BANK OF ENGLAND

Staff Working Paper No. 826

Trend and cycle shocks in Bayesian unobserved components models for UK productivity

Marko Melolinna and Máté Tóth

September 2019

BANK OF ENGLAND

Staff Working Paper No. 826

Trend and cycle shocks in Bayesian unobserved components models for UK productivity

Marko Melolinna,[1] and Máté Tóth[2]

Abstract

This paper presents a range of unobserved components models to study productivity dynamics in the United Kingdom. We introduce a set of univariate and bivariate models that allow for shocks between the trend and the cycle to be correlated, and use Bayesian sampling techniques to estimate the models. We show that the size of the priors on the trend and cycle shock has an effect on the results, suggesting that a range of priors need to be considered for policy-making purposes. If the prior is set to a smooth trend, then models with little correlation between the trend and cycle shocks are the likeliest to fit the data. On the other hand, if there is a prior belief that the trend shock is allowed to vary relatively freely, the results suggest that there is a negative correlation between trend and cycle shocks to UK productivity. This is consistent with real-business cycle type narratives, where trend shocks are the main driver of productivity dynamics. Finally, our evidence suggests that the trend productivity growth rate in the UK has been weaker since the financial crisis. There is also a significant positive correlation between shocks to UK trend productivity and those of other advanced economies.

Key words: Business cycle, Markov Chain Monte Carlo, productivity puzzle.

JEL classification: C11, C32, E32.

(1) Bank of England. Email: marko.melolinna@bankofengland.co.uk
(2) European Central Bank. Email: mate.toth@ecb.int

The views expressed in this paper are those of the authors, and not necessarily those of the Bank of England or its committees, or the European Central Bank. We are grateful to George Kapetanios, Patrick Schneider, an anonymous referee and participants at a Bank of England Research seminar and at the 2019 Asia Meeting of the Econometric Society in Seoul, Korea, for comments on an earlier version of this paper.

The Bank's working paper series can be found at www.bankofengland.co.uk/working-paper/staff-working-papers

Bank of England, Threadneedle Street, London, EC2R 8AH
Email publications@bankofengland.co.uk

ISSN 1749-9135 (on-line)

1 Introduction

Labour productivity growth in the UK has been weak since the financial crisis; not only has there been a downward shift in the level of productivity, the growth rate of productivity has also struggled to reach its pre-crisis pace. Hence, it is unsurprising that this recent weakness in UK productivity - or "productivity puzzle" - has been analysed in a number of studies using a wide range of approaches. The existing literature has mainly concentrated on a production function-based approach for identifying the relevant drivers of productivity growth in the UK, often using firm-level data. However, econometric methods to measure the effects of different types of shocks on aggregate productivity, or to compare aggregate dynamics across countries have been applied less widely. This study aims to make a contribution to this literature by examining unobserved (transitory) cyclical and (permanent) trend components of UK productivity dynamics over time, allowing for the relative importance of these factors to be driven by the data dynamics. We focus on labour productivity because it is straightforward to calculate and one does not need to make (possibly incorrect) assumptions on the functional form of the production process. We compare models with different specifications for the trend component. We then construct narratives on UK productivity dynamics based on the likeliest models, and analyse their forecasting performance. Furthermore, we study indicators of global productivity dynamics to see how they correlate with shocks to UK productivity dynamics.[1]

Methodologically, we use a framework of uncorrelated and correlated unobserved components (UC) models. This framework has the advantage of making the importance of allowing for correlation between the trend and the cycle shocks a testable empirical question. While the more traditional non-correlated UC models do not allow for this cross-correlation, the correlated UC model relaxes this assumption. This makes it potentially ideal for enriching the analysis of both UK-specific trend and cycle components as well as studying correlations of UK productivity trend and cycle components with those of other variables. The study will use a single-variable specification as a benchmark, and then extend to a bivariate framework, where productivity shocks in other economies can be correlated with productivity shocks in the UK.

We also conduct a sensitivity analysis on the priors on the trend and cycle shocks in our models. This has largely been ignored in previous literature, but as we show, it can be crucial for the posterior results in a Bayesian setting. We consider an example of setting the relative size of the trend and cycle shock priors based on results from a structural time series model, but the more general point is the need to form a prior view on the shock processes before using these types of models. We also show a potential way of informing the priors, based on previous forecasting performance of the models, as well as some evidence from a Monte Carlo simulation.

[1] This is, of course, not to say that other factors could not be relevant for UK productivity dynamics. We want to concentrate on a small selection of key variables for which the data needed for the analysis is readily available. We leave the relevance of other variables to other studies and approaches.

The main results of the study are the following. The most important insight is that the relative size of the priors on the standard deviation of trend and cycle shocks has an effect on the results, so the motivation of this prior is crucial for using these types of models. If the prior is set to be consistent with a smooth trend, then non-correlated UC models, with no parameterised correlation between the trend and cycle shocks, or correlated models with moving average smoothed trends, are found to be the likeliest based on their marginal data density. From a forecasting perspective, for our data on UK productivity, models with smooth trends are preferable. Furthermore, a Monte Carlo experiment with simulated data suggests that in the absence of knowledge of the true data-generating process, priors with smooth trends are a safer choice to capture the true cyclical component. Nevertheless, using results from a structural VAR approach to inform our priors, we find evidence for relatively volatile trend shocks. If one takes this prior belief to the UC models, then by far the most likely models, given that data, are generally the ones allowing for correlations between the trend and the cycle shocks, rather than traditional non-correlated UC models. In the correlated UC models, the results also suggest that there is a negative correlation between trend and cycle shocks to UK productivity. This is consistent with the narrative of real shocks being the dominant force in driving productivity dynamics. The likeliest models imply substantially weaker trend growth since the financial crisis. There is also evidence to suggest that there is a significant positive correlation between shocks to UK trend productivity and those of other advanced economies. These positive correlations between trends appear to have become stronger since the financial crisis, which is consistent with the view of synchronised global shocks affecting macroeconomic dynamics more than before the crisis.

Literature Review. The current study links to both empirical literature studying recent productivity dynamics, as well as theoretical literature that has developed the UC modelling framework. In terms of the former, the crux of the matter has typically been whether the productivity puzzle is mainly attributable to permanent or temporary factors. Oulton and Sebastia-Barriel (2017) emphasise the long-term effects of financial factors in a cross-country panel study. Barnett et al. (2014a) suggest that while temporary factors, like labour hoarding, could explain the most of the puzzle in the aftermath of the financial crisis, structural factors were more important later on. Barnett et al. (2014b), using a model for firm-level contributions to productivity growth, attribute a large proportion of the productivity puzzle to sluggish reallocation of resources between firms and sectors, while Riley et al. (2015) emphasise the role played by an adverse credit shock in causing frictions in the resource reallocation process. Using a decomposition of sector-level data up to 2011, Goodridge et al. (2014) conclude that the productivity puzzle is mainly driven by a total factor productivity rather than a capital or labour shock. Overall, the literature tends to find evidence more in favour of long-term structural rather than cyclical explanations for the productivity puzzle in the UK, although the relative importance of different factors has probably shifted in different periods after the financial crisis.

A strand of the productivity literature has also looked at links in productivity dynamics across different countries. In particular, a number of recent studies have investigated the prospects of a catching up process between non-frontier and frontier economies and firms, some of them also using data for the UK. Using a firm-level cross country dataset, Andrews et al. (2016) find that productivity divergence between countries has mainly been driven by weaker diffusion from frontier to laggard firms rather than frontier firms' productivity growth slowing. Bartelsman et al. (2008) conclude that UK firms were able to learn from the domestic frontier, even if they were not able to catch up with global frontier firms. Bartelsman et al. (2013) report significant cross-country variation in the correlation between firm size and productivity, while Bergeaud et al. (2016a) and (2016b) study the drivers of productivity convergence across large advanced economies over longer time periods.

