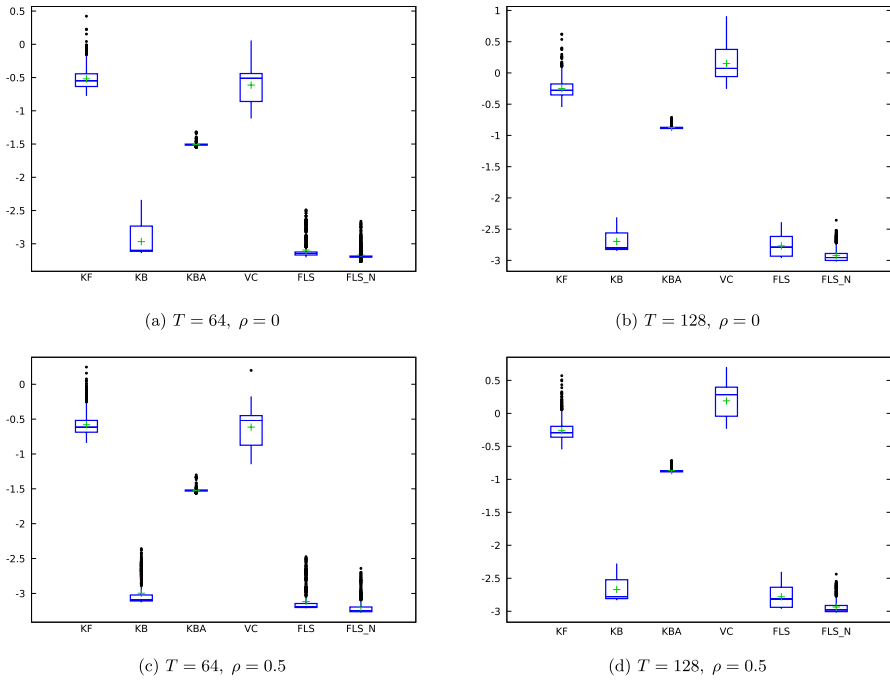


Fig. 4 Distribution of the CPU times, in \log_{10} seconds

In this case, the estimators exhibit different behaviours. The KF, KBA and VC estimators show substantially smaller values of both MD and MAD than the FLS and KB estimators, the difference being more pronounced for MAD. This is in line with what we saw in the previous section, namely a marked difference across the estimators in the behaviour of coefficients in the neighborhood of jumps in the true parameters. Since this is a systematic effect, we end up with large deviations from the truth if the estimates are “too smooth” (or in other words the true parameters are “too jumpy”). Overall, it looks as if estimators relying on parametric assumptions, such as KF and VC, are relatively robust even when those assumptions are violated; and the data-driven smoothing of KBA also delivers relatively robust results compared to simple FLS and KB.

As one referee pointed out, there is an asymmetry here. All the methods under consideration allow—in some form or other—for gradual changes in the parameter estimates, but one subset of the methods also allows for abrupt changes and another subset does not. So if the true parameters do in fact change abruptly, the second subset will under-perform relative to the first in a way that is not counter-balanced by systematically better performance in case of gradual change in the parameters.

Table 3 Simulation results: $\beta_{j,t}$ with jumps

T		$\beta_{1,t}$		$\beta_{2,t}$	
		MD	MAD	MD	MAD
64	KF	0.0017	0.2247	- 0.0010	0.2553
	KB	0.0858	0.6201	- 0.0159	0.8280
	KBA	- 0.0004	0.1755	- 0.0004	0.1750
	VC	0.0013	0.2172	- 0.0005	0.2489
	FLS	0.0680	0.4243	- 0.0468	0.5708
	FLSN	0.0680	0.4155	- 0.0409	0.5617
128	KF	- 0.0050	0.1684	0.0010	0.2013
	KB	- 0.1326	0.3440	- 0.4093	0.6566
	KBA	- 0.0022	0.1411	- 0.0012	0.1433
	VC	- 0.0052	0.1654	0.0007	0.1989
	FLS	- 0.0393	0.1637	- 0.1437	0.2744
	FLSN	- 0.0383	0.1619	- 0.1410	0.2716

