



# Bayesian comparison of production function-based and time-series GDP models

Jacek Osiewalski<sup>1</sup> · Justyna Wróblewska<sup>1</sup> · Kamil Makiela<sup>1</sup> 

Received: 7 November 2016 / Accepted: 4 September 2018 / Published online: 20 October 2018  
© The Author(s) 2018

## Abstract

A purely Bayesian vector autoregression (VAR) framework is proposed to formulate and compare tri-variate models for the logs of the economy-wide aggregates of output and inputs (physical capital and labour). The framework is derived based on the theory of the aggregate production function, but at the same time, accounts for the dynamic properties of macroeconomic data, which makes it particularly appealing for modelling GDP. Next, using the proposed framework we confront a-theoretical time-series models with those that are based on aggregate production function-type relations. The common knowledge about capital and labour elasticities of output as well as on their sum is used in order to formulate prior distribution for each tri-variate model, favouring the linearly homogenous Cobb–Douglas production function-type relation. In spite of this, production function-based co-integration models fail empirical comparisons with simple VAR structures, which describe the three aggregates by three stochastic trends.

**Keywords** Bayesian inference · VAR models · Economic growth models · Co-integration analysis · Aggregate production function · Potential output

**JEL Classification** C11 · C51 · C52 · O40

## 1 Introduction

Aggregate production functions, which link aggregated product of an economy with its aggregated physical capital and labour—treated as the factors of production—have been widely used in macroeconomics, economic growth literature in particular. It has been so, despite serious objections formulated by many economists against the

---

✉ Kamil Makiela  
kamil.makiela@uek.krakow.pl

<sup>1</sup> Department of Econometrics and Operations Research, Cracow University of Economics, Kraków, Poland

very concept of aggregate production functions; this critique is summarized in Felipe and McCombie (2013). While microeconomic production functions are reasonable models that describe technology of individual producers, meaningful aggregation of physical capital, labour and output is hardly possible, and any simple relation between aggregates looks very suspicious—see, for example, Fisher (1969). Even in modern attempts to derive aggregate production functions from micro-foundations, the underlying assumptions are extremely specific—see, for example, Jones (2005) and Growiec (2008, 2013). On the other hand, it is often argued that even though the link between the aggregate production function (APF hereafter) and its microeconomic foundations is not clear it still presents a reasonable, well-defined mathematical relationship.

In both theoretical and empirical models of economic growth the Cobb–Douglas functional form is prevailing due to historical reasons, its simplicity, a relatively good data fit as well as some formal justifications—e.g. it is a first-order local approximation of any smooth production function expressed in terms of logs of inputs and output (provided such a relation exists on the level of some aggregates of output and inputs). A more general and quite popular form is translog, which is a second-order local approximation. Nowadays, APF is used, for example, as the world technology frontier in international growth comparisons and in decomposition of economic growth into its main sources, as proposed in Koop et al. (1999, 2000), where Bayesian analysis of stochastic production frontiers is applied; see Makięła (2014) for a more recent study. Analyses of the production function-type relations between aggregates are mainly performed by econometricians and empirical macroeconomists. Their focus is more on statistical issues of modelling and inference than on theoretical economic foundations of such empirical models.

Stochastic production frontier models extend traditional production functions by adding non-negative unobservable random terms that represent technical inefficiency. Such models either are completely static, or use time trends (or time-varying parameters) to capture technology changes. On the other hand, however, GDP and GNP data (mainly for the USA and some other developed economies) were used long ago to demonstrate important models and techniques of modern econometrics; see, for example, Koop et al. (1997) for a Bayesian approach to joint modelling and testing of long memory and a unit root. Describing dynamic properties of economic time series is a crucial issue, and co-integration analysis constitutes a standard framework to establish stable, long-run relations among macroeconomic variables—variables that can usually be described by integrated stochastic processes. Co-integration techniques are useful tools to avoid *spurious regressions*. So far, the stochastic frontier—or, more generally, a comparative—analysis of economic growth has ignored these issues when analysing panel data, even those covering longer periods, like 30 years. Although our initial motivation was rooted in stochastic frontier production functions for panel data, we focus here on a simpler, purely time-series analysis with no panel context.

Since production function specifications use data on three variables, we have to build appropriate tri-variate models for Bayesian comparisons. Thus, all our competing models consist of three equations. We start from basic VAR( $p$ ) specifications—their first two equations (for the logarithms of inputs aggregates) are also used in production function-type models; only the form of the third equation (for the log of GDP) is

specific. Empirical comparisons are based on individual economies (not on a panel of countries), namely the USA and Poland, but also on the UK and Hungary—in order to check whether the results are similar for different countries, developed ones as well as those undergoing economic transformation. Since we use annual data from relatively short periods,  $p=3$  is taken as the maximum lag.

The Bayesian VAR framework enables us to calculate posterior probabilities of many nested and non-nested models, in particular specifications that assume co-integrating relations (for Bayesian co-integration analysis, see, for example, Koop et al. 2004, 2010; Wróblewska 2009). In order to analytically obtain basic results, we use as much as possible the conjugate prior distributions. We also resort to the concept of Bayesian sequential cut (see, for example, Florens and Mouchart 1985; Osiewalski and Steel 1996), which allows us to separate the analysis of aggregate inputs dynamics from the main issue of analysing aggregate output given current aggregate inputs levels. This main issue is closely related to the formal econometric status of any APF-type relation. Thus, we would like to look from an a-theoretical, purely time-series co-integration perspective at the issue of existence and empirical adequacy of APF-type relations.

The aim of this study is to provide an unbiased, scientific evidence in favour or against APF, and thus to shed more light on the topic that has been on the research agenda for decades. Our contribution in this regard is threefold. Firstly, there are two main approaches used by practitioners to calculate potential output or output gap in macroeconomics: (i) a purely dynamic, time-series approach and (ii) a theory-driven approach based on APF; see, for example, Turner et al. (2016) or Alichì et al. (2017) for a survey of methods used to model macroeconomic output. We present a modelling framework that, on the one hand, addresses the dynamic properties of macroeconomic data and, on the other hand, encompasses models deeply rooted in economic theory of aggregate production and productivity. Secondly, we present a formal Bayesian model comparison, which brings together these two pieces of the literature that describe aggregate production. Thirdly, we examine the role of APF-type relations, especially Cobb–Douglas and translog, within a dynamic econometric framework and analyse if APF can be viewed as an empirically valid co-integration relation.

We are fully aware that our study requires an explicit assumption that APF is a testable, falsifiable construct. This, however, may not be feasible in some fields of macroeconomics, which greatly rely on the very existence of APF. We do not wish to dispute such studies since, given the Bayesian perspective, they merely imply a strong prior belief as regards APF—a belief that, via prior odds, simply outweighs any information that may come from (sometimes messy and distorted) macroeconomic data. We are also aware that our purely statistical approach to the relation between economy-wide aggregates of output and inputs can be criticized—even if we find a co-integration equation that can be interpreted as the APF. One should be aware of the argument given in Shaikh (1974), also discussed in Felipe and McCombie (2013), which shows that a well-fitted empirical Cobb–Douglas-type relation can be nothing more but a straightforward consequence of a simple identity between constant-price value added and the sum of wages and profits, provided that the observed shares of wages and profits are constant over time. That is, even if we find a co-integration relation with properties attributed to traditional production functions, the latter can

still be considered *spurious* in a fundamental, economic sense. Such relation would only show that the conditions formulated in Shaikh (1974) are approximately met. However, if such co-integration relation is unlikely in view of the data, we obtain an empirical evidence against conditions that would make the APF-type equation a useful approximation of reality.

The structure of the paper is as follows: Sects. 2 and 3 present the technical details of the VAR modelling framework, Bayesian model comparison and the role of the Cobb–Douglas relation. Readers more interested in the empirical results may move forward to Sect. 4 where competing model classes are presented. Section 5 focuses on the data and empirical results of model comparison. Section 6 is devoted to the best specifications and posterior results on crucial parameters. Final remarks are summarized in Sect. 7.

## 2 Bayesian VAR modelling framework

We describe annual data on GDP (i.e. aggregate product  $Q_t$ ) as well as aggregates of capital ( $K_t$ ) and labour ( $L_t$ ) using a tri-variate VAR( $p$ ) model for the natural logarithms of both inputs ( $k_t = \ln K_t$ ,  $l_t = \ln L_t$ ) and output ( $q_t = \ln Q_t$ ). The basic form of a model for  $x_t = [k_t l_t q_t]$  can be written as

$$x_t = \sum_{i=1}^p x_{t-i} \Phi_i + d_t \Gamma_o + \varepsilon_t, \quad (1a)$$

where  $\Phi_i$  are  $(3 \times 3)$  matrices of parameters accompanying lags  $x_{t-i} = [k_{t-i} l_{t-i} q_{t-i}]$  ( $i = 1, \dots, p$ ),  $\Gamma_o$  is a  $(m \times 3)$  matrix of coefficients that post-multiplies  $d_t$ , the row vector of  $m$  deterministic components (e.g.  $m = 2$  and  $d_t = [1 \ t]$ ), and  $\{\varepsilon_t\}$  is a tri-variate Gaussian white noise process with contemporaneous covariance matrix  $\Sigma$ , a  $(3 \times 3)$  symmetric positive definite matrix.

Denoting  $z_t = [x_{t-1} \dots x_{t-p} d_t]$  and  $\Phi = [\Phi'_1 \dots \Phi'_p \Gamma'_o]'$  we can rewrite (1a) as

$$x_t = z_t \Phi + \varepsilon_t, \quad (1b)$$

where  $\Phi$  is  $(3p + m) \times 3$  and can be represented by its 3 columns as  $\Phi = [\varphi^{(1)} \varphi^{(2)} \varphi^{(3)}]$ , where  $\varphi^{(j)}$  groups the  $3p + m$  unknown coefficients appearing in equation  $j$  ( $j = 1, 2, 3$ ).