In the methodological literature related to our paper, traditionally, correlated UC models have been used to study GDP dynamics either in a one-variable (see, for example, Morley et al. (2003), Morley (2007) and Weber (2011)) or a multiple-variable setting (see Basistha (2007), Trenkler and Weber (2016), Sinclair (2009) and Mitra and Sinclair (2012)). The latter study has some relevance for the current study, as it decomposes GDP dynamics in seven advanced economies into trend and cycle components, finding trend components to be highly variable and correlations between trend and cycle components to be important both within and across countries. Given the exceptionally weak UK productivity dynamics in a world with synchronised business cycles, one interesting question is the correlation of UK productivity shocks with those of other countries.

One important difference between most studies using correlated UC models and our study is that we use Bayesian rather than maximum likelihood methods for estimation. This allows us to take into account prior information in an efficient way as well as to make the estimation more stable. Typically, maximum likelihood methods can lead to difficulties in finding solutions for the models when the number of parameters to be estimated is large relative to the number of observations (as is often the case in correlated UC models). In doing so, we use recent techniques on Bayesian methods for correlated UC models developed by Chan and Eisenstat (2015), Chan and Grant (2016) and Grant and Chan (2017a) and (2017b). Furthermore, we also present some extensions and modifications to the latter modelling framework by allowing for different specifications for the trend component of the models.

The rest of the paper is organised as follows. Section 2 introduces the theoretical framework used in the analysis, including the estimation methodology. Section 3 presents the empirical estimation results and finally, Section 4 concludes. Most of the technical details are covered in the appendices.

2 Modelling framework

2.1 Univariate models

Traditionally, unobserved components (UC) models – whether simple univariate or more complex structural ones – assume a particular structure for splitting output (or, in our case, productivity, i.e. output per 1 unit of labour input) dynamics into a cycle and a trend component.

In the simplest form of the univariate model, the split of the observable variable, (productivity, y_t) into unobservable trend and cycle components (τ_t and c_t, respectively) is as follows:

$$y_t = \tau_t + c_t \tag{1}$$

with the following dynamics for the trend and the cycle components:

$$\begin{aligned} \tau_t &= \mu + \tau_{t-1} + \eta_t \\ c_t &= \phi_1 c_{t-1} + \phi_2 c_{t-2} + \epsilon_t \end{aligned} \tag{2}$$

with i.i.d. error terms η_t and ϵ_t. The error terms have the following variance-covariance matrix:

$$\Sigma_{\epsilon\eta} = \begin{bmatrix} \sigma_\eta^2 & \sigma_{\eta\epsilon} \\ \sigma_{\epsilon\eta} & \sigma_\epsilon^2 \end{bmatrix} \tag{3}$$

The formulation of equation (2) sets a random walk with deterministic drift (μ) for the trend component, as well as an AR(2) structure for the cycle component. The number of lags required in the cycle component can be tested empirically (although typically, a longer lag structure than AR(2) does not improve the fit).

The key feature of this framework that we want to concentrate on is the form of the variance-covariance matrix. UC models traditionally only allow for a diagonal matrix with orthogonal trend and the cycle shocks (i.e., $\sigma_{\epsilon\eta} = 0$). However, the key distinguishing feature of the correlated UC model compared to a standard UC model is to relax this assumption and allow for the cross-correlation to be non-zero. In our framework, we allow for both non-correlated and correlated model versions, and then study the dynamics and posterior likelihoods of the models.[2]

We also allow for a richer trend structure in another version of the model.[3] We note that the model in equation (2) (called models **NC** and **C** in Table 1 below) is fairly restrictive,

[2]In practice, we parameterise σ_η, σ_ϵ and the correlation coefficient $\rho = \sigma_{\eta\epsilon}/\sigma_\eta\sigma_\epsilon$. All results below are reported for these parameters.

[3]We also experimented with models where drift term shocks are not allowed to have permanent effects on the trend. For example, Lewis and Vazquez-Grande (2017) use a structure where the drift term shock has a decaying pattern. However, as these models turn out to be no likelier (with the criteria set out below) nor provide dramatically different results, we do not explore these models any further here.

as it imposes a trend with a constant slope on the data. This restriction can be relaxed by allowing for a time-varying trend with the following structure (models **NC-2M** and **C-2M**):

$$\begin{aligned} \tau_t &= \mu_{t-1} + \tau_{t-1} + \eta_t \\ \mu_t &= \mu_{t-1} + \omega_t \end{aligned} \tag{4}$$

where μ_t is a random walk drift term for the trend and ω_t is an i.i.d. error term.

Table 1 shows a summary of the main selection of univariate models, for which we report results below. The models allowing for time-varying trend drift are abbreviated with 2M, as they are effectively 2-period moving averages of the trend.

Table 1: univariate models

Model	**Abbreviation**
non-correlated errors	NC
correlated errors	C
random walk trend drift, non-correlated errors	NC-2M
random walk trend drift, correlated errors	C-2M

2.2 Bivariate models

For the bivariate models, we complement the univariate case with three additional observable variables, modelling each one of them as the second variable in turn. These variables are US productivity and a proxy for G7 productivity (see next section for details on the data). With the inclusion of these observables we aim to capture in highly stylised way the effect of the global productivity frontier and technology diffusion on UK productivity growth. The benchmark cases, corresponding to univariate model **C**, are called **C-US,** and **C-Gprod** (see Table 2 below). Obviously, since we are interested in correlations between the two observable variables in the models, non-correlated UC models are not applicable in the bivariate setup.

Table 2: bivariate models

Model	**Abbreviation**
correlated errors, UK and US productivity	C-US
correlated errors, UK and global productivity proxy	C-Gprod
random walk trend drift, correlated errors, UK and US productivity	C-2M-US
random walk trend drift, correlated errors, UK and global productivity	C-2M-Gprod

The basic structure for the bivariate model is exactly the same as for the univariate one. However, in this case, we stack the observable and unobservable vectors in equations (1) and (2) by time, i.e.

$$\mathbf{y}_t = \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix}, \boldsymbol{\tau}_t = \begin{bmatrix} \tau_{1t} \\ \tau_{2t} \end{bmatrix} \text{ and } \mathbf{c}_t = \begin{bmatrix} c_{1t} \\ c_{2t} \end{bmatrix}$$

The structure for the covariance matrix becomes slightly more complicated than in the univariate case. In particular, the covariance matrix for the trend and cyclical shocks in the bivariate model (where the two variables are denoted by y1 (UK) and y2) is the following:

$$E\left(\begin{bmatrix} \eta_t \\ \epsilon_t \end{bmatrix} \begin{bmatrix} \eta_t & \epsilon_t \end{bmatrix}\right) = \begin{bmatrix} \Sigma_\eta & \Sigma_{\eta\epsilon} \\ \Sigma_{\epsilon\eta} & \Sigma_\epsilon \end{bmatrix} \tag{5}$$

where Σ_η is the covariance matrix for the trend component:

$$\Sigma_\eta = \begin{bmatrix} \sigma^2_{\eta 1} & \sigma_{\eta 1 \eta 2} \\ \sigma_{\eta 1 \eta 2} & \sigma^2_{\eta 2} \end{bmatrix} \tag{6}$$

Σ_ϵ is the covariance matrix for the cycle component:

$$\Sigma_\epsilon = \begin{bmatrix} \sigma^2_{\epsilon 1} & \sigma_{\epsilon 1 \epsilon 2} \\ \sigma_{\epsilon 1 \epsilon 2} & \sigma^2_{\epsilon 2} \end{bmatrix} \tag{7}$$

and $\Sigma_{\eta\epsilon} = \Sigma'_{\epsilon\eta}$ is the covariance matrix for cross-covariance terms between the trend and the cycle components across the two variables:

$$\Sigma_{\eta\epsilon} = \begin{bmatrix} \sigma_{\eta 1 \epsilon 1} & \sigma_{\eta 1 \epsilon 2} \\ \sigma_{\eta 2 \epsilon 1} & \sigma_{\eta 2 \epsilon 2} \end{bmatrix} \tag{8}$$

Based on Grant and Chan (2017b), we use the following version of the 2M (time-varying trend) model (models **C-2M-US, C-2M-Gprod**):

$$\Delta\boldsymbol{\tau}_t = \Delta\boldsymbol{\tau}_{t-1} + \boldsymbol{\eta}_t \tag{9}$$

which can be written as:

$$\boldsymbol{\tau}_t = 2\boldsymbol{\tau}_{t-1} - \boldsymbol{\tau}_{t-2} + \boldsymbol{\eta}_t \tag{10}$$

and - as shown in Grant and Chan (2017b) - this is equivalent to (4) when the variance of the error term η_t is set to 0. Hence, (10) should be viewed as the bivariate version of the 2M model, with a form that facilitates the estimation process.