3.3 CPU time and problem size

To conclude our analysis on simulated data we explore the sensitivity of the CPU-time to sample size and number of regressors. Here we go back to the one-sample experiment based on Eq. (9) where we let the number of regressors to vary in $J = \{2, 4, 8\}$ and the sample size in $T = \{128, 256, 512\}$. For each design we report the CPU-times for the estimators in Table 4. Here, the experiment run on a Debian Linux computer with two Intel Xeon CPUs E5-2640 v4 2.40 GHz, 250 GB of RAM. The Gretl version is 2023b.

The clearest result is that the time required by each estimator depends, quite obviously, on whether they require numerical optimization or not. This explains why KF and VC emerge as the most costly methods, whereas FLS and KB are the least demanding. The adaptive version of KB, KBA, sits somewhere in the middle, since numerical optimization is restricted to one parameter only (the bandwidth), so it does not scale with the number of parameters.

According to the results, KB, KBA and FLS methods seem to show an order of complexity roughly linear in both J and T . Conversely, KF and VC generally show a quadratic relation in J and T with negligible second-order effects. Overall, this simple experiment confirms that the rank, in terms of CPU time, is stable with larger T and J .

In order to put compute time into perspective from the practitioner’s point of view, it should be noted that, for most time-series data, sample sizes are not very large (except perhaps with high-frequency financial market data). The results we obtain seem to indicate that CPU times are unlikely to be prohibitive in most practical applications, with the possible exception of VC. Nevertheless, CPU time may have to be taken in consideration if one is running these procedures iteratively, maybe in the context of a simulation experiment. For these cases, the two

Table 4 CPU time (sec)

		$T = 128$	$T = 256$	$T = 512$
$J = 2$	KF	0.4216	1.0180	4.2650
	KB	0.0021	0.0055	0.0175
	KBA	0.1018	0.3816	1.7728
	VC	0.6658	4.9813	26.056
	FLS	0.0015	0.0024	0.0045
	FLSN	0.0012	0.0022	0.0041
$J = 4$	KF	2.2921	9.0386	7.3593
	KB	0.0028	0.0084	0.0284
	KBA	0.1298	0.5256	2.4405
	VC	4.4118	32.152	123.83
	FLS	0.0014	0.0027	0.0061
	FLSN	0.0013	0.0025	0.0054
$J = 8$	KF	29.814	124.88	157.59
	KB	0.0060	0.0186	0.0734
	KBA	0.2153	0.9062	6.1115
	VC	80.680	578.91	3988.3
	FLS	0.0020	0.0038	0.0074
	FLSN	0.0018	0.0033	0.0066

KB variants are advisable if standard errors are also needed, and the two variants of FLS otherwise.

4 Application: Okun's law

This section includes an analysis based on real data. As is well known, Okun's law links unemployment to the level of economic activity; in his original article (Okun 1963), Okun noted that a 1% increase in unemployment is typically associated with a 2% change in the deviation of GDP from its potential.

We consider estimation of Okun's law using US quarterly data from 1948:02 to 2023:01 from the St. Louis Fed's FRED database. We use the seasonally adjusted unemployment rate (UNRATE) and the output gap, calculated as the residual from the HP filter ($\lambda = 1600$) on the log of the constant-price GDP (GDPC1).

Figure 5 is very clear in this respect, since the time paths of the two variables are clearly opposite. However, it has been often suggested in the literature (see *eg* Knotek 2007), that Okun's law may be unstable through time.⁵ In fact, after splitting the sample into three subperiods, with breakpoints at 1974:1 and 2000:1, Fig. 6

⁵ Jalles (2019), in particular, estimated a time-varying response of unemployment to output for several countries, although his work is not directly comparable to ours in that only a static specification was considered.