For the purposes of co-integration analysis, the particularly important representation of (1a) is the one in terms of first differences:

$$x_t = x_{t-1}(I_3 + \Pi) + \sum_{j=1}^{p-1} \Delta x_{t-j} \Gamma_j + d_t \Gamma_o + \varepsilon_t, \quad (2a)$$

where  $\Pi = -I_3 + \sum_{i=1}^p \Phi_i$  and, for  $p > 1$ ,  $\Gamma_j = - \sum_{h=j+1}^p \Phi_h$ . Denoting  $\tilde{z}_t = [x_{t-1} \Delta x_{t-1} \dots \Delta x_{t-(p-1)} d_t]$  and  $\tilde{\Phi} = [\Pi' \Gamma_1' \dots \Gamma_{p-1}' \Gamma_o']'$  we can rewrite (2a) as

$$x_t = x_{t-1} + \tilde{z}_t \tilde{\Phi} + \varepsilon_t, \tag{2b}$$

where, again,  $\tilde{\Phi} = [\tilde{\varphi}^{(1)} \tilde{\varphi}^{(2)} \tilde{\varphi}^{(3)}]$  is  $(3p + m) \times 3$  with  $\tilde{\varphi}^{(j)}$  grouping the  $3p + m$  unknown coefficients of equation  $j$  in (2a) ( $j=1, 2, 3$ ). Note that  $\tilde{\Phi} = R\Phi - F$ , where

$$R = \begin{bmatrix} I_3 & I_3 & I_3 & \dots & I_3 & 0 \\ 0 & -I_3 & -I_3 & \dots & -I_3 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & -I_3 & 0 \\ 0 & 0 & 0 & \dots & 0 & I_m \end{bmatrix}, \quad F = \begin{bmatrix} I_3 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}; \quad RF = F.$$

For the parameter matrices  $\Phi$  and  $\Sigma$  in (1a,b) we assume the conjugate joint prior distribution, i.e.  $\Sigma$  has an inverted Wishart marginal prior distribution and  $\Phi$  has a matrix-variate normal conditional distribution (given  $\Sigma$ ), that is

$$\Sigma \sim IW(A, s), \quad \text{vec}(\Phi) | \Sigma \sim N(\text{vec}(F), \Sigma \otimes W).$$

The hyper-parameters  $A$  (a  $3 \times 3$  symmetric, positive definite scale matrix) and  $s$  (degrees of freedom,  $s > 2$ ) will be elicited in the next section. In the conditional prior of  $\Phi = [\varphi^{(1)} \varphi^{(2)} \varphi^{(3)}]$  we reflect the idea that  $x_t$  is, most likely, just the tri-variate random walk, i.e.  $x_t = x_{t-1} + \varepsilon_t$ . Thus, we take  $F = [I_3 \ 0]'$  as the prior mean. The precision of our prior assumptions on  $\Phi$  is reflected by the choice of  $W$ , as  $\Sigma \otimes W$  is the covariance matrix of  $\text{vec}(\Phi)$  given  $\Sigma$ . We assume that  $W$ , a square matrix of order  $3p + m$ , is block-diagonal with  $I_3, \frac{1}{2}I_3, \dots, \frac{1}{p}I_3, W_0$  as diagonal blocks. Our prior resembles the so-called Minnesota (Litterman) prior (see, for example, Doan et al. 1984; Litterman 1986), except that we use a purely Bayesian approach and our prior is exactly the marginal distribution of the parameters; thus, it does not use hyper-parameters based on the data being modelled.

The induced conditional prior for the matrix  $\tilde{\Phi} = R\Phi - F$ , which groups the conditional mean parameters in (2b), is matrix-variate normal:  $\text{vec}(\tilde{\Phi}) | \Sigma \sim N(0, \Sigma \otimes \tilde{W})$ , where

$$\tilde{W} = RWR' = \begin{bmatrix} \left(1 + \frac{1}{2} + \dots + \frac{1}{p}\right)I_3 & -\left(\frac{1}{2} + \dots + \frac{1}{p}\right)I_3 & \dots & -\frac{1}{p}I_3 & 0 \\ -\left(\frac{1}{2} + \dots + \frac{1}{p}\right)I_3 & \left(\frac{1}{2} + \dots + \frac{1}{p}\right)I_3 & \dots & \frac{1}{p}I_3 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ -\frac{1}{p}I_3 & \frac{1}{p}I_3 & \dots & \frac{1}{p}I_3 & 0 \\ 0 & 0 & \dots & 0 & W_0 \end{bmatrix}.$$

The simple prior distribution described here enables us to use purely analytical approach in order to obtain basic posterior results and the marginal data density (MDD hereafter). Of course, our conjugate prior is defined for completely free, unconstrained parameters of (1a, b) or, equivalently, (2a, b). In particular, we do not impose stability conditions that are imposed in, for example, Wróblewska (2009).

There are many ways of generating competing models with the use of this general framework. Obvious special cases are obtained by assuming different orders  $p$  of  $\text{VAR}(p)$  or imposing  $\Pi = 0$  for each  $p$  and considering  $\text{VAR}(p - 1)$  for  $\Delta x_t$ . But our main task is to reconsider the production function-type relation within dynamic framework; hence, the crucial transformation of  $\text{VAR}(p)$  is the one in terms of the marginal sampling model for  $x_t^{(1)} = [k_t \ l_t]$ , given the past of all three variables in  $x_t$ , and the conditional sampling model for  $q_t$  given  $x_t^{(1)}$  and the past of  $x_t$ . This enables us to directly consider exogeneity issues and to propose many other competing models.

Starting from (1a), we obtain the following forms of the sampling distribution of  $x_t^{(1)}$  given the past of  $x_t$  (denoted as  $\psi_{t-1}$ ) and the sampling distribution of  $q_t$  given  $x_t^{(1)}$  and  $\psi_{t-1}$ :

$$x_t^{(1)} | \psi_{t-1}, \Phi, \Sigma \sim N(z_t \Phi^{(1)}, \Sigma_{11}), \tag{3}$$

$$q_t | \psi_{t-1}, x_t^{(1)}, \Phi, \Sigma \sim N(x_t^{(1)} \beta + z_t \delta, \omega), \tag{4}$$

where  $\Phi^{(1)} = [\varphi^{(1)} \varphi^{(2)}]$  groups the coefficients of the first two equations of  $\text{VAR}(p)$ ,  $\Sigma_{11}$  is the  $(2 \times 2)$  left upper block of

$$\Sigma = \begin{bmatrix} \sigma_{kk} & \sigma_{kl} & \sigma_{kq} \\ \sigma_{kl} & \sigma_{ll} & \sigma_{lq} \\ \sigma_{kq} & \sigma_{lq} & \sigma_{qq} \end{bmatrix} = \begin{bmatrix} \Sigma_{11} & \Sigma_{21} \\ \Sigma_{21} & \sigma_{qq} \end{bmatrix},$$

and  $\beta = \Sigma_{11}^{-1} \Sigma_{12}$ ,  $\omega = \sigma_{qq} - \Sigma_{21} \beta > 0$ ,  $\delta = \varphi^{(3)} - \Phi^{(1)} \beta$ .

Due to the well-known properties of inverted Wishart distributions,  $\Sigma_{11}$  and  $(\beta, \omega)$  are stochastically independent,  $\Sigma_{11} \sim IW(A_{11}, s - 1)$ ,  $\beta | \omega \sim N(A_{11}^{-1} A_{12}, \omega A_{11}^{-1})$ , and  $\omega \sim IW(a_{qq.1}, s)$  or, equivalently,  $\tau = \omega^{-1} \sim \text{Gamma}(\frac{s}{2}, \frac{a_{qq.1}}{2})$ , where we use the obvious partitioning of the hyper-parameter matrix

$$A = \begin{bmatrix} a_{kk} & a_{kl} & a_{kq} \\ a_{kl} & a_{ll} & a_{lq} \\ a_{kq} & a_{lq} & a_{qq} \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & a_{qq} \end{bmatrix},$$

and  $a_{qq.1} = a_{qq} - A_{21} A_{11}^{-1} A_{12} > 0$ . Thus, the decomposition of (1a) into (3) and (4) has led to a one-to-one re-parameterization of the conditional (contemporaneous) covariance matrix  $\Sigma$  into two groups of parameters,  $\Sigma_{11}$  and  $(\beta, \omega)$ , which are a priori independent;  $\Sigma_{11}$  appears only in the marginal model for inputs (given the past of  $x_t$ ) and  $(\beta, \omega)$  together with  $\delta$  parameterizes the conditional model of output (given current inputs and  $\psi_{t-1}$ ). The one-to-one re-parameterization of  $\Phi$ , which appears in (3) and (4), leaves  $\Phi^{(1)} = [\varphi^{(1)} \varphi^{(2)}]$  in the marginal model for inputs and uses

$\delta = \varphi^{(3)} - \Phi^{(1)}\beta$  in the conditional model for output. Note that the joint conditional prior of  $\text{vec}[\Phi^{(1)}\delta] = Q_\beta \text{vec}(\Phi)$ , given  $\Sigma$ , is normal with mean  $Q_\beta \text{vec}(F)$  and covariance matrix  $Q_\beta(\Sigma \otimes W)Q'_\beta$ , where

$$Q_\beta = \begin{bmatrix} I & & 0 \\ -\beta' \otimes I & I & I \end{bmatrix}; \quad \text{thus} \quad Q_\beta \text{vec}(F) = \text{vec} \begin{bmatrix} 1 & 0 & -\beta_1 \\ 0 & 1 & -\beta_2 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix},$$

and  $Q_\beta(\Sigma \otimes W)Q'_\beta$  is block-diagonal with matrices  $\Sigma_{11} \otimes W$  and  $\omega W$  as diagonal blocks.