In order to conduct the Kalman filtering exercise, which will yield the path of unobservable variables, the UCM needs to be cast in state-space form and its parameters have to be set up. A more specific presentation of the state space equations is set out in Appendix A.

2.3 Setup for correlations between trend and cycle shocks

The correlations between the trend and the cycle shocks are the key feature of the modelling framework. In particular, this allows for setting up some hypotheses for the different combinations of the shocks. It also relates to the choice of prior narrative for the smoothness of the trend, as discussed below. The different cases for the interactions between the trend and cycle shocks are summarised in Table 3 and explained in more detail below. However, when interpreting these cases, it is important to keep in mind that the results exhibit dominant average patterns rather than definite proof of causality, since the individual shocks are not identified in our modelling framework.

Table 3: shock correlation cases

		Sign of correlation		
		0	+	-
Direction of shock	$\eta \rightarrow \epsilon$	1	3	5
	$\epsilon \rightarrow \eta$	2	4	6

Notes: table shows the numbering of the different cases. For explanations for each case, see main text.

There are six distinct cases of correlations in the models, depending on whether the original shock is a trend shock (like e.g. an innovation that lifts potential productivity) or a cycle shock (a temporary shock, like e.g. a strike). To set the scene for these six cases, it helps to think of the model setup (applicable to all of the specifications developed above) in the following set of equations:

$$\begin{aligned} \Delta y_t &= \Delta\tau_t + \Delta c_t \\ \Delta\tau_t &= f(\eta_t) \\ \Delta c_t &= f(\epsilon_t) \end{aligned} \tag{11}$$

where the period-to-period changes due to shocks in the trend and the cycle components are a function of the error term in both equations, and the change in the observable series is the sum of these shocks at each period, *ceteris paribus.*

- **In cases 1 and 2**, there is no correlation between the shocks, which means that trend shocks do not affect cycle shocks, and vice versa. This corresponds to the uncorrelated UC model.

- **In case 3**, the shock originates from the trend equation, and this has a positive correlation with the cycle shock. In terms of equations (11), if, at time t, there is a shock to η_t of size, say, x_t, which causes the trend to change by ax_t, and it also causes ϵ_t to change by z_t and consequently the cycle changes by bz_t (which has the same sign as ax_t, given the positive correlation assumed). Consequently, $\Delta y_t = ax_t + bz_t$, which is larger in absolute value than ax_t. We term this the "*overshooting*" case;

trend shocks cause larger than original shocks to productivity, for example, because agents take a positive cyclical signal from an observed trend shock and increase production by more than the original cyclical shock. In the case of a positive trend shock, this could lead to overheating and inflation pressures.

- **In case 4**, the shock originates from the cycle equation, and this has a positive correlation with the trend shock. In equation (11), if, at time t, there is a shock to ϵ_t of size x_t, which causes the cycle to change by ax_t, and it also causes η_t to change by z_t and consequently the trend changes by bz_t (which has the same sign as ax_t, given the positive correlation assumed). Consequently, as in case 3, $\Delta y_t = ax_t + bz_t$, which is larger in absolute value than ax_t. We call this the "*hysteresis*" case; the cycle shock has scarring effects on potential productivity.

- **In case 5**, the shock originates from the trend equation, and this has a negative correlation with the cycle shock. In equation (11), there is a shock to η_t of size x_t, which causes the trend to change by ax_t, and it also causes ϵ_t to change by z_t and consequently the cycle changes by bz_t (which has the opposite sign to ax_t, given the negative correlation assumed). Consequently, $\Delta y_t = ax_t + bz_t$, which is smaller in absolute value than ax_t. We call this the "*real business cycle (RBC)*" case; shocks to potential productivity have immediate effects smaller than their actual size on observed productivity. For example, a positive real shock to productivity leads to a higher immediate trend level to productivity, but actual productivity only tends to catch up later, causing a negative cyclical term shock. It is also worth noting that in this case, it is possible that the variance of the first difference of the trend could be greater than the variance of the first difference of the productivity series.

- **In case 6**, the shock originates from the cycle equation, and this has a negative correlation with the trend shock. In equation (11), if, at time t, there is a shock to ϵ_t of size x_t, which causes the cycle to change by ax_t, and it also causes η_t to change by z_t and consequently the trend changes by bz_t (which has the opposite sign to ax_t, given the negative correlation assumed). Consequently, as in case 5, $\Delta y_t = ax_t + bz_t$, which is smaller in absolute value than ax_t. This is called the "*creative destruction*" case; cyclical shocks have the opposite effect on potential productivity. This could happen, for example, if a negative cyclical shock led to an immediate folding of weak-productivity projects, causing the average productivity, measured at the whole-economy level, to fall.

We return to the empirics of the size and sign of the correlation in the Results section below. We do this by examining the statistical significance of the cross-correlation terms in the likeliest models, and while we cannot say anything definite on whether shocks have originated in the trend or the cycle component, the relative size of the error terms in the trend and cycle equations should give us some guidance on it.

2.4 Estimation

Most of the literature on correlated UC models have used maximum likelihood methods for estimation. The Kalman-filter recursions described above can be used to evaluate the (log)-likelihood function of the UCM and thus, in principle, maximum likelihood estimation of the parameters is possible in our setup as well. However, conventional likelihood methods based on numerical optimization algorithms often struggle to identify certain regions of the parameter space, and can fail to converge if the number of parameters to estimate is large compared to the number of available observations. Hence, we regularize the likelihood surface and make estimation of parameters feasible by adopting the Bayesian approach, while we also conduct a sensitivity analysis with respect to the chosen priors. Bayesian methods also allow for explicitly accounting for prior beliefs, and help to avoid the so-called "pile-up problem". This problem (which essentially causes a downward bias on the standard deviations of the parameter estimates) has been documented by, for example, Stock and Watson (1998), and Kim and Kim (2013) show how the problem can be avoided by using Bayesian methods.

To carry out the estimations, we use recently developed MCMC methods developed by Chan and Jeliazkov (2009) and Grant and Chan (2017a).[4] These have the advantage of using band and sparse matrix algorithms for state space models, which have been shown to be more efficient than conventional Kalman filter based algorithms. The efficiency of these methods is a great advantage when performing some of the grid search estimations for the different priors (as detailed below). This method also parameterises the starting point of the unobservable trend (τ_0), which helps in pinning down the trend and cycle paths for the model. More details on the modelling strategy are provided in Appendix C, including the extensions we have incorporated for the bivariate versions of our models compared to the original model by Grant and Chan (2017a).

Given that we use a Bayesian approach, we also need priors for all the parameters in the models. Table 4 details the priors we use for our models. They are relatively uninformative, and in line with those of Grant and Chan (2017a), with one exception. As discussed below, we want to study the robustness of the results to changing the prior on the relative size of the trend and cycle shocks. The priors for univariate and bivariate models are similar, where applicable.