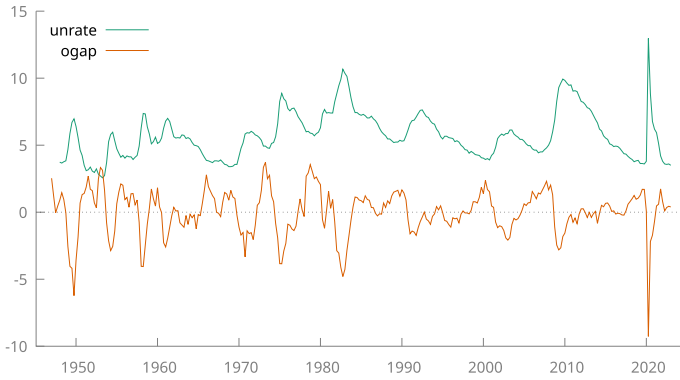


Fig. 5 Time plot of ogap and unrate

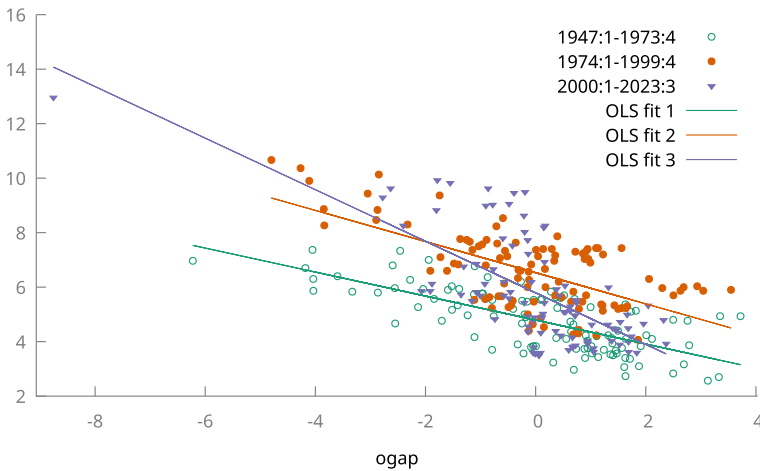


Fig. 6 Scatterplot of \tilde{y} (ogap) and u_t (unrate), by period

suggests that the long-run relationship between unemployment and output has become progressively steeper.

To investigate this possibility, we applied all the methods considered in Sect. 3. We begin by considering the simplest possible statistical model for Okun’s law, that is a static model of the form

$$u_t = \beta_{1,t} + \beta_{2,t}\tilde{y}_t + \varepsilon_t, \tag{11}$$

where u_t is unemployment and \tilde{y}_t is the output gap; the time-invariant version reads

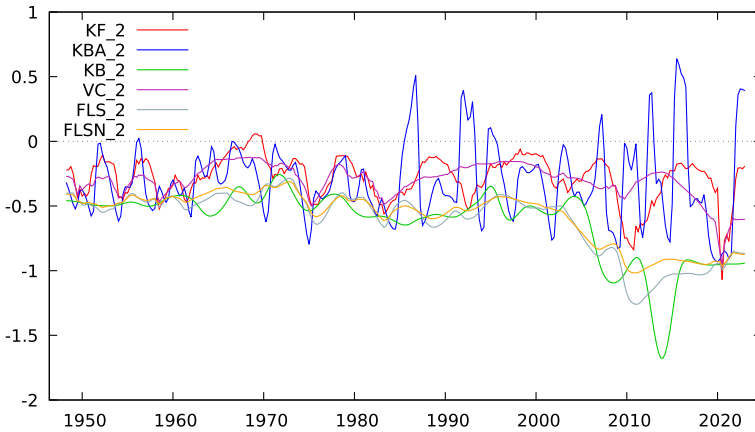


Fig. 7 Estimated semielasticity for Okun's Law—static model

Table 5 Correlation matrix

	KF_2	KBA_2	KB_2	VC_2	FLS_2	FLSN_2
KF_2	1.0000	0.2125	0.3863	0.6594	0.5170	0.5253
KBA_2	0.2121	1.0000	-0.0037	0.3430	-0.0165	0.0138
KB_2	0.3863	-0.0037	1.0000	0.3717	0.8547	0.8581
VC_2	0.6594	0.3430	0.3717	1.0000	0.4887	0.5726
FLS_2	0.5170	-0.0165	0.8547	0.4887	1.0000	0.9840
FLSN_2	0.5253	0.0138	0.8581	0.5726	0.9840	1.0000