All this means that  $\theta^{(1)} = (\Phi^{(1)}, \Sigma_{11})$  is a sufficient parameterization of (3),  $\theta^{(2)} = (\delta, \beta, \omega)$  is a sufficient parameterization of (4), and both groups of parameters are independent a priori, which guarantees the Bayesian sequential cut (see, for example, Florens and Mouchart 1985; Osiewalski and Steel 1996), leading not only to posterior independence between  $\theta^{(1)}$  and  $\theta^{(2)}$  (i.e. their conditional independence given observations), but also to a useful factorization of MDD given initial observations:

$$p(x_1, \dots, x_T | \psi_0) = \int_{\Theta_1} \int_{\Theta_2} p(\theta^{(1)})p(\theta^{(2)}) \prod_{t=1}^T [p(x_t^{(1)} | \psi_{t-1}, \theta^{(1)}) p(q_t | \psi_{t-1}, x_t^{(1)}, \theta^{(2)})] d\theta^{(2)} d\theta^{(1)} = C_1 C_2$$

where

$$C_1 = \int_{\Theta_1} p(\theta^{(1)}) \prod_{t=1}^T p(x_t^{(1)} | \psi_{t-1}, \theta^{(1)}) d\theta^{(1)},$$

$$C_2 = \int_{\Theta_2} p(\theta^{(2)}) \prod_{t=1}^T p(q_t | \psi_{t-1}, x_t^{(1)}, \theta^{(2)}) d\theta^{(2)},$$

$p(\theta^{(1)})$  and  $p(\theta^{(2)})$  are prior densities and  $T$  is the number of tri-variate observations. Thus, our prior assumptions lead to complete separation of the Bayesian analysis in the marginal and conditional models (3) and (4).

The parameterization in terms of  $\theta^{(1)}$  and  $\theta^{(2)}$  is obviously not the only one that leads to the sequential cut. Instead of (1a, b), we can start from (2a, b) and obtain the following forms of the marginal and conditional sampling models:

$$x_t^{(1)} | \psi_{t-1}, \tilde{\Phi}, \Sigma \sim N(x_{t-1}^{(1)} + \tilde{z}_t [\tilde{\varphi}^{(1)} \tilde{\varphi}^{(2)}], \Sigma_{11}), \tag{5}$$

$$q_t | \psi_{t-1}, x_t^{(1)}, \tilde{\Phi}, \Sigma \sim N(x_t^{(1)}\beta + q_{t-1} - x_{t-1}^{(1)}\beta + \tilde{z}_t \tilde{\delta}, \omega), \tag{6}$$

where  $\tilde{\delta} = \tilde{\varphi}^{(3)} - [\tilde{\varphi}^{(1)} \tilde{\varphi}^{(2)}]\beta$  and (6) can be equivalently written as

$$\Delta q_t | \psi_{t-1}, x_t^{(1)}, \tilde{\Phi}, \Sigma \sim N\left(\Delta x_t^{(1)} \beta + \tilde{z}_t \tilde{\delta}, \omega\right).$$

Now  $\tilde{\theta}^{(1)} = (\tilde{\varphi}^{(1)}, \tilde{\varphi}^{(2)}, \Sigma_{11})$  and  $\tilde{\theta}^{(2)} = (\tilde{\delta}, \beta, \omega)$  are sufficient parameterizations of the marginal and conditional models, respectively. Since  $\text{vec}(\tilde{\Phi}) | \Sigma \sim N(0, \Sigma \otimes \tilde{W})$  and  $\text{vec}[\tilde{\varphi}^{(1)} \tilde{\varphi}^{(2)} \tilde{\delta}] = Q_\beta \text{vec}(\tilde{\Phi})$ ,

$$\text{vec}[\tilde{\varphi}^{(1)} \tilde{\varphi}^{(2)} \tilde{\delta}] | \Sigma_{11}, \beta, \omega \sim N\left(Q_\beta 0, Q_\beta (\Sigma \otimes \tilde{W}) Q'_\beta\right) = N\left(0, \begin{bmatrix} \Sigma_{11} \otimes \tilde{W} & 0 \\ 0 & \omega \tilde{W} \end{bmatrix}\right).$$

The conditional prior of  $\text{vec}[\tilde{\varphi}^{(1)} \tilde{\varphi}^{(2)}]$  given  $\Sigma_{11}$  is  $N(0, \Sigma_{11} \otimes \tilde{W})$ , the conditional prior of  $\tilde{\delta}$  given  $(\beta, \omega)$  is  $N(0, \omega \tilde{W})$ , which does not depend on  $\beta$ , and  $\tilde{\theta}^{(1)}$  and  $\tilde{\theta}^{(2)}$  are a priori independent. Thus, under our prior structure, the complete separation of posterior inference and MDD calculations is preserved not only for (3) and (4), but also for (5) and (6). This is useful for the model formulation strategy. As long as we do not impose restrictions that link parameters of the marginal and conditional models, we can separately consider their competing variants. Consequently, if we consider  $n_1$  variants of the marginal model and  $n_2$  variants of the conditional model,  $n_1 \cdot n_2$  models for  $x_t$  can be formulated and compared. Note that it is formally irrelevant which particular parameterization we choose—(3) or (5) for the marginal model, (4) or (6) for the conditional model. For instance, it may be easier to formulate separate restrictions on  $\theta^{(1)}$  and  $\tilde{\theta}^{(2)}$  (than on  $\theta^{(1)}$  and  $\theta^{(2)}$ ) in order to create an interesting (interpretable) model for  $x_t$ . Incidentally, remind that  $\theta^{(h)}$  and  $\tilde{\theta}^{(j)}$  are a priori independent ( $h, j = 1, 2; h \neq j$ ). Since  $\theta^{(1)} \perp \theta^{(2)}$  and  $[\tilde{\varphi}^{(1)} \tilde{\varphi}^{(2)}]$  is a (one-to-one) transformation of  $[\varphi^{(1)} \varphi^{(2)}]$ , it is clear that  $\tilde{\theta}^{(1)} \perp \theta^{(2)}$ ; starting from  $\tilde{\theta}^{(1)} \perp \tilde{\theta}^{(2)}$ , by the same argument we get  $\theta^{(1)} \perp \tilde{\theta}^{(2)}$ .

### 3 The role of the Cobb–Douglas-type relation in the modelling strategy

In this section we focus on the conditional sampling model for output and present the use of the relation between current levels of all three variables in both creating specific models and eliciting prior distribution of the conditional covariance matrix  $\Sigma$ .

In both equivalent formulations of the conditional sampling model for output given inputs, (4) and (6), the logarithm of output is linearly dependent on the logarithms of current input levels (plus some linear function of past values of  $k_t, l_t, q_t$ ). In view of the traditional concept of the APF, the presence of the relation between output and current inputs in the conditional model could be called the contemporaneous or immediate APF. Since we work within the VAR model framework for logs of  $Q_t, K_t$  and  $L_t$ , this relation takes the Cobb–Douglas form. It is nothing to be empirically discovered. It is obtained by construction, as the nonzero vector  $\beta = \Sigma_{11}^{-1} \Sigma_{12}$  is a consequence of nonzero conditional correlation between output and current inputs, i.e. nonzero  $\Sigma_{12}$  in the contemporaneous covariance matrix  $\Sigma$  (conditional with respect to the past of  $x_t$ ).



As we consider standard VAR models, the conditional covariance matrix is constant over time, so are the immediate capital and labour elasticities of output. Making this covariance matrix vary over time would lead to time-varying elasticities in the spirit of Koop et al. (2000). That is, the immediate Cobb–Douglas APF with either constant or variable parameters is an obvious part of the standard VAR or a more general (TVP VAR) model.

Although the static Cobb–Douglas form is just a matter of re-parameterization of the basic VAR framework, there are some deeper, important questions, which can be answered through empirical investigations only. First, we are interested whether the immediate APF-type relation is the only important part of the conditional sampling model. If so, the very old idea of a purely static Cobb–Douglas APF would seem still relevant. If not, we may enquire about empirical validity of its different extensions, like generalizing its functional form, keeping deterministic trend in the equation, adding one-sided inefficiency terms, replacing  $\beta$  by a time-varying vector of parameters. From the viewpoint of dynamic econometrics, the most interesting issue amounts to testing if the immediate Cobb–Douglas APF-type relation is also a co-integration relation linking all three aggregate variables, if they are treated as integrated stochastic processes. If the issues mentioned above can be formalized as restrictions on—or extensions of—the conditional sampling model only, such modelling strategy preserves the Bayesian sequential cut. In this case weak exogeneity of inputs (for inference on any function of the conditional model parameters) is assumed. Of course, weak exogeneity can be tested within the full, three-equation framework—as well as the presence of any co-integration relation among the three (obviously non-stationary) variables. These issues require Bayesian analysis of the joint model for inputs and output—without any separation of MDD calculations—and thus they are computationally more demanding.

Before we present the competing models and their Bayesian comparison using observations from four countries, let us use the immediate Cobb–Douglas APF idea to elicit the hyper-parameters ( $A$  and  $s > 2$ ) of the inverted Wishart prior of  $\Sigma$ . Since, given  $\omega$ ,  $\beta$  is normal with mean  $A_{11}^{-1}A_{12}$  and covariance matrix  $\omega A_{11}^{-1}$ , and  $\tau = \omega^{-1}$  has the gamma distribution with mean  $sa_{qq.1}^{-1}$  and variance  $2sa_{qq.1}^{-2}$ , the marginal prior distribution of  $\beta$  is Student's  $t$  with  $s > 2$  degrees of freedom, mean  $A_{11}^{-1}A_{12}$  and covariance matrix  $\frac{a_{qq.1}}{s-2}A_{11}^{-1}$ . Although, following Fisher (1969), Shaikh (1974) and Felipe and McCombie (2013), we are quite sceptical about the concept of the APF, we do favour the traditional Cobb–Douglas case (with constant returns to scale) through prior elicitation. Thus, this disputable concept, so popular in many economic models, stands in the centre of our prior elicitation and  $\beta$  is treated as the vector of capital and labour elasticities of output that, most likely, are equal and sum up to unity, with prior standard deviations 0.2 for each elasticity and 0.1 for their sum. This leads to five equations with seven unknowns; the sixth equation is obtained by assuming that the diagonal elements of  $A$  are equal. Finally, we get  $s = \frac{26}{3}$  and

$$A = \frac{a}{15} \begin{bmatrix} 16 & 14 & 15 \\ 14 & 16 & 15 \\ 15 & 15 & 16 \end{bmatrix}, \quad a > 0;$$

since  $a_{qq.1} = \frac{a}{15}$ , the corresponding marginal prior for  $\tau = \omega^{-1}$  is gamma  $(\frac{13}{3}, \frac{a}{30})$ . We assume  $a=1$ , which results in the prior mean 130 and prior standard deviation  $10\sqrt{39}$  for the precision parameter  $\tau$ .