[4]Estimation is executed by modifying a Matlab code kindly provided by Joshua Chan.

Table 4: **Model priors**

Parameter	Model	Prior density	Hyperparameters (mean,stdev)
ϕ_1	univariate	Log-normal	(1.3, 1)
	bivariate	Log-normal	(1.3, 1)
ϕ_2	univariate	Log-normal	(-0.7, 1)
	bivariate	Log-normal	(-0.7, 1)
μ_0	univariate	Log-normal	(0.5, 1)
	bivariate	Log-normal	(0.5, 1)
σ_ϵ	univariate	Uniform	(0,1)
σ_η	univariate	Uniform	(0,RS)
ρ	univariate	Uniform	*(min=-1, max=1)*
τ_0	bivariate	Log-normal	(500,100)
τ_{-1}	bivariate	Log-normal	(500,100)
Σ	bivariate	Inv. Wishart	(I_4, ∞] with RS trend shocks

Notes: for the bivariate models, relevant parameter priors are identical for both variables in the models. RS=relative size of shock, varies as described below.

We want to compare the likelihood of the different model options, and for this, we need to calculate marginal data likelihoods (also known as marginal data densities) as well as Bayes factors for the models. To proceed with this comparison, we define the Bayes factor between models i and j as follows (see Grant and Chan (2017a) and Kroese and Chan (2014) for more technical details):

$$BF_{ij} \equiv \frac{p(\mathbf{y}|M_i)}{p(\mathbf{y}|M_j)} \tag{12}$$

where

$$p(\mathbf{y}|M_k) = \int p(\mathbf{y}|\boldsymbol{\theta}_k, M_k)p(\boldsymbol{\theta}_k|M_k)d\boldsymbol{\theta}_k \tag{13}$$

for models M_k, $k = i, j$. $p(\mathbf{y}|M_k)$ is the marginal data likelihood under model k, $p(\mathbf{y}|\boldsymbol{\theta}_k, M_k)$ is the likelihood function depending on the model-specific parameter vector $\boldsymbol{\theta}_k$ and $p(\boldsymbol{\theta}_k|M_k)$ is the prior density.

We can also define the so-called posterior odds ratio between models i and j as follows:

$$\frac{\mathbf{P}(M_i|\mathbf{y})}{\mathbf{P}(M_j|\mathbf{y})} = \frac{\mathbf{P}(M_i)}{\mathbf{P}(M_j)} \times BF_{ij} \tag{14}$$

where $\mathbf{P}(M_i)/\mathbf{P}(M_j)$ is the prior odds ratio. If both models are equally likely *a priori*, the prior odds ratio equals 1 and the posterior odds ratio equals the Bayes factor. A Bayes factor value of X would indicate than model i is X times likelier than model j, given the data.

In general, the marginal likelihood in equation (13) cannot often be calculated analytically. Instead, classic cross-entropy methods have been developed to estimate the marginal likelihood, and more recently, Chan and Eisenstat (2015) introduced an MCMC importance sampling method, which we use for the bivariate models. The basic idea with this method is to locate a density that is close to an ideal importance sampling density within a convenient family of distributions, then minimise the cross-entropy distance to the ideal density, and then take an average of draws from the chosen sampling density. This method also has the added advantage of allowing for calculating the integrated marginal likelihood of the UK productivity data in non-nested models, where the second variable can differ. We can hence compare Bayes factors across the different versions of the bivariate models in a consistent way.

2.5 Prior identification

Even though the six different cases described above, in principle, provide an intuitive framework for thinking about the different UC models, it cannot be ruled out that the view on the preferred model can crucially depend on what is assumed as priors on the standard deviations of the trend and cycle shocks (σ_η and σ_ϵ, respectively). In particular, this view will be affected on the *prior relative size* (RS) of the shocks, which we define as follows:

$$RS = \frac{\sigma_\eta}{\sigma_\epsilon} \tag{15}$$

Typically, in previous literature on correlated UC models, RS has been assumed to be constant (usually 1), with no attempt to examine how its variation can affect the results.[5] We want to explicitly allow for this variation. Of course, the range of values that RS can take is infinite. But the Bayesian nature of our UC models allows us to use information from outside the models to inform the estimation process in the form of priors. There are various options on how to do this, but we use structural vector autoregression (SVAR) identification methods introduced by Cover et al. (2003). This method is well suited to the assumptions underlying our UC models, since it explicitly allows for correlations between the trend and cycle shocks. Cover et al. (2003) argue for various reasons why demand (cyclical) and supply (trend) shocks could be correlated, and this is in line with the principles of correlated UC models. More technical details about the identification methods in Cover et al. (2003) are set out in Appendix B.

We use the information from the identified historical decompositions of the SVAR models as priors on RS for our UC models. More specifically (as described in Appendix

[5] For example, Grant and Chan 2017(a) claim that the fact that they use different priors for the covariance matrix for the univariate and bivariate versions of their model, and still get similar results, proves that "...priors do not play an important role in driving the results." (footnote 6, p. 543). However, $RS = 1$ in all cases they consider and hence, the importance of its variation is not explored.

B), we use a SVAR model (SVAR1) to identify a case where demand shocks are caused by supply shocks, and another model (SVAR2) to identify a case where the causality runs in the opposite direction. The bivariate SVAR models, following Cover et al. (2003), have data on (the first differences of) UK GDP and CPI from 1991 to 2018. After estimating the models, we use the historical decomposition of GDP into demand and supply shocks, and then use the standard deviation of innovations in these shocks over time to measure the relative volatility of demand and supply shocks. For SVAR1, where the variation of the supply shock is relatively smaller, RS is 0.86, and for SVAR2, it is 1.46. This will give us a range of priors within which to compare the models.[6]

Even though modern macroeconomic models do tend to allow for correlations between trend and cycle shocks, there is nothing to force this correlation in practice. A researcher or a policy-maker may hold a prior according to which trends are relatively smooth and cyclical shocks are more volatile manifestations of temporary factors. Typically, for example, monetary policy would aim to mitigate the effects of the latter, while having little or no effect on the former. Hence, we also want to explore these types of priors. An easy, agnostic way of doing this is to allow for a prior that sets the relative size of the trend and cycle shocks to be equal to the Hodrick-Prescott filter (i.e., 1/40 with quarterly data). We highlight the difference that the SVAR priors make to a case where $RS = 0.025$.[7]

3 Empirical application

3.1 Data

The empirical analysis is carried out with publicly available data on productivity. For the UK and the US, we use productivity per hour data, as released by the Office for National Statistics (ONS) and the US Bureau of Economic Analysis (BEA), respectively.

We also want to use a proxy of a measure of global, or advanced economy productivity. These estimates are not readily available, but we construct a proxy of G7 productivity by using annual productivity (per worker), in levels (PPP-based) from the U.S. Conference Board Total Economy Database (TED). We then intrapolate this data into quarterly frequency, and calculate a GDP-weighted average of G7 (excl. UK) productivity levels.

All data used in the analysis are quarterly, with a sample period of 1991Q1 to 2018Q2.

[6] There is, of course, a confidence interval related to the point estimates from the SVAR models. However, given the way we search for the likeliest models in a wide grid, this is not a problem for our exercise. There could also be another way of defining the relative importance of the demand and supply shocks in the SVAR models, namely the relative contributions to the forecast error variance decomposition (FEVD) at long horizons. As it turns out, the relative FEVD contributions of the demand and supply shocks are fairly close to the ones from the historical decompositions, so again, our results are not affected by this choice.

[7] We do not explore the possibility of using a a prior from the SVAR models for ρ, even though in principle this would be possible. However, given that the SVAR model prior for ρ is around 0.1, i.e., relatively close to zero, and given the uniform prior we use, this does not make any difference in our setup. We explore the robustness of our results to some changes in the tightness of the prior for ρ below.