$$\widehat{\text{unrate}} = \underset{(0.078627)}{5.72058} - \underset{(0.047636)}{0.614341} \text{ogap}$$

$$T = 300 \quad \bar{R}^2 = 0.3560 \quad F(1, 298) = 166.32 \quad \hat{\sigma} = 1.3618 \quad (12)$$

(standard errors in parentheses)

It is important to note that the model (12) suffers from considerable autocorrelation in the residuals: the DW statistic is around 0.12 and the p value for the LM autocorrelation test with 8 lags is $3.73848e-133$.

Application of the various methods for time-varying estimation yields the estimates of $\beta_{2,t}$ depicted in Fig. 7. In this case, the results are much less reassuring than those obtained with simulated data. Overall, the time-varying estimated coefficients float around stable values over the time span considered but are strongly erratic. We observe sizable discrepancies between all methods. Table 5 reports the sample correlation matrix for the time series in Fig. 7. If results had been homogeneous across methods, one would have expected correlations close to 1. Instead, most correlations appear very weak. The only exception are the two variants of FLS and the fixed-bandwidth version of KB.

Table 6 Error-correction model—time constant coefficients

	Coefficient	Std. Error	p-value
β_1	0.303	0.109	0.006
β_2	-0.054	0.018	0.004
β_3	-0.089	0.020	0.000
β_4	-0.519	0.025	0.000
β_5	-0.079	0.025	0.002
R^2	0.642		
s^2	0.415		
LM test p-value	0.164		

“LM test” is Godfrey’s autocorrelation test with 4 lags

It could be thought that the disappointing performance of the various estimators considered (at least in terms of their ability to deliver a consistent picture) may depend on the gross misspecification inherent in the static formulation of model (12). For this reason, we now consider a dynamic equation so as to take into consideration the time-series properties of the two variables, and accommodate short-run persistence accordingly.

Therefore, we set up a simple ECM model of the form

$$\Delta u_t = \beta_{1,t} + \beta_{2,t}u_{t-1} + \beta_{3,t}\tilde{y}_{t-1} + \beta_{4,t}\Delta\tilde{y}_t + \beta_{4,t}\Delta\tilde{y}_{t-1} + \eta_t. \tag{13}$$

where we focus on the time-varying long run multiplier (LRM), which can be immediately computed as

$$\hat{\kappa}_t = -\frac{\hat{\beta}_{3,t}}{\hat{\beta}_{2,t}}, \tag{14}$$

and is expected to be roughly -0.5 according to the traditional version of Okun’s law. The time-invariant estimates of Eq. (13) are shown in Table 6; note that the estimated value of the long-run multiplier is -1.67, much larger than expected.

Figure 10 shows the estimated time-varying coefficients for the model parameters.⁶ Again, we fail to find homogeneous results going from one estimator to another.

The most worrisome finding, however, concerns the estimated long-run multiplier: while for some of the estimators (KF, VC and FLS, see Fig. 8) the estimates are relatively similar, and conform to the predictions from economic theory, other methods display very wide swings (see Fig. 9). The fact that KB and KBA produce estimates of $\beta_{2,t}$ and $\beta_{3,t}$ with non-coincident sign-switches induce implausible fluctuations, with apparently random sign changes of the long-run multiplier. This result is in line with the findings in Lucchetti and Valentini (2023) about the sensitivity of kernel-based estimators to single data points.