Finally, let us focus on the prior distribution of the parameters of just one equation in (1b). Without loss of generality, we consider only the first equation, where  $\varphi^{(1)}$  given  $\sigma_{kk}$  is normal (with mean  $[1 \ 0]'$  and covariance matrix  $\sigma_{kk}W$ ) and  $\sigma_{kk}^{-1}$  is gamma (with parameters  $\frac{s-2}{2}$  and  $\frac{a_{kk}}{2}$ , i.e. with mean  $\frac{s-2}{a_{kk}}$  and standard deviation  $a_{kk}^{-1}\sqrt{2(s-2)}$ ). Hence, the marginal prior distribution of  $\varphi^{(1)}$  is the Student's  $t$  distribution with  $s-2$  degrees of freedom, location vector  $[1 \ 0]'$  and precision matrix  $\frac{s-2}{a_{kk}}W^{-1}$ ; our choice of the hyper-parameters values leads to the prior covariance matrix  $V(\varphi^{(1)}) = \frac{a_{kk}}{s-4}W = \frac{8}{35}\text{Diag}(I_3, \frac{1}{2}I_3, \dots, \frac{1}{p}I_3, W_0)$ . In the case of  $m=2$  and  $d_t = [1 \ t]$  (unrestricted constant and trend) we take  $W_0 = 100I_2$ .

## 4 Competing models

### 4.1 Type A models: standard VAR and VEC specifications

Since we work with annual data, we only consider VAR( $p$ ) models with  $p \in \{1, 2, 3\}$  and use the parameterization (2a, b), which enables us to impose rank restrictions on the  $(3 \times 3)$  long-run multiplier matrix  $\Pi$ . Assuming that  $r = \text{rank}(\Pi)$  is a particular number from the set  $\{0, 1, 2, 3\}$ , we generate four (very different) cases of VAR( $p$ ).

In the case of  $r=0$  (VAR( $p-1$ ) for  $\Delta x_t$ ,  $\Pi=0$ ) or  $r=3$  (VAR( $p$ ) for  $x_t$ , unrestricted non-singular  $\Pi$ ) the sequential Bayesian cut is preserved when we work with appropriate sufficient parameterizations of the marginal model for inputs and the conditional model for output given current inputs, like in (5) and (6). Basic results can be obtained analytically, and the marginal data density factorizes as  $\text{MDD} = C_1 C_2$ .

The cases of  $r=1$  and  $r=2$  amount to one or two co-integration relations with no restrictions other than the basic identifying ones. Such specifications assume that the same parameters (e.g. co-integrating vectors) appear in all three equations and thus Bayesian cut is precluded. There is no MDD factorization, and numerical tools have to be used (see "Appendix" for details of Bayesian co-integration analysis adopted in this study).

All type A models belong to the VAR family; they differ in terms of  $p$ ,  $r$  and the presence and character of the deterministic variables  $d_t$  in (2a).

### 4.2 Type B models: based on production function

Type B models preserve (by their construction) the Bayesian sequential cut and thus MDD factorization. In fact, in all type B models, as the marginal model for current inputs, we take the first two equations (for  $x_t^{(1)}$ ) of this type A model (with  $r=0$  or  $r=3$ ) that leads to the highest value of  $C_I$ , the first factor of MDD. Type B specifications differ in the form of the third equation, i.e. the conditional model for output given

current inputs. The basic (and the simplest) one corresponds to zero restrictions on the whole vector  $\delta$  in (4) except for the intercept; that is, the third equation consists of the static Cobb–Douglas APF and nothing else. In other type B models the conditional part (for output given current inputs) is richer; its more general form can be written as

$$q_t = \beta_1 k_t + \beta_2 l_t + \gamma_1 k_t^2 + \gamma_2 k_t l_t + \gamma_3 l_t^2 + \delta_1 t + \delta_0 + v_t - z_t, \quad (7)$$

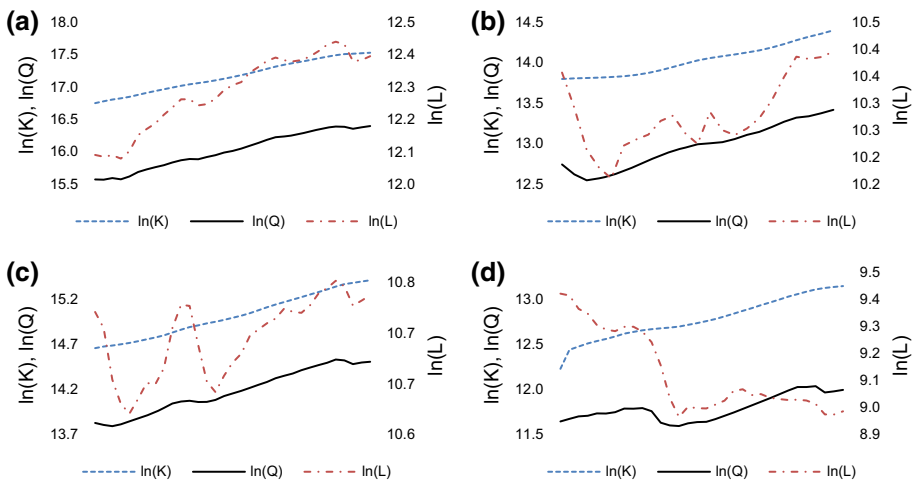
where  $v_t \sim iiN(0, \omega)$  and  $z_t \equiv 0$  or  $z_t \sim iiExp(\lambda)$ ; if they appear, the exponential terms  $z_t$  (with mean and standard deviation  $\lambda$ ) are independent of symmetric errors  $v_t$  and represent period  $t$  technical inefficiency. The parameters  $\beta = [\beta_1 \beta_2]'$ ,  $\delta_0$ ,  $\delta_1$  and  $\omega$  appear also in (4), the conditional part of type A models with  $r = 0$  or  $r = 3$ , so we keep for  $(\beta_1, \beta_2, \delta_0, \delta_1, \omega)$  the same conjugate prior distribution as in (4).

The translog parameters  $(\gamma_1, \gamma_2, \gamma_3)$  are a priori normally distributed (given  $\omega$  and the remaining parameters) with zero mean and covariance matrix  $60^{-1} \omega I_3$ ; higher variances have also been tried in the empirical part, always leading to worse Bayesian models in terms of MDD values. If  $z_t = 0$  for all  $t$ , (7) can be treated as the classical normal linear regression model with a conjugate prior;  $C_2$  and basic posterior results can be obtained analytically.

In cases where exponential inefficiency terms are introduced, we assume that  $\lambda$  and the vector of other parameters are independent and  $\lambda^{-1}$  is a priori exponential with mean  $(-\ln r^*)^{-1}$ , where  $r^* \in (0, 1)$  is the prior median of technical efficiency (see van den Broeck et al. 1994). Posterior inference is based on Gibbs sampling, as described in Osiewalski and Steel (1998), where  $nT$  inefficiency terms are assumed conditionally independent not only over  $n$  units, but also over  $T$  periods; this is directly useful now, as in the present work we have  $n = 1$  (each economy is treated separately.) Evaluating  $C_2$  in models with inefficiency is a non-trivial numerical task; we use a very quick and efficient method, the modified arithmetic mean estimator proposed in Pajor (2017).

Although the conditional part of any type B model is static (except for a possible deterministic trend), the marginal part is always dynamic. Thus, the first two equations (taken from some VAR( $p$ ) model) describe the dynamics of input aggregates (conditionally not only on their own past, but also on the past of the aggregated output) and the third equation postulates a strong (almost deterministic) relation between the aggregates of output and current inputs.

Note that sampling models that can be obtained either by certain restrictions in class A or, equivalently, by appropriate restrictions in class B are particularly important for fair comparison of these two classes. Keeping, for the parameters of such a common specification, the same marginal distribution in B as in A makes prior elicitation in class B easier and enables making Bayesian model comparison as coherent as possible. The full-efficiency ( $z_t \equiv 0$ ) Cobb–Douglas production function-type relations with and without linear trend are two obvious special cases that can be obtained from either (4) or (7); thus, they belong to both classes A and B.



**Fig. 1** Annual data on GDP (solid line), capital (dashed line) and labour (dash-dot line); natural logs; **a** the USA (1982–2011,  $T=30$ ); **b** Poland (1992–2011,  $T=20$ ); **c** the UK (1982–2011,  $T=30$ ); **d** Hungary (1983–2011,  $T=29$ )

## 5 The data and results of initial model comparison

We use annual data prepared for international comparisons and published in Penn World Tables, version 8.0;  $K_t$  is capital stock (with PWT 8.0 code *rkna*),  $L_t$  denotes total hours worked by engaged persons (given as *emp\*avh*), and  $Q_t$  is real GDP (*rgdpna*). Four countries are considered. For the USA and UK, the observations since 1979 till 2011 are used, but the first three data points serve as initial conditions; thus, we model  $T=30$  tri-variate observations from the period 1982–2011. In the case of Poland, we model only  $T=20$  observations, covering the period 1992–2011, in order to avoid modelling pre-transformation data; the data from 1989 to 1991 are used as initial conditions. In the case of Hungary, which was a more developed economy under the communist regime, we use the data from the period 1980–2011 and focus on modelling  $T=29$  observations from 1983 to 2011. Main interest is in the results for the USA, the world leader, and Poland, an example of successful economic transformation. The results for the UK and Hungary serve as a sensitivity check—whether similar findings can be obtained for different economies.

The data are presented in Fig. 1. Due to scale differences between capital and product variables ( $k_t = \ln K_t$ ,  $q_t = \ln Q_t$ ), relatively smooth upward trends constitute a dominating picture, with fluctuations and short-term deviations being much less visible. Labour ( $l_t = \ln L_t$ ) has its own scale on the right-hand side in each chart, so its fluctuations are more visible. However, the stochastic and dynamic character of all three aggregates is obvious when we look at the results of model comparison.