The starting point of the sample is dictated by the availability of comparable data on the global productivity proxy.[8]

3.2 Results from prior sensitivity analysis

We start by looking at how the marginal data densities of the different models depend on the RS and what that implies for the likeliest model in each case. To facilitate these comparisons, Figure 1 shows the marginal likelihood value for each of the univariate (LHS panel) and bivariate (RHS panel) models by searching through a grid of RS values from 0.02 to 1.6, at 0.02 intervals. This range is large and dense enough to accommodate all the feasible cases we are interested in, i.e., $RS = 0.86$, $RS = 1.46$ and $RS = 0.025$.

The results from the grid search suggest that the prior value of the trend and cycle shocks indeed matters for what is the likeliest model. For the univariate case, there is a clear split between low values of RS, where the non-correlated UC models are likelier, and higher values of RS, where the correlated model C is very robustly the likeliest model for all priors within the range of the SVAR1 and SVAR2 estimates.[9]

For the bivariate case, the results are not as clear-cut, but some conclusions emerge nevertheless. First, for low RS values, the 2M models, allowing for a smoother trend, are likelier. Second, for higher RS values, the C-model with US productivity as the second variable is generally the likelier.

We also examine how the trend component of the models changes, when RS is changed. Figure 2 depicts the results for the whole grid search range (LHS panel) and the SVAR range (RHS panel) for the likeliest correlated models. There is a lot of variation in the estimated trend, however, the smoother (and in some extreme cases, linear) trends all relate to the lower RS values for both of these models. When the RS value is in the SVAR range, i.e., when cross-correlation between the trend and cycle shocks is presumed, the resulting trend is very robust to changing the RS within the range.

The conclusion from the grid search is obvious - it matters whether a low or high RS prior is presumed, but once the prior is implemented, the results are robust to small variations in the RS. In other words, there needs to be a prior decision on whether trend and cycle shocks can be correlated or not. We are relatively agnostic about this decision, and hence, where applicable, we will report the results from the UC models for all three cases below (i.e., $RS = 0.86$, $RS = 1.46$ and $RS = 0.025$).

[8] We also experimented with a longer sample starting in 1980Q1 for those data that are available. The results are qualitatively similar to what we report below (although the marginal data density and posterior odds ratio estimates are not comparable across different sample periods).

[9] Note that it is not possible to compare the the likelihood of the different models across different prior assumptions. In other words, the comparisons in the charts should be read vertically, not horizontally.

3.3 UC model results

Table 5 reports the parameter estimates as well as the marginal data likelihood value and table 6 shows the posterior odds (PO) ratios from the various **univariate models,** separately for $RS = 0.86$ and $RS = 1.46$.[10] We also show results from a Hodrick-Prescott filter as well as a linear trend filter to compare the likelihood of the UC models with these simple models. A number of results are worth highlighting.

Table 5: Univariate models - parameter estimates

		Coefficients			*Standard deviations/correlations*			
Model	***RS***	**μ**	**ϕ1**	**ϕ2**	**ϵ**	**η**	**ρ**	**ML**
NC	**0.86**	0.37	0.85	-0.25	0.06	0.40		-120.50
		0.06	*0.42*	*0.32*	*0.05*	*0.08*		
	1.46	0.37	0.82	-0.24	0.06	0.40		-120.99
		0.06	*0.42*	*0.32*	*0.05*	*0.08*		
C	**0.86**	0.38	0.72	-0.13	0.22	0.63	-0.67	-117.85
		0.08	*0.38*	*0.31*	*0.21*	*0.13*	*0.22*	
	1.46	0.39	0.77	-0.08	0.36	0.86	-0.76	-117.98
		0.09	*0.32*	*0.28*	*0.27*	*0.28*	*0.17*	
NC_2M	**0.86**		0.76	-0.03	0.32	0.02		-122.00
			0.15	*0.11*	*0.06*	*0.01*		
	1.46		0.76	-0.03	0.32	0.02		-122.55
			0.15	*0.11*	*0.06*	*0.01*		
C_2M	**0.86**		0.77	-0.03	0.31	0.02	0.08	-122.09
			0.17	*0.11*	*0.06*	*0.01*	*0.55*	
	1.46		0.77	-0.03	0.31	0.02	0.06	-122.63
			0.17	*0.11*	*0.06*	*0.01*	*0.54*	
DT								-125.14
HP								-165.75

Notes: the table shows posterior median point estimates (first row) and standard deviations (second row) for each model for the selected parameters. ML is the marginal log data likelihood. RS=relative size of trend and cycle shocks (see text for details).

[10] As already established above, when $RS = 0.025$, non-correlated UC models are more likely. For brevity, we do not report the results for that case separately here, but show the resulting trend and cycle dynamics in charts below.

Table 6: Univariate models - posterior odds ratios

Model i / Model j	*RS*	NC	C	NC_2M	C_2M	DT	HP
NC	**0.86**	1.0					
	1.46	1.0					
C	**0.86**	0.1	1.0				
	1.46	0.0	1.0				
NC_2M	**0.86**	4.5	63.2	1.0			
	1.46	4.7	96.2	1.0			
C_2M	**0.86**	4.9	69.4	1.1	1.0		
	1.46	5.2	104.3	1.1	1.0		
DT		103.8	1470.3	23.3	21.2	1.0	
		63.6	1288.0	13.4	12.4	1.0	
HP		inf	inf	inf	inf	inf	1.0
		inf	inf	inf	inf	inf	1.0

Notes: the table shows the posterior odds ratio of model i vs model j.
Value above 1 means the model in column i is likelier than model in row j.
RS=relative size of trend and cycle shocks (see text for details).

First, looking at the marginal data densities suggests that correlated models are the likelier for the data than uncorrelated ones, as already highlighted in the grid search above. Furthermore, the UC models are always likelier than enforcing an H-P filter or a deterministic trend on the data.[11]

Second, allowing for the large, non-informative priors for the trend and cycle component shocks results in the posterior estimates of the trend shocks being relatively large compared to the cycle shocks (for similar results with US GDP, see Grant and Chan (2017a)).

Third, there appears to be a significant negative relationship between trend and cycle shocks in the likeliest model (model C), as also confirmed by the posterior distributions for the estimates of the correlation coefficient ρ (for $\sigma_{\epsilon\eta}$) in the top left-hand panel of Figure 3.[12] This finding would be consistent with either the "RBC" or the "creative destruction" hypotheses detailed above (cases 5 and 6 in Table 3). To investigate this further, we plot the shock terms from the trend and cycle equations from C (and also the C-US) model in Figure 4. The charts suggest that trend shocks are much more important for productivity dynamics than cycle shocks in these models, and this is especially apparent during the financial crisis. Hence, this finding suggests that RBC type real shocks have been more

[11] We also cross-checked the Bayesian results with relevant maximum likelihood parameter estimates for selected models. As expected, the point estimates are broadly similar, but the standard deviations of the ML estimates tend to be much smaller, in line with the pile-up problem mentioned above.

[12] This is in line with typical findings in the literature for US GDP (see e.g. Morley et al. (2003)), but, to our knowledge, has not been examined for UK productivity before.

important in driving productivity dynamics than other types of trend or cyclical shocks.

We also experimented with a version of the C-model that allows for a break for the pre- and post financial crisis periods (in practice, this is achieved by introducing a dummy that has the value 0 before 2008Q1 and 1 after it). This model also confirms the weakness seen in trend productivity since the financial crisis in the UK, as the estimate for the post-financial crisis trend growth rate exactly offsets the estimate for the growth rate for the entire sample. However, the ML estimate and PO ratios for the model suggests that it is substantially less likely to be the correct model for the data than any of the other models, so we take no further guidance from this model.