⁶ We do not report the time path of the constant term for reason of space. We limit to report that the relative estimated coefficients exhibit very low time variability, regardless of the estimation technique.

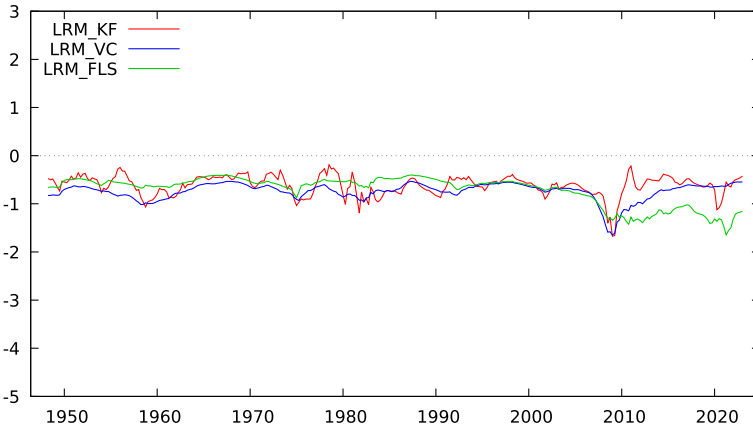


Fig. 8 Estimated long-run multiplier. Methods: KF, VC, FLS

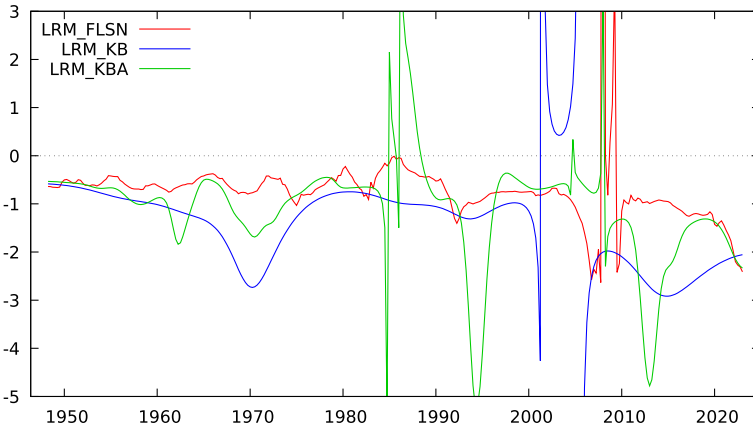


Fig. 9 Estimated long-run multiplier. Methods: KB, KBA, FLSN

The estimate of the long-run multiplier from kernel-based methods turns out to be numerically quite unstable in several cases between 1980 and 2010, as well as the FLSN in the first decade of 2000. All in all, it is worth noting that this behaviour is not ascribable to computational difficulties but is due to the fact that the estimated parameter $\beta_{2,t}$ approaches zero and therefore the LRM takes large values as shown in Fig. 10.

Finally, it is interesting to note that time-varying estimates of the model parameters are not necessarily similar to time-constant ones, even on average. As a result, the LRMs markedly differ if we allow the parameters to adjust over time (whatever estimation method is employed) and even small adjustments in the parameters estimates may result in large discrepancies in the estimation of the quantities of economic interest, such as the LRM.