The natural Bayesian characteristic of how well a model fits the data is  $p(x_1, \dots, x_T | \psi_0)$ , the marginal data density (MDD) value for the observations at hand. MDD values constitute the main ingredient to the formal Bayesian procedure of model comparison; see, for example, Osiewalski and Steel (1993). Using MDD values,

each reader can use his or her own prior distribution on the model space and then easily calculate the resulting posterior model probabilities. In order to avoid selecting prior model probabilities, we are not going to report posterior probabilities of competing specifications. Tables 1 and 2 contain the decimal logarithms of MDD values for type A and B models, respectively. Of course, these values are based on particular prior distributions on model parameters that we have proposed.

The MDD values, presented in Table 1 for different VAR( $p$ ) models ( $p=1,2,3$ ), indicate that simple specifications with non-stationarity, namely VAR(0), VAR(1) and VAR(2) without constants for the tri-variate differenced series  $\Delta x_t$ , are the best ones. The second best are VAR(1), VAR(2) and VAR(3) models for levels, with one co-integration relation and a constant restricted to this relation (or even without any constant). Thus, Bayesian co-integration analysis leads to the conclusion that the annual data on aggregate inputs and output are best described by three stochastic trends (in particular, just by a tri-variate random walk), with some possibility of only two stochastic trends and a long-run relation linking all three variables. However, three stochastic trends—precluding any long-run relation—are much more likely. Under equal prior probabilities for all models, specifications based on first differences are more than 10 times more probable a posteriori than co-integrated VAR models for levels. Also, these co-integration relations do not show any features of a regular production function. All other specifications appearing in Table 1 can be omitted in further considerations as they are unlikely in view of the data. The results presented above are strikingly similar for all four economies, and, in view of the relatively short time span of the data, they are strong even for the Polish economy, represented by 20 data points only. Table 1 is based on a very diffuse prior for the linear trend parameters, as we use  $W_0 = 100I_2$  in the prior covariance matrix for the coefficients corresponding to  $d_t = [1 \ t]$ . It is not surprising that the models with time variable  $t$  are less likely—in particular, models with an unrestricted trend become worse as the rank of  $\Pi$  increases; the worst models appear at the bottom of Table 1. Using much more precise prior for the deterministic components, with 1 and 0.01 as the diagonal elements of  $W_0$ , leads to the same general (qualitative) conclusions, but makes models with deterministic terms slightly more likely a posteriori. In our final results in the next section we take advantage of this increased role of the constant term under the more precise prior.

The MDD values for the APF (or type B) models, presented in Table 2, are based on our initial assumption that the joint tri-variate type B model always uses the first two equations from one of the best VAR( $p$ ) models. So, we have taken these equations of the VAR(1) specification for  $\Delta x_t$  (without constants) as the marginal sampling model for inputs, and we have assumed different specifications of the general form (7) as conditional models for output given current inputs. All our type B models preserve a Bayesian sequential cut by construction, so  $MDD = C_1 C_2$ —where, for example,  $\log(C_1) = 22.64$  for the USA. However, the values of  $\log(C_2)$  are so small for most type B models that we cannot treat them as competitors to the best type A models. The basic, static, full-efficiency Cobb–Douglas APF model is completely unlikely in view of the data. Extending it to the translog specification also does not seem to provide any substantial difference (with some exception to Hungary) and adding exponential inefficiency terms even worsens the initial situation, since such extension is clearly not supported by the data. The only model extension that is worth noting amounts



Table 1 continued

The USA ( $T=30$ )			Poland ( $T=20$ )			The UK ( $T=30$ )			Hungary ( $T=29$ )		
VAR( $p$ )	Case d	Rank $r$	Log(MDD)	VAR( $p$ )	Case d	Rank $r$	Log(MDD)	VAR( $p$ )	Case d	Rank $r$	Log(MDD)
1	3	2	41.84	1	3	2	22.65	1	3	2	41.61
2	3	2	41.68	2	3	2	22.50	2	3	2	41.45
3	3	2	41.59	3	3	2	22.39	3	3	2	41.35
1	5	3	41.30	1	5	3	22.20	1	5	3	41.08
1	3	3	41.01	1	2	1	21.99	1	3	3	40.83
2	5	3	40.99	1	3	3	21.97	1	2	1	40.75
1	2	1	40.98	2	5	3	21.87	2	2	2	40.73
2	2	1	40.84	2	2	1	21.84	2	2	1	40.61
3	5	3	40.82	3	2	1	21.74	3	5	3	40.54
3	2	1	40.75	3	5	3	21.69	2	3	3	40.53
2	3	3	40.73	2	3	3	21.68	3	2	1	40.53
3	3	3	40.57	3	3	3	21.52	3	3	3	40.36
1	2	2	39.49	1	2	2	20.82	1	2	2	39.32
2	2	2	39.21	2	2	2	20.55	2	2	2	39.05
3	2	2	39.07	3	2	2	20.40	3	2	2	38.91
...	...	...	...	...	...	...	...	...	...	...	...
1	1	3	33.34	1	1	3	15.15	1	1	3	33.31
2	1	3	33.10	2	1	3	14.91	2	1	3	33.08
3	1	3	32.98	3	1	3	14.78	3	1	3	32.95

$d = 1$ : unrestricted linear trend,  $d = 2$ : restricted trend (only  $r = 1, 2$ ),  $d = 3$ : unrestricted constant,  $d = 4$ : restricted constant (only  $r = 1, 2$ ),  $d = 5$ : no deterministic components; log(MDD) is the decimal log of marginal data density; under equal prior odds, a difference in log(MDD) equal to one indicates that one model is ten times more likely than the other (in view of the data)

**Table 2** Decimal logarithms of MDD values for different type B (production function) models

	The USA ( $T=30$ )				Poland ( $T=20$ )				The UK ( $T=30$ )				Hungary ( $T=29$ )			
	$u_t$	$r^*$	APF	Log(MDD)	$u_t$	$r^*$	APF	Log(MDD)	$u_t$	$r^*$	APF	Log(MDD)	$u_t$	$r^*$	APF	Log(MDD)
0	-	-	TR-t	42.25	0	-	CD-t	22.84	0	-	TR	41.10	0	-	TR	35.87
0	-	-	CD-t	42.13	0	-	TR-t	22.50	0	-	TR-t	40.77	0	-	TR-t	35.17
0	-	-	TR	41.54	0	-	TR	19.88	0	-	CD-t	40.70	0	-	CD	34.91
0	-	-	CD	37.64	0	-	CD	18.07	0	-	CD	39.22	0	-	CD-t	34.35
Exp	0.7	TR	TR	34.53	Exp	0.7	CD-t	17.49	Exp	0.7	TR	34.27	Exp	0.7	CD	30.32
Exp	0.7	CD-t	CD-t	34.13	Exp	0.7	TR-t	17.13	Exp	0.7	CD-t	33.48	Exp	0.7	TR	30.27
Exp	0.7	TR-t	TR-t	34.07	Exp	0.85	CD-t	16.67	Exp	0.7	TR-t	33.35	Exp	0.85	TR	29.37
Exp	0.85	CD-t	CD-t	33.14	Exp	0.7	TR	16.60	Exp	0.85	TR	33.32	Exp	0.85	CD	29.02
Exp	0.85	TR-t	TR-t	33.04	Exp	0.85	TR-t	16.28	Exp	0.7	CD	33.04	Exp	0.7	TR-t	29.01
Exp	0.85	TR	TR	32.99	Exp	0.85	TR	15.83	Exp	0.85	CD-t	32.28	Exp	0.7	CD-t	28.70
Exp	0.7	CD	CD	32.39	Exp	0.7	CD	15.67	Exp	0.85	TR-t	32.20	Exp	0.85	TR-t	28.06
Exp	0.85	CD	CD	31.39	Exp	0.85	CD	15.33	Exp	0.85	CD	31.95	Exp	0.85	CD-t	27.72

CD, Cobb–Douglas; TR, translog; “-t” indicates time trend in the production function;  $r^*$ , prior median of technical efficiency;  $u_t$ , type of inefficiency prior; log(MDD) is the decimal log of marginal data density; APF stands for aggregate production function; under equal prior odds, a difference in log(MDD) equal to one indicates that one model is ten times more likely than the other (in view of the data)



to adding linear trend to the basic APF (again, with the exception of Hungary), but such primitive dynamics cannot compete with the best co-integrated VAR models. We have also considered—among many other type B specifications, not mentioned in Table 2—the Cobb–Douglas APF with time-varying parameters, proposed in Koop et al. (2000). However, describing Cobb–Douglas elasticities by a certain bivariate AR(1) process has appeared to be a worse modelling strategy than just adding a linear trend.

### 6 A synthesis and final results

Our empirical results presented in the previous section suggest that a chance to obtain a good type B specification, hopefully competing with VAR (i.e. type A) models, is in richer dynamics in the APF equation. In order to achieve this goal, we extend a simple Cobb–Douglas model (with or without trend) in an old-fashioned manner, by adding an autocorrelation term. Hence, let us consider the following specification:

$$q_t = \beta_1 k_t + \beta_2 l_t + \rho(q_{t-1} - \beta_1 k_{t-1} - \beta_2 l_{t-1}) + \delta_0 + v_t. \tag{8}$$

It can be treated as a member of class B, if we appropriately extend (7) to cover autocorrelation. On the other hand, (8) results from (4) if (taking  $\delta_1 = \rho$ ) we impose the restrictions:  $\delta_2 = -\delta_1 \beta_1$ ,  $\delta_3 = -\delta_1 \beta_1$  and  $\delta_i = 0$  for  $i = 4, \dots, 3p$ . In fact, (8) can be obtained as the conditional part in the co-integrated VAR(1) model with one co-integrating relation,  $[-\beta_1 - \beta_2 1]'$  as the normalized co-integrating vector and weakly exogenous inputs:

$$\begin{cases} \Delta k_t = \gamma_{0k} + \varepsilon_{tk} \\ \Delta l_t = \gamma_{0l} + \varepsilon_{tl} \\ \Delta q_t = (\rho - 1)(q_{t-1} - \beta_1 k_{t-1} - \beta_2 l_{t-1}) + \gamma_{0q} + \varepsilon_{tq} \end{cases} \tag{9}$$

The co-integrated VAR(1) model in (9) has a very interesting interpretation. In (9) it is assumed that

- (1) All three variables are conditionally correlated (given their past), as it is in any type A model where matrix  $\Sigma$  has no zeros off the diagonal;
- (2) The inputs formation is described by two stochastic trends;
- (3) Both inputs are weakly exogenous for inference on the parameters in (8);
- (4) The level of output, given current levels of inputs and the past of all variables, is determined by the Cobb–Douglas-type relation, both through its “immediate” impact and through deviations from it, which occurred in the previous period;
- (5) The “immediate” Cobb–Douglas-type relation is a co-integrating relation as well.