Figure 5 shows the cycle estimates (i.e., c_t in equation (2)) from the different univariate models. There are some noteworthy differences between the models. In particular, the C model implies that the financial crisis caused a largely structural shock to the trend productivity growth, rather than a cyclical shock. This is in contrast to the results from the NC model, which has a more volatile cycle component (close to the H-P filter). As is also typical for these types of models, the standard deviations around the point estimates for the unobservable variables are large.

Table 7 reports the parameter estimates as well as the marginal data likelihood values (for the UK productivity data), table 8 shows the posterior odds ratios across the models and Figure 6 shows selected cycle components from the various **bivariate models**. The results tend to favour the models where US rather than global productivity is the second variable, which is not surprising given the strong links between the US and the UK economies.

Table 7: Bivariate models - parameter estimates

		Coefficients				*Standard deviations/correlations*						
Model	***RS***	**μ_UK**	**μ_y2**	**ϕ1_UK**	**ϕ2_UK**	**ϵ_UK**	**η_UK**	**ηϵ_UK**	**ηϵ_y2**	**ϵ_UK_y2**	**η_UK_y2**	**ML**
C_US	**0.86**	0.38	0.50	0.52	0.15	0.39	0.68	-0.67	-0.43	0.21	0.54	-130.74
		0.08	*0.07*	*0.24*	*0.19*	*0.26*	*0.28*	*0.17*	*0.29*	*0.34*	*0.22*	
	1.46	0.38	0.51	0.53	0.13	0.37	0.68	-0.66	-0.45	0.21	0.48	-128.74
		0.08	*0.07*	*0.25*	*0.20*	*0.26*	*0.28*	*0.17*	*0.26*	*0.34*	*0.21*	
C_Gprod	**0.86**	0.39	0.29	0.54	0.12	0.55	0.76	-0.74	-0.67	0.48	0.59	-130.94
		0.08	*0.04*	*0.24*	*0.19*	*0.43*	*0.39*	*0.18*	*0.16*	*0.31*	*0.30*	
	1.46	0.39	0.28	0.58	0.09	0.44	0.73	-0.68	-0.66	0.44	0.53	-130.09
		0.08	*0.05*	*0.26*	*0.22*	*0.35*	*0.35*	*0.19*	*0.16*	*0.31*	*0.27*	
C_2M_US	**0.86**			0.55	0.01	0.23	0.08	0.00	0.07	0.02	0.16	-131.39
				0.21	*0.14*	*0.06*	*0.03*	*0.26*	*0.26*	*0.13*	*0.22*	
	1.46			0.47	0.04	0.20	0.11	0.00	0.07	0.00	0.19	-133.85
				0.24	*0.15*	*0.06*	*0.04*	*0.24*	*0.24*	*0.14*	*0.20*	
C_2M_Gprod	**0.86**			0.48	0.02	0.21	0.09	0.03	-0.14	0.04	0.35	-132.86
				0.24	*0.14*	*0.05*	*0.03*	*0.27*	*0.19*	*0.16*	*0.19*	
	1.46			0.44	0.05	0.19	0.12	0.05	-0.15	0.00	0.33	-135.98
				0.25	*0.16*	*0.05*	*0.04*	*0.25*	*0.18*	*0.17*	*0.18*	

Notes: the table shows posterior median point estimates (first row) and standard deviations (second row) for each model for the selected parameters. ML is the marginal log data likelihood.
RS=relative size of trend and cycle shocks (see text for details).

Table 8: Bivariate models - posterior odds ratios

Model i Model j	*RS*	C_US	C_Gprod	C_2M_US	C_2M_Gprod
C_US	**0.86**	1.0			
	1.46	1.0			
C_Gprod	**0.86**	1.2	1.0		
	1.46	3.9	1.0		
C_2M_US	**0.86**	1.9	1.6	1.0	
	1.46	166.2	42.8	1.0	
C_2M_Gprod	**0.86**	8.3	6.9	4.4	1.0
	1.46	1395.1	358.9	8.4	1.0

Notes: the table shows the posterior odds ratio of model i vs model j.
Value above 1 means the model in column i is likelier than model in row j.
RS=relative size of trend and cycle shocks (see text for details).

There is also some evidence that the trend shocks between UK and US/global productivity are strongly positively correlated for the C-models (see Figure 3). Apart from the C-Gprod model, this is not the case for the correlations between the cycle shocks. This is a noteworthy finding, and provides evidence on long-term trend productivity shocks between the UK and other advanced economies being positively correlated, while the case is less so for short-term cyclical shocks.

There are relatively large differences between the cycle dynamics of the different bivariate models. The models allowing for time-varying trends have more volatile cycle components over time (which was also the case for the univariate models), but the results from the 2M models are also more robust to changes in the y2 variable. Given the differences in the ML estimates, the results tend to favour the C rather than the 2M models for the bivariate version of the models. It is possible that the number of parameters becomes a more detrimental factor to the estimation in the more complex 2M models in the bivariate case, where the number of parameters to be estimated is large in any case.

For comparing trend and cycle estimates across selected models, figures 7 and 8 show the trends and cycles of two univariate and two bivariate models. Despite the differences in the model structures and the detailed results presented above, some common themes emerge. The estimates of the cycle differ significantly during the financial crisis, but there does not appear to be evidence in any of the models on there being a negative "productivity gap" at the end of the sample. The trend growth rate estimates are relatively similar across the different C models, but they are much smoother in the NC-2M and C-2M-Gprod models that have a lower RS value. This further emphasises the importance of the prior assumption on the correlation between the trend and the cycle shocks. Nevertheless, all the models paint a fairly pessimistic picture of UK productivity dynamics since the crisis; the growth rates are much lower than before the crisis, although there are some signs of a pick-up towards the end of the sample.

Figure 9 depicts a rolling end-point estimate of the trend correlation (η_UK_y2) term between the two variables in selected bivariate cases. These are estimates based on keeping the start of the sample constant (1991Q1) and rolling the end point forward quarter-by-quarter from 2007Q4 to 2018Q2. When prior correlation between the trend and cycle shocks is assumed, the results for the likeliest models suggest that the correlation of the trend shocks has become more positive and more significant in recent years. This supports the fairly common narrative of productivity shocks having become more synchronised since the financial crisis. However, for the likeliest model with the lower RS prior, the correlation is not statistically significant, again highlighting the importance of the priors.

3.4 Forecasting experiment

In this study, we are relatively agnostic about the choice of priors, as the discussion above implies. The econometrician may have different priors in different situations and for different purposes. However, evaluating the out-of-sample forecasting performance of UC models is an important way of testing their usefulness for policy purposes. To examine this, we run a pseudo-real time forecasting experiment or the likeliest models discussed above for $RS = 0.86$, $RS = 1.46$ and $RS = 0.025$. When real-time vintage data is available, we report these results (vintage RT) along with the version with current data only (pseudo RT).[13] In this experiment, we estimate the models with data from 1991Q1 to 1999Q4, then construct a forecast for 1 to 12 quarters ahead, and then roll the estimation period forward quarter-by-quarter, generating the 1-12 quarter forecasts each time. We do this until the end of the sample.

The results of this experiment are shown in Figure 10 and Table 9. There is a clear difference between the models with the higher RS priors versus the models with the H-P prior; the latter outperform the former at all horizons. In fact, as the table and the RHS chart indicate, the pseudo-RT performance of the univariate non-correlated model as well as the C-2M-Gprod model with the low priors against a pure random walk model is very good at all horizons. These models appear to be able to forecast the UK productivity dynamics relatively well in "normal" times, but even they do not perform particularly well around turning points, like the financial crisis (Figure 10, RHS panel). For the vintage-RT models, the forecasting performance is weaker than for the pseudo-RT versions. However, again, it is clear that the NC-2M model performs well, whereas the C-models have almost

[13] For some of the data (like the global productivity series), real-time vintages of the data are not available. Hence, the results here should be seen as illustrative rather than being accurate on the actual real-time performance. Nevertheless, the results should be accurate on the *relative* performance between the models.

no forecasting power apart from the first 1-2 quarters.