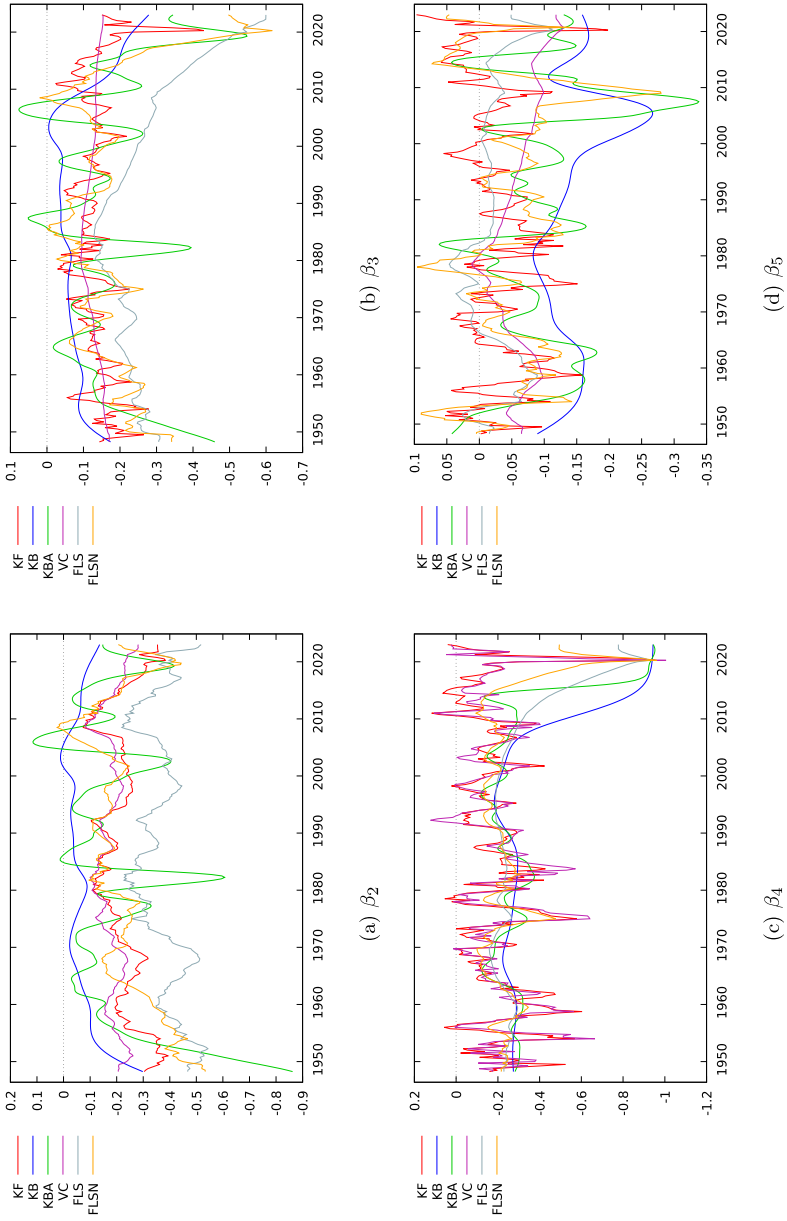


Fig. 10 Okun's law: estimated coefficients. Note: estimates refer to Eq. (13)

5 Final remarks

We have compared several techniques for linear models with time-varying parameters in *gretl*: the candidates are the Kalman filter apparatus, moments and kernel-based methods and flexible least squares. We compared these estimators both via simulations and on real data. Although they perform similarly under controlled and regular scenarios, the behaviour on real data could be different and so the practitioner's choice could be far from obvious. Apart from their statistical properties, we also show that in terms of CPU time these methods offer very different performance, where methods that do not employ numerical optimization, such as kernel-based methods and flexible least squares, outperform the alternatives.

Acknowledgements We thank the Editor, two anonymous Referees and the participants to the 8th *gretl Conference* for their remarks and suggestions.

Funding Open access funding provided by Università Politecnica delle Marche within the CRUI-CARE Agreement.