Thus, the joint model (9), with its conditional part (8), reconciles the traditional Cobb–Douglas APF (“immediate” and static) with the concept of co-integration, fundamental in modern econometrics. Note that this model significantly differs from standard type A models with  $r = 1$  (see Sect. 4). Firstly, it imposes weak exogeneity of inputs. Secondly, the normalized co-integrating vector  $[-a_1 - a_2 1]'$  reflects the

conditional (contemporaneous) correlation structure between output and inputs, as  $a_1 = \beta_1$ ,  $a_2 = \beta_2$  and  $\beta = \Sigma_{11}^{-1} \Sigma_{12}$  (or the other way round: conditional correlation reflects the co-integration relation).

The joint specification (9) is of particular interest, but it may occur too parsimonious to win the competition with richer structures. However, it can be easily generalized, keeping all its crucial properties. As in the case of any type B model, the first two equations of (9)—i.e. the marginal model for inputs—can be replaced by two equations with the highest value of  $C_1$  in class A. As the VAR(2) model with  $\Pi = 0$  (assuming three stochastic trends) is one of the three best specifications of type A, it is crucial to consider its close VAR(2) alternative with only two stochastic trends and the “immediate” Cobb–Douglas-type relation as the co-integration relation:

$$\begin{cases} \Delta k_t = \Delta x_{t-1} \gamma^{(1)} + \gamma_{0k} + \varepsilon_{tk} \\ \Delta l_t = \Delta x_{t-1} \gamma^{(2)} + \gamma_{0l} + \varepsilon_{tl} \\ \Delta q_t = (\rho - 1)(q_{t-1} - \beta_1 k_{t-1} - \beta_2 l_{t-1}) + \Delta x_{t-1} \gamma^{(3)} + \gamma_{0q} + \varepsilon_{tq} \end{cases} \quad (10)$$

The first two equations in (10) give the assumed value of  $C_1$ , the first component of MDD, if both their intercepts are zero. The conditional model for output given current inputs can be written in the following form, with the Bayesian sequential cut being preserved:

$$q_t = \beta_1 k_t + \beta_2 l_t + \rho(q_{t-1} - \beta_1 k_{t-1} - \beta_2 l_{t-1}) + \Delta x_{t-1} \tilde{\delta}^* + \tilde{\delta}_0 + v_t, \quad (11)$$

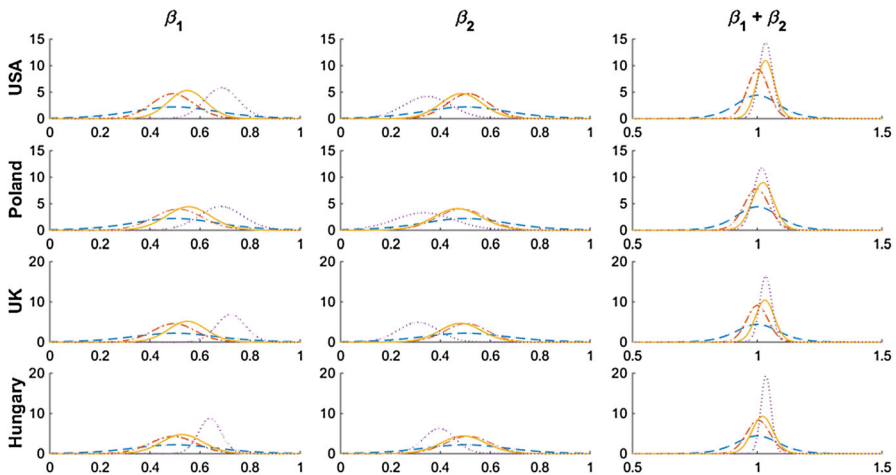
$v_t$  is a Gaussian white noise (with variance  $\omega$ ); given  $\omega$ , the parameters  $\beta$ ,  $\tilde{\delta}^*$  and  $\tilde{\delta}_0$  are a priori independent, with conditional normal prior distributions presented in Sect. 2, just below formula (6). In particular,  $\tilde{\delta}^* \sim N(0, \frac{1}{2} \omega I_3)$  and  $\tilde{\delta}_0 \sim N(0, \omega w_{00})$ . Note that the zero restriction on  $\tilde{\delta}^*$  leads to the simpler specification (8). In Table 3 we present the decimal logarithms of MDD values based on the assumed  $C_1$  and on  $C_2$  corresponding to different values of  $\rho$  in the conditional model in (11). Since no dynamic stability constraints have been imposed in type A models, we consequently do not restrict  $\rho$  to the unit interval and the highest MDD values can be obtained even for  $\rho$  slightly above 1. As they do not lead to interesting results, we do not focus on specifications with  $\rho$  free, but on models with fixed values of this parameter, namely  $\rho = 1$  (i.e. VAR with three stochastic trends) and  $\rho$  just below 1 (i.e. VAR with two stochastic trends, “immediate” Cobb–Douglas relation as the co-integration relationship and a slow adjustment). Models with and without constant are compared; if  $\tilde{\delta}_0$  is present, then  $w_{00} = 1$  is assumed in the prior distribution, because  $w_{00} = 100$  led to lower value of  $C_2$ . Our task is now to find close competitors to VAR specifications (without constants) for first differences of logs of the original data;  $w_{00} = 100$  would preclude finding them, as it is clear from Table 1.

Under equal prior probabilities, three stochastic trends are more likely a posteriori than the co-integrating relation of the Cobb–Douglas form, but by less than one order of magnitude for the USA and the UK (and even less for Hungary and Poland). If an economist has strong prior beliefs that a stable, long-run APF is an adequate and useful concept for macroeconomic analyses, and the case of three stochastic trends has only small prior probability (say, less than 0.1), then—looking at our results of model

**Table 3** Decimal logarithms of MDD values for different versions of the best models

	The USA ( $T=30$ )		Poland ( $T=20$ )		The UK ( $T=30$ )		Hungary ( $T=29$ )	
	No const.	Constant	No const.	Constant	No const.	Constant	No const.	Constant
$\rho = 0.95$	46.183	46.408	27.113	27.251	45.673	46.046	41.663	41.672
$\rho = 0.96$	46.260	46.525	27.110	27.287	45.725	46.145	41.742	41.726
$\rho = 0.97$	46.350	46.646	27.091	27.319	45.793	46.242	41.843	41.775
$\rho = 0.98$	46.470	46.760	27.050	27.344	45.902	46.327	41.982	41.815
$\rho = 0.99$	46.696	46.851	27.012	27.357	46.148	46.382	42.198	41.841
$\rho = 1$	47.330	46.898	27.185	27.351	46.795	46.393	42.489	41.847
$\rho$ free	45.940	45.995	26.490	26.541	45.476	45.532	41.033	41.087

“ $\rho$  free”—exogeneity is assumed, but co-integration parameters are not related to  $\beta_1$  and  $\beta_2$ ;  $\rho$  is integrated with the use of its  $N\left(1, \omega \sum_{i=1}^p \frac{1}{i}\right)$  conditional prior, following from the assumptions presented in Appendix; under equal prior odds, a difference in decimal logarithm of MDD (marginal data density) values equal to one indicates that one model is ten times more likely than the other (in view of the data)

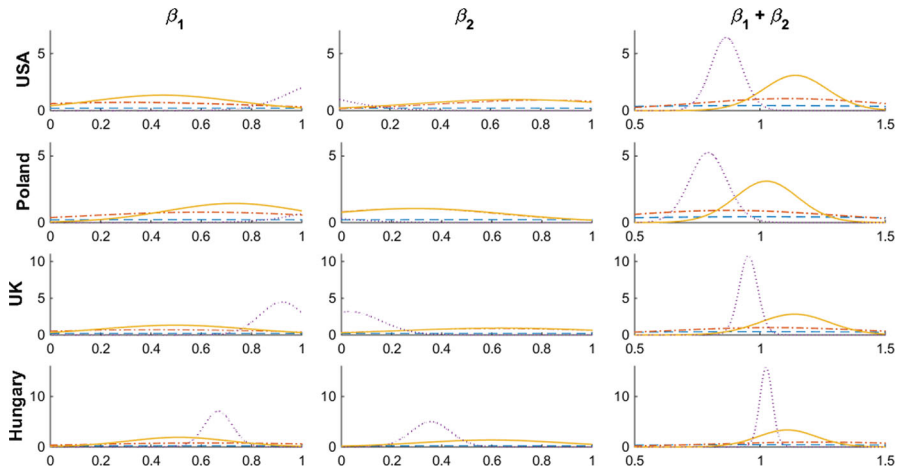


**Fig. 2** Marginal posterior densities of  $\beta_1$ ,  $\beta_2$  in (11) and their sum—informative priors; prior densities are plotted using dashed line; posterior densities are plotted using: dash-dot line for  $\rho = 1$ ; solid line for  $\rho = 0.95$ ; dotted line for  $\rho = 0.5$ ;  $\beta_1$ ,  $\beta_2$  and  $\beta_1 + \beta_2$  are capital, labour and scale elasticities, respectively; model with  $\rho = 1$  has the highest marginal data density (i.e. under equal prior odds it is the most likely of the three presented given the data); model with  $\rho = 0.5$  has the lowest marginal data density

comparison—such economist still chooses (11) with  $\rho$  just below 1 as a better model than the specification in terms of first differences alone. Now the situation is different than in the case of comparison between pure type A models (see Table 1), where we did not impose any extra restrictions on the specifications with one co-integration relation and they were more than one order of magnitude worse than the models assuming three stochastic trends. As we can see, the additional assumptions (weak exogeneity of inputs and equivalence between the co-integration relation and the “immediate” Cobb–Douglas relation, together with fixing  $\rho$  just below 1 and introducing the constant term of high prior precision), imposed jointly, are helpful to make the co-integration VAR model a competitor to the VAR model for first differences.