Overall, if forecasting performance of the models is an important factor, there is a strong case for using low priors and non-correlated rather than correlated versions of the univariate model for UK productivity. In the case of the bivariate models, versions that smooth the trend component are preferable. This is an important result largely ignored in previous literature.

Table 9: Forecast performance of selected models

Quarters	NC 2M (RS=0.025)		C (RS=0.86)		C (RS=1.46)		C-US (RS=1.46)		C-2M-Gprod (RS=0.025)
	pseudo RT	vintage RT	pseudo RT	vintage RT	pseudo RT	vintage RT	pseudo RT	vintage RT	pseudo RT
1	0.68 ***	0.99	0.68 ***	0.93 *	0.68 ***	0.94	0.68 ***	0.93 *	0.67 ***
2	0.67 ***	0.90 **	0.72 ***	0.86 **	0.71 ***	0.86 **	0.71 ***	0.85 **	0.67 ***
3	0.71 ***	0.87 **	0.78 ***	0.89	0.77 ***	0.89	0.76 ***	0.88	0.70 ***
4	0.75 ***	0.87 **	0.85 **	0.96	0.84 **	0.95	0.83 **	0.95	0.74 ***
5	0.77 ***	0.89 **	0.89	1.01	0.88	1.00	0.87	1.00	0.76 ***
6	0.81 ***	0.92 *	0.95	1.05	0.95	1.04	0.93	1.04	0.80 ***
7	0.83 ***	0.94	1.02	1.07	1.02	1.06	1.01	1.06	0.84 ***
8	0.85 **	0.91 **	1.06	1.03	1.07	1.03	1.05	1.03	0.86 **
9	0.81 ***	0.88 ***	0.99	0.99	1.00	1.00	0.99	0.99	0.82 ***
10	0.77 ***	0.84 ***	0.92	0.97	0.92	0.97	0.91	0.97	0.78 ***
11	0.75 ***	0.82 ***	0.88 *	1.00	0.89 *	1.00	0.88 *	1.00	0.76 ***
12	0.76 ***	0.83 ***	0.89 *	1.08	0.89 *	1.08	0.89 *	1.08	0.76 ***

Notes: the table shows the Theil U statistic for forecasts at different horizons.
Statistical significance: *** 1%, ** 5%, * 10%.

3.5 Monte Carlo experiment

Another way to examine the validity of the different types of models is to carry out a Monte Carlo experiment on hypothetical data generating processes (DGP) and see how the different UC models capture the relevant features of these DGPs.[14] We do this for the univariate UC models.[15] More specifically, we consider four true DGPs. The first two (DGP1 and DGP2) are of type NC and C, the last two (DGP3 and DGP4) are of type _2M. DGP1 and DGP3 are cases where the trend is relatively volatile (high σ_η and $RS = 1$) and there is a negative correlation between the trend and cycle shocks (C and C_2M type DGPs). DGP2 and DGP4 are cases where the trend is relatively smooth (low σ_η and $RS = 0.025$) and there is no correlation between the trend and cycle shocks (NC and NC_2M type DGPs). The first column of Table 10 sets out the four DGPs and their relevant parameter values.

Our experiment then proceeds as follows:

1. We generate the true y_t for each of the DGPs with the structure of the univariate models (as in equations (1) to (4)) by summing up sequentially the trend and cycle

[14] See Kamber et al. (2018) for an example of an experiment that is similar in spirit.

[15] There is no analytical advantage of adding the bivariate case for this experiment.

components that are produced with the true DGP parameter values and by setting $\tau_0 = c_0 = 0$ as the initial values. To generate the time series, we draw the error terms for each time period from the $\Sigma_{\epsilon\eta}$ multivariate normal variance-covariance matrix. We run the procedure for two cases, $T = 250$ (small T) and $T = 10,000$ (large T) to study whether "population" properties can be captured in a small sample.

2. We then take the four true DGPs for y_t and run four UC models on the "data". In each case, we use priors corresponding to the true DGP. This produces four UC models for each DGP, so 16 models altogether.

3. We compare the relevant estimated parameter values to the true parameter values in the DGPs to see how well the different models capture the true parameter values in each case. We are especially interested in σ_η, σ_ϵ and ρ, as well as the correlation coeffcient between the true DGP and UC model cycle components ("cycle correlation" in Table 10).

The main results of the experiment are reported in Table 10. The two different samples have parameter values that are close to the true DGP that generated those values (columns true DGP vs true data), although DGP2 and DGP4 have somewhat higher in-sample standard deviations of the trend (σ_η) than the true DGP. Also, the marginal data likelihood almost always picks the UC model that is based on the true DGP, apart from DGP4 in the large T case. In other words, if the true RS was known, the ML would be a fairly good measure of the likeliest model in our experiment.

Looking at the parameter values for the UC models that are based on the true DGPs (green cells in Table 10) suggests that especially for large T, the models capture the properties of the true DGPs relatively well. However, ρ turns out to be difficult to capture, and the sign is often wrong, especially in the small T. This suggests that we should be careful when interpreting the trend/cycle correlations in the correlated UC models.

In terms of the cycle correlation, for large T, the correlation in the case of the correct UC model on the underlying true DGP is highest (values in green cells are higher than other cells on the same row). But it is also worth noting that especially for small T, the _2M models (DGP3 and DGP4) outpeform the other two models; they tend to find a cycle that has a relatively high correlation with the true cycle, even if the true DGP is different.

The results of the Monte Carlo experiment are necessarily tentative, as, of course, we will never know what the true DGP in our UK productivity data is. However, the results suggest some guidelines on which UC models to prefer. In small samples (like in our case), if one does not have strong beliefs that the true DGP is of type 1, with a relatively volatile trend component, it is "safer" to pick a UC model of type C_2M or NC_2M with a relatively low prior trend volatility. This is because even if the true DGP turns out to be different from the chosen model, the latter can still capture important properties of the data relatively well.

Table 10: Monte Carlo simulation results

TRUE MODEL				ESTIMATED MODELS			
T=250		True DGP	True data	DGP1	DGP2	DGP3	DGP4
DGP1	σ(η)	1	1.056	0.871	0.023	0.016	0.012
C	σ(ϵ)	1	0.972	0.215	0.967	0.907	0.919
	ρ	-0.5	-0.531	0.115		-0.055	
	cycle correlation			0.433	0.005	0.503	0.523
	ML			-377.523	-384.785	-386.453	-383.097
DGP2	σ(η)	0.025	0.167	0.104	0.023	0.010	0.010
NC	σ(ϵ)	1	0.951	0.722	0.914	0.915	0.911
	ρ	0	0.013	0.418		0.536	
	cycle correlation			0.942	0.945	0.924	0.953
	ML			-367.094	-364.426	-380.840	-378.513
DGP3	σ(η)	1	1.056	0.965	0.027	0.893	0.020
C_2M	σ(ϵ)	1	0.972	0.992	1.005	0.675	0.989
	ρ	-0.5	-0.531	0.783		-0.046	
	cycle correlation			-0.002	0.049	0.472	0.318
	ML			-872.246	NA	-457.249	-568.776
DGP4	σ(η)	0.025	0.167	0.725	0.029	0.050	0.016
NC_2M	σ(ϵ)	1	0.951	0.731	1.002	0.879	0.937
	ρ	0	0.013	0.035		0.110	
	cycle correlation			0.007	0.000	0.865	0.807
	ML			-429.400	NA	-398.316	-395.725
T=10,000		True DGP	True data	DGP1	DGP2	DGP3	DGP4
DGP1	σ(η)	1	0.991	0.961	0.027	0.011	0.009
C	σ(ϵ)	1	0.996	0.819	1.004	0.991	0.991
	ρ	-0.5	-0.493	-0.414		0.560	
	cycle correlation			0.622	0.093	0.331	0.505
	ML			-15027.1	-15186.5	-15210.4	-15214.6
DGP2	σ(η)	0.025	0.157	0.029	0.024	0.009	0.008
NC	σ(ϵ)	1	1.000	0.915	0.995	0.990	0.976
	ρ	0	0.005	0.320		0.799	
	cycle correlation			0.937	0.946	0.795	0.876
	ML			-14561.9	-14557.4	-15074.6	-15110.4
DGP3	σ(η)	1	0.991	0.995	0.030	0.977	0.020
C_2M	σ(ϵ)	1	0.996	0.994	1.005	0.844	0.995
	ρ	-0.5	-0.493	0.929		-0.919	
	cycle correlation			0.047	0.217	0.661	0.254
	ML			NA	NA	-18121.4	NA
DGP4	σ(η)	0.025	0.157	0.995	0.030	0.024	0.017
NC_2M	σ(ϵ)	1	1.000	0.890	1.005	0.971	0.990
	ρ	0	0.005	-0.510		-0.418	
	cycle correlation			0.074	0.060	0.787	0.796
	ML			-175965.2	NA	-15828.0	-15833.5