Data availability Data used in Sect. 4 are publicly available at <https://fred.stlouisfed.org/>. Replication material is available at https://github.com/fravale/replication_tvp_gretl.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Athans M (1974) The importance of Kalman filtering methods for economic systems. In: *Annals of economic and social measurement*, vol 3, no 1. NBER, pp 49–64
- Bai J, Perron P (2003) Critical values for multiple structural change tests. *Econ J* 6(1):72–78
- Bjørnland HC, Thorsrud LA (2019) Commodity prices and fiscal policy design: procyclical despite a rule. *J Appl Econ* 34(2):161–180
- Byrd RH, Lu P, Nocedal J, Zhu C (1995) A limited memory algorithm for bound constrained optimization. *SIAM J Sci Comput* 16(5):1190–1208
- Cai Z (2007) Trending time-varying coefficient time series models with serially correlated errors. *J Econ* 136(1):163–188
- Cai Y, Wu Y (2021) Time-varying interactions between geopolitical risks and renewable energy consumption. *Int Rev Econ Finance* 74:116–137
- Chan KS, Tong H (1986) On estimating thresholds in autoregressive models. *J Time Ser Anal* 7(3):179–190
- Chow GC (1960) Tests of equality between sets of coefficients in two linear regressions. *Econometrica* 28:591–605
- Cimadomo J, D'Agostino A (2016) Combining time variation and mixed frequencies: an analysis of government spending multipliers in Italy. *J Appl Econ* 31(7):1276–1290

- Commandeur JFF, Koopman SJ, Ooms M (2011) Statistical Software for State Space Methods. *J Stat Softw* 41(1):1–18. <https://doi.org/10.18637/jss.v041.i01>
- Cooley TF, Prescott EC (1976) Estimation in the presence of stochastic parameter variation. *Econ J Econ Soc* 167–184
- Fan J, Zhang W (1999) Statistical estimation in varying coefficient models. *Ann Stat* 27(5):1491–1518
- Giraitis L, Kapetanios G, Yates T (2014) Inference on stochastic time-varying coefficient models. *J Econ* 179(1):46–65. <https://doi.org/10.1016/j.jeconom.2013.10.009>
- Giraitis L, Kapetanios G, Yates T (2018) Inference on multivariate heteroscedastic time varying random coefficient models. *J Time Ser Anal* 39(2):129–149. <https://doi.org/10.1111/jtsa.12271>
- Giraitis L, Kapetanios G, Marcellino M (2021) Time-varying instrumental variable estimation. *J Econ* 224(2):394–415
- Hamilton JD (1989) A new approach to the economic analysis of nonstationary time series and the business cycle. *Econ J Econ Soc* 357–384
- Harvey AC (1990) Forecasting, structural time series models and the Kalman filter. Cambridge University Press, Cambridge
- Jalles JT (2019) On the time-varying relationship between unemployment and output: what shapes it? *Scott J Polit Econ* 66(5):605–630
- Johnson E, Moggridge D (eds) (1987) The collected writings of John Maynard Keynes: Volume 14, the general theory and after: Part II. Defense and Development. Cambridge University Press, Cambridge
- Kalaba R, Tesfatsion L (1989) Time-varying linear regression via flexible least squares. *Comput Math Appl* 17(8–9):1215–1245
- Kalaba R, Tesfatsion L (1990) Flexible least squares for approximately linear systems. *IEEE Trans Syst Man Cybern* 20(5):978–989
- Knotek ES (2007) How useful is Okun’s law? *Econ Rev Fed Reserve Bank Kansas City* 92(4):73
- Lucchetti R (2011) State space methods in Gretl. *J Stat Softw* 41(11):1–22. <https://doi.org/10.18637/jss.v041.i11>
- Lucchetti R, Valentini F (2023) Kernel-based time-varying IV estimation: handle with care. *Empir Econ* 1–26
- Lütkepohl H, Herwartz H (1996) Specification of varying coefficient time series models via generalized flexible least squares. *J Econ* 70(1):261–290
- Okun AM (1963) Potential GNP: its measurement and significance. Cowles Foundation for Research in Economics at Yale University
- Pelagatti MM (2015) Time series modelling with unobserved components. CRC Press, Boca Raton
- Schlicht E (2021) VC: a method for estimating time-varying coefficients in linear models. *J Korean Stat Soc* 1–33
- Teräsvirta T (1994) Specification, estimation, and evaluation of smooth transition autoregressive models. *J Am Stat Assoc* 89(425):208–218

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.