Finally, in Figs. 2 and 3 we present plots of the marginal prior and posterior densities of  $\beta_1$ ,  $\beta_2$  and their sum; the posterior densities correspond to three values of  $\rho$ : namely  $\rho = 1$ ,  $\rho = 0.95$  and  $\rho = 0.5$ . The value  $\rho = 0.99$  (that yields a particularly good model) leads to almost the same posterior densities as  $\rho = 1$ , so there is no point in plotting them separately. Figure 2 corresponds to our original prior specification for the parameters of (11), and Fig. 3 is based on 100 times smaller prior precision for  $\beta_1$ ,  $\beta_2$  and their sum.

As our model specification gets better (i.e. the closer we are to the case  $\rho = 1$ , which amounts to three stochastic trends and lack of any co-integration relation), the marginal posterior distributions of  $\beta_1$ ,  $\beta_2$  and their sum become very similar to the marginal prior distributions. This empirical result, the same for all four different data sets, indicates that our inference—on the parameters that have clear interpretation within the APF context—is very fragile and based mainly on prior assumptions, if we restrict to relevant dynamic models. For other specifications, which we can treat as completely inadequate, our posterior inference seems to be driven by the data.



**Fig. 3** Marginal posterior densities of  $\beta_1$ ,  $\beta_2$  in (11) and their sum—diffuse priors; prior densities are plotted using dashed line; posterior densities are plotted using: dash-dot line for  $\rho = 1$ ; solid line for  $\rho = 0.95$ ; dotted line for  $\rho = 0.5$ ;  $\beta_1$ ,  $\beta_2$  and  $\beta_1 + \beta_2$  are capital, labour and scale elasticities, respectively; model with  $\rho = 1$  has the highest MDD (i.e. under equal prior odds it is the most likely of the three presented, given the data); model with  $\rho = 0.5$  has the lowest marginal data density

However, any conclusions based on a highly misspecified model can be misleading. It seems that, for our data covering the decades around the year 2000, the APF cannot be confirmed even as a purely empirical relation explained in Shaikh (1974). The dynamics of the output and inputs aggregates is best described by a tri-variate random walk.

## 7 Concluding remarks

The concept of modelling an economy-wide output via aggregate production function, APF, although fundamentally questionable and criticized, has been very important for both the theory of economic growth and empirical growth studies (e.g. growth accounting, development accounting). However, the studies have not focused on the empirical justification of APF within the modern dynamic econometrics. For this reason, we have used annual data on total production and two input aggregates in order to formally compare basic time-series and more traditional models. The crucial issue in modelling this tri-variate, non-stationary time series is to adequately capture its dynamics. We start within the VAR framework and stress that the “instantaneous” or “short-run” Cobb–Douglas-type relation appears through an obvious re-parameterization and representation of VAR in terms of the conditional and marginal models—for the output and inputs, respectively. The important empirical question is whether such APF-looking relation is also a co-integration relation (linking aggregate output with aggregate inputs) based on modern econometric techniques. If so, it could be treated as a valid tool to describe such links and interpreted at least as in Shaikh (1974).

For the purpose of formulation and comparison of our competing models we have devised a fully Bayesian modelling and inference framework. The framework is rooted in both a purely statistical, time-series analysis of the data and the macroeconomic theory of APF. This makes it particularly appealing to practitioners and statistical agencies who seek a well-founded framework for modelling the aggregate output. Furthermore, since Bayesian model comparison is sensitive to prior distributions, we have conducted our research carefully, keeping the same Bayesian Cobb–Douglas model in the two general classes (A and B); this model appears as a special case in class A, and it is the starting point for extensions in class B. In order to make the comparison fully operational, reduce the computational burden and avoid the possibility of critiques from the purely numerical perspective, we have restricted to the simplest prior distribution classes, like the natural conjugate one in the case of basic VAR models. Of course, other Bayesian models considered (co-integrated VAR models, stochastic frontier specifications) required much more effort, but they were built around the basic structures.

In fact, we have favoured a priori the regular, linearly homogenous Cobb–Douglas relation by using it to partially elicit the prior for the VAR covariance matrix. Despite the central role of the Cobb–Douglas case in our Bayesian modelling strategy, posterior results have not confirmed the empirical relevance of the APF-type relation in modelling yearly input and output data from individual economies. It is worth noting that we have also explored possible extensions to traditional Cobb–Douglas, translog in particular. However, an APF based on the translog specification has turned out to be less probable a posteriori.

Among all specifications under consideration, the best ones are VAR(0), VAR(1) and VAR(2) models for the logarithmic growth rates of output and inputs aggregates. Thus, our tri-variate series is (most likely) best described by three stochastic trends with only “instantaneous” Cobb–Douglas-type relation, reflecting possible conditional correlation between current output and inputs. The second best specifications are the specially constructed models with the error correction term in the output equation only. Therefore, less likely than the three stochastic trends, but still probable, are two stochastic trends, which are responsible for the dynamic formation of exogenous inputs, along with the co-integrating Cobb–Douglas-type relation, the same as the “instantaneous” one.

This main result of model comparison is not totally conclusive (in favour of three stochastic trends), which is rather obvious given the relatively short observation periods. However, the empirical role of APF-type relation is further diminished by poor identification of its crucial economic characteristics within the best specifications with co-integration; i.e. the better the model is, the less informative our posterior inference. Our results are strikingly similar for all four data sets—for the leading developed economies (the USA and the UK), as well as for the two transforming economies (Poland and Hungary)—which makes the findings more interesting.

An extension of this study could, perhaps, consider panel data or longer time frames. In particular, panel data with their cross section component can bring a new dimension of aggregate product’s variation which could be relevant in examining the APF. Nonetheless, the results presented in this study suggest that APF may be difficult to support empirically, at least if we look at it from dynamic econometrics perspective.

This can be an additional argument in the debate summarized in Felipe and McCombie (2013).

**Acknowledgements** This work was supported by research funds granted to the Faculty of Management at Cracow University of Economics, within the framework of the subsidy for the maintenance of research potential.

**Data and computer code availability** The data used in the study can be downloaded from Penn World Tables. Information about exact variables and database version used is provided in Sect. 5. Computer code has been written in MATLAB and GAUSS and can be made available upon request.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

### Appendix: Bayesian co-integration analysis

In order to introduce prior distributions employed in the analysis of VEC models, we rewrite representation (2b)  $(x_t = x_{t-1} + \tilde{z}_t \tilde{\Phi} + \varepsilon_t)$  as

$$\Delta x_t = \tilde{z}_t \tilde{\Phi} + \varepsilon_t = x_{t-1} \Pi + \tilde{z}_t^2 \Gamma + \varepsilon_t, \tag{12}$$

where  $\Gamma = [\Gamma'_1, \Gamma'_2, \dots, \Gamma'_{p-1}, \Gamma'_0]'$  and  $\tilde{z}_t^2 = [\Delta x_{t-1}, \Delta x_{t-2}, \dots, \Delta x_{t-(p-1)}, d_t]$ . In the case of co-integration, the matrix  $\Pi$  is of reduced rank ( $r(\Pi) = r, r = 1, 2$ ) and can be decomposed as the product of two full column rank matrices. Following the idea of Koop et al. (2010) to perform the analysis, we consider two equivalent parameterizations of  $\Pi$ :

$$\beta \alpha' \equiv B_\Pi A'_\Pi \tag{13}$$

where  $\beta$  has orthonormal columns, whereas  $B_\Pi \in \mathbb{R}^{(3+l)r}$  ( $l = 1$  when a constant or a linear trend is added to the co-integration relation,  $l = 0$  when there are not any deterministic components in the co-integration relations). As  $\beta$  ( $B_\Pi$ ) and  $\alpha$  ( $A_\Pi$ ) may have different dimensions we start (as suggested by Koop, León-González and Strachan, 2010) with the  $A_\Pi, B_\Pi$  parameterization.

$$\Delta x_t = x_{t-1} B_\Pi A'_\Pi + \tilde{z}_t^2 \Gamma + \varepsilon_t = (x_{t-1} B_\Pi \tilde{z}_t^2) \begin{pmatrix} A'_\Pi \\ \Gamma \end{pmatrix} + \varepsilon_t = (x_{t-1} B_\Pi \tilde{z}_t^2) \Gamma_A + \varepsilon_t, \tag{14}$$

We have decided to assume that  $B_\Pi$  and  $\Gamma_A$  are a priori independent and impose the following matrix normal distributions for the considered parameters:

$$\begin{aligned} \text{vec}(\mathbf{B}_\Pi) &\sim N\left(0, \frac{1}{3+l} I_{(3+l)r}\right), \\ \text{vec}(\Gamma_A) &\sim N\left(0, \Sigma \otimes \tilde{W}_A\right), \end{aligned} \tag{15}$$

where  $\tilde{W}_A = M\tilde{W}M'$ , with  $M = \begin{bmatrix} M_{11} & 0 \\ 0 & I \end{bmatrix}$   $M_{11} = [I_r \ 0_{r \times (3-r)}]$ .

From  $\text{vec}(\mathbf{B}_\Pi) \sim N(0, \frac{1}{3+l} I_{(3+l)r})$  follows that the orientation of  $\mathbf{B}_\Pi$  (i.e.  $\mathbf{B}_\Pi(\mathbf{B}'_\Pi \mathbf{B}_\Pi)^{-\frac{1}{2}}$ ) is uniformly distributed over the Stiefel manifold (see, for example, Chikuse 1990) and so is the space spanned by this matrix (which is the element of the Grassmann manifold). In other words, by such formulated prior distribution we impose uniform prior over the set of all possible co-integration spaces, we do not favour any direction. Keeping that in mind and knowing that the uniform distribution over the Stiefel manifold is right orthogonal invariant, we can, without loss of generality, assume that  $\beta = \mathbf{B}_\Pi(\mathbf{B}'_\Pi \mathbf{B}_\Pi)^{-\frac{1}{2}} \mathbf{O}_r$  and  $\alpha = A_\Pi(\mathbf{B}'_\Pi \mathbf{B}_\Pi)^{\frac{1}{2}} \mathbf{O}_r$ , where the  $r \times r$  matrix  $\mathbf{O}_r = \text{diag}(\pm 1)$  is constructed in such a way that it controls the element of the first row of  $\beta$  to be positive, so it helps us to deal with the many-to-one relationship between the Stiefel and the Grassmann manifolds.

Summing up our prior assumptions we obtain the following joint prior density function:

$$\begin{aligned} p(\mathbf{B}_\Pi, \Gamma_A, \Sigma) &= p(\mathbf{B}_\Pi)p(\Gamma_A|\Sigma)p(\Sigma) = c_1 \exp\left\{-\frac{1}{2} \text{tr}[(3+l)\mathbf{B}'_\Pi \mathbf{B}_\Pi]\right\} \\ &\times c_2 |\Sigma|^{-\frac{r+3(p-1)+m}{2}} \exp\left\{-\frac{1}{2} \text{tr}\left[\Sigma^{-1} \Gamma'_A \tilde{W}_A^{-1} \Gamma_A\right]\right\} \times c_3 |\Sigma|^{-\frac{s+3+l}{2}} \exp\left\{-\frac{1}{2} \text{tr}\left[\Sigma^{-1} A\right]\right\} \\ &= c |\Sigma|^{-\frac{s+r+m+3(p-1)+4}{2}} \exp\left\{-\frac{1}{2} \text{tr}\left[\Sigma^{-1} (A + \Gamma'_A \tilde{W}_A^{-1} \Gamma_A)\right]\right\} \exp\left\{-\frac{1}{2} \text{tr}[(3+l)\mathbf{B}'_\Pi \mathbf{B}_\Pi]\right\} \end{aligned} \tag{16}$$

where  $|C|$  denotes the determinant of  $C$ ,  $\text{tr}(C)$ —the trace of  $C$ ,  $c_1, c_2$ —the normalizing constants of the matrix-variate normal distribution,  $c_3$ —the normalizing constant of the inverted Wishart distribution, and

$$\begin{aligned} c &= c_1 c_2 c_3 = (2\pi)^{-\frac{(3+l)r}{2}} \left|\frac{1}{3+l} I_{(3+l)r}\right|^{-\frac{1}{2}} (2\pi)^{-\frac{(r+3(p-1)+m)3}{2}} \left|\tilde{W}_A\right|^{-\frac{3}{2}} |A|^{\frac{s}{2}} 2^{-\frac{3s}{2}} \pi^{-\frac{3(3-1)}{4}} \\ &\times \left[\prod_{i=1}^3 \Gamma\left(\frac{s+1-i}{2}\right)\right]^{-1} = (2\pi)^{-\frac{(6+l)r+3(3(p-1)+m)}{2}} 2^{-\frac{3s}{2}} \pi^{-\frac{3}{2}} (3+l)^{\frac{(3+l)r}{2}} \\ &\times \left[\prod_{i=1}^3 \Gamma\left(\frac{s+1-i}{2}\right)\right]^{-1} \left|\tilde{W}_A\right|^{-\frac{3}{2}} |A|^{\frac{s}{2}}. \end{aligned}$$



Combining the joint prior with the likelihood density function leads us to the joint posterior:

$$p(B_{\Pi}, \Gamma_A, \Sigma|X) = \tilde{c}|\Sigma|^{-\frac{s+r+m+3(p-1)+4+T}{2}} \exp\left\{-\frac{1}{2}tr\left[\Sigma^{-1}\left(A + \Gamma'_A \tilde{W}_A^{-1} \Gamma_A + E'E\right)\right]\right\} \\ \times \exp\left\{-\frac{1}{2}tr\left[(3+l)B'_{\Pi}B_{\Pi}\right]\right\}, \tag{17}$$

where  $\tilde{c} = c(2\pi)^{-\frac{3T}{2}}$  and  $E = (\varepsilon'_1 \ \varepsilon'_2 \ \dots \ \varepsilon'_T)'$ . To obtain MDD we have to integrate out  $B_{\Pi}$ ,  $\Gamma_A$  and  $\Sigma$  from the likelihood according to the imposed priors. The parameters  $\Gamma_A$  and  $\Sigma$  can be integrated out analytically, which leads to the following conditional (on  $B_{\Pi}$ ) data density:

$$p(y|B_{\Pi}) = (2\pi)^{-\frac{3T}{2}} \prod_{i=1}^3 \frac{\Gamma\left(\frac{T+s+1-i}{2}\right)}{\Gamma\left(\frac{s+1-i}{2}\right)} \left|\tilde{W}_A\right|^{-\frac{3}{2}} |A|^{\frac{3}{2}} \left|\tilde{W}_A^{-1} + Z'_B Z_B\right|^{-\frac{3}{2}} |A + R_A + D_A|^{-\frac{s+T}{2}},$$

where  $Z_B = (Z_1 B Z_2)$ ,  $R_A = (Z_0 - Z_B \hat{\Gamma}_A)'(Z_0 - Z_B \hat{\Gamma}_A)$ ,  $\hat{\Gamma}_A = (Z'_B Z_B)^{-1} Z'_B Z_0$ ,  $D_A = \tilde{\Gamma}'_A \tilde{W}_A^{-1} (\tilde{W}_A^{-1} + Z'_B Z_B)^{-1} Z'_B Z_B \hat{\Gamma}_A$ ,  $Z_0 = (\Delta x'_1 \ \Delta x'_2 \ \dots \ \Delta x'_T)'$ ,  $Z_1 = (x'_0 \ x'_1 \ \dots \ x'_{T-1})'$ ,  $Z_2 = (\tilde{z}'_1 \ \tilde{z}'_2 \ \dots \ \tilde{z}'_T)'$ . To integrate out  $B_{\Pi}$  we have to use Monte Carlo methods, mainly the arithmetic mean estimator.

**References**

Aliche A, Bizimana A, Laxton D et al (2017) Multivariate filter estimation of potential output for the United States. IMF Working Paper WP/17/106

Chikuse Y (1990) The Matrix Angular Central Gaussian distribution. J Multivar Anal 33:265–274

Doan T, Litterman RB, Sims CA (1984) Forecasting and conditional projection using realistic prior distributions. Econom Rev 3:1–100

Felipe J, McCombie JSL (2013) The aggregate production function and the measurement of technical change. Edward Elgar, Cheltenham

Fisher FM (1969) The existence of aggregate production functions. Econometrica 37:553–577

Florens J-P, Mouchart M (1985) Conditioning in dynamic models. J Time Ser Anal 6:15–34

Growiec J (2008) A new class of production functions and an argument against purely labor-augmenting technical change. Int J Econ Theory 4:483–502

Growiec J (2013) A microfoundation for normalized CES production functions with factor-augmenting technical change. J Econ Dyn Control 37:2336–2350

Jones CI (2005) The shape of production functions and the direction of technical change. Q J Econ 120:517–549

Koop G, Ley E, Osiewalski J, Steel MFJ (1997) Bayesian analysis of long memory and persistence using ARFIMA models. J Econom 76:149–169

Koop G, Osiewalski J, Steel MFJ (1999) The components of output growth: a stochastic frontier analysis. Oxf Bull Econ Stat 61:455–487

Koop G, Osiewalski J, Steel MFJ (2000) Modelling the sources of output growth in a panel of countries. J Bus Econ Stat 18:284–299

Koop G, Strachan R, Dijk H, Villani M (2004) Bayesian approaches to cointegration. In: Mills TC, Patterson K (eds) The Palgrave handbook of econometrics, vol 1. Econometric theory. Palgrave-Macmillan, Basingstoke, pp 871–898

- Koop G, León-González R, Strachan R (2010) Efficient posterior simulation for cointegrated models with priors on the cointegration space. *Econom Rev* 29:224–242
- Litterman RB (1986) Forecasting with Bayesian vector autoregressions—five years of experience. *J Bus Econ Stat* 4:25–38
- Makiela K (2014) Bayesian stochastic frontier analysis of economic growth and productivity change in the EU, USA, Japan and Switzerland. *Cent Eur J Econ Model Econom* 6:193–216
- Osiewalski J, Steel MFJ (1993) Una perspectiva bayesiana en selección de modelos. *Cuadernos Economicos* 55/3:327–351 (original English version available at: <http://www.cyfronet.krakow.pl/~eeosiewa/pubo.htm>)
- Osiewalski J, Steel MFJ (1996) A Bayesian analysis of exogeneity in models pooling time series and cross sectional data. *J Stat Plan Infer* 50:187–206
- Osiewalski J, Steel MFJ (1998) Numerical tools for the Bayesian analysis of stochastic frontier models. *J Prod Anal* 10:103–117
- Pajor A (2017) Estimating the marginal likelihood using the arithmetic mean identity. *Bayesian Anal* 12:261–287
- Shaikh A (1974) Laws of production and laws of algebra: the humbug production function. *Rev Econ Stat* 56:115–120
- Turner D, Cavalleri M, Guillemette Y et al (2016) An investigation into improving the real-time reliability of OECD output gap estimates. *OECD Working Paper ECO/WKP(2016)18*
- van den Broeck J, Koop G, Osiewalski J, Steel MFJ (1994) Stochastic frontier models: a Bayesian perspective. *J Econom* 61:273–303
- Wróblewska J (2009) Bayesian model selection in the analysis of cointegration. *Cent Eur J Econ Model Econom* 1:57–69