Notes: the table shows selected parameter values for the true DGP (in rows) and the simulated UC model for each case (in columns). ML is the marginal data likelihood. ML value marked NA indicates large values and hence, very unlikely models.

3.6 Robustness of trend and cycle shocks

As shown in Table 4, for the bivariate models, an inverse Wishart prior distribution for the covariance matrix is used, centered at the identity matrix. Hence, in the bivariate case, there is natural tendency for the cross-correlations between trends and cycles to be centered around zero, and for the variances to be centered around 1. An obvious question is, how do the univariate model results change if one assumes a normal prior, rather than the uninformative uniform prior that we use in the benchmark case.

We conduct two exercises to study the robustness to the relevant priors. First, we let the prior for ρ to be normally distributed. Figure 11 shows the prior and posterior distributions for two such choices for the likeliest univariate model (C), where the uniform prior has been replaced by a normal prior. In general, the results are robust to reasonable choices of the normal priors (i.e., unless the prior is very tightly centered at zero). Indeed, if one wants to set a prior that is tightly centered at zero, then the lower RS values can be used, which will result in the NC model as the likeliest choice.

Second, we allow for normal N(0,1) priors for both ρ and the standard deviations of the trend and cycle shocks (σ_ϵ and σ_η) and then carry out the grid search for different values of RS described above. The results look relatively similar to the benchmark case; Figure 12 shows that for the univariate models, the C-model is still by far the likeliest for the high RS values, while the smoother trend models (NC-2M, C-2M) are more likely for low RS values. The posterior parameter estimates for the likeliest model (C) are also very similar compared to the benchmark case with the uniform priors.

4 Conclusions

This paper presents a new way of looking at productivity data through the lens of a Bayesian correlated unobserved components model. We allow for different specifications of the trend and the size of the trend and cycle shocks, using new sampling techniques introduced in recent literature.

Generally, it has to be kept in mind that there is a potential identification problem in any UC model with a non-diagonal covariance matrix of the error terms (see e.g. Clark (1987)). One way of overcoming this issue is to use Bayesian methods, where the trade-off is that the posterior results may be heavily driven by the prior assumptions.

In this spirit, the main contribution of the study is to examine the robustness of the results to different priors of the shock components and thus to provide a general "health warning" against assuming particular priors in Bayesian UC models. To use these models properly, one needs to have a prior narrative about how the trend and cycle shocks are correlated and consequently, how smooth the trend should be. Our suggestion, based on the results with UK productivity data, is to examine a wide range of the most relevant priors before drawing conclusions from these types of models for policy-making purposes. In our case, it also turns out that if forecasting performance of the models is a priority, there is a strong case for using low priors and non-correlated rather than correlated versions of the univariate model. In the case of the bivariate models, versions with a smooth trend component perform better.

According to the results, when smooth, H-P filter type priors are assumed for the variation of the trend shock component, non-correlated UC models, or models that allow for a smoother trend, are the likeliest choice. However, it is worth noting that this is not the same as using an H-P filter; the UC models allow for "the data to speak" freely, and are much more likely than a strictly H-P filtered model.

On the other hand, we do find some evidence from a structural VAR exercise informing our priors that trend shocks are relatively volatile in the UK data. When this relatively large variation of the trend shock is assumed in the prior, the most likely UC models allow for correlation between trend and cycle components, and they suggest that this correlation is negative for UK productivity data. This is consistent with real shocks being the dominant force in driving productivity dynamics, as has been suggested in the literature studying GDP dynamics for advanced economies. The likeliest models also imply substantially weaker trend growth since the financial crisis. Finally, some of the evidence suggests that there is a significant positive correlation between shocks to UK trend productivity and those of other advanced economies. These correlations have become stronger since the financial crisis.

Whichever priors or models are used, the results are consistent with a relatively pessimistic view on post-financial crisis productivity dynamics in the UK. The weakness of trend productivity growth rate appears to be consistent with a secular stagnation type narrative. On the other hand, it is possible that positive real shocks could quickly lead to an improvement in trend productivity. More structural models and views on future technological progress are needed to formulate forecasts for that; our model - like any time series model, no matter how sophisticated - is a reflection of past, not future data dynamics. Nevertheless, it would seem to be the case that structural policies that foster positive productivity shocks are key to any future pick-up, and cyclical policies have a smaller role to play.

While we believe this study provides a fresh view on how to study productivity dynamics, there are plenty of ways to broaden the analysis. Within the modelling framework, further research could explore, for example, more complicated structures for productivity dynamics and the shock processes. In particular, it would be interesting to study a non-linear version of the model, by, for example, allowing for a Markov-switching process to distinguish for dynamics before and after the financial crisis. One could also expand the multivariate version of the model by including other relevant variables, like trade openness and population aging.

A Appendix

This appendix describes the state space equations in detail. For completeness, the specification given here is for a model with two observable endogenous variables (i.e., the bivariate version), but this can be reduced to the univariate case by dropping the subscript i.

The state space form of the dynamics for country i (=1 (UK) or =2 (world)) over time t is the following:

$$y_{it} = A' s_{it} \tag{A.1}$$

$$s_{it} = C + B s_{it-1} + F v_{it} \tag{A.2}$$

where (A.1) is the observation equation and (A.2) is the state equation, y_{it} is the level of observed productivity for country i at time t, A, B, C, and F are the parameter matrices to be estimated and s_{it} is the state vector.

More specifically, the observation equation takes the following form (for definitions of the variables, see main text):

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \tau_{1t} \\ c_{1t} \\ c_{1t-1} \\ \tau_{2t} \\ c_{2t} \\ c_{2t-1} \end{bmatrix} \tag{A.3}$$

and the state equation is the following:

$$\begin{bmatrix} \tau_{1t} \\ c_{1t} \\ c_{1t-1} \\ \tau_{2t} \\ c_{2t} \\ c_{2t-1} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ 0 \\ 0 \\ \mu_2 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \phi_{11} & \phi_{12} & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & \phi_{21} & \phi_{22} \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \tau_{1t-1} \\ c_{1t-1} \\ c_{1t-2} \\ \tau_{2t-1} \\ c_{2t-1} \\ c_{2t-2} \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \eta_{1t} \\ \epsilon_{1t} \\ 0 \\ \eta_{2t} \\ \epsilon_{2t} \\ 0 \end{bmatrix} \tag{A.4}$$

and the variance-covariance matrix has the following form:

$$\Sigma_{\eta\epsilon} = \begin{bmatrix} \sigma^2_{\varepsilon 1} & \sigma_{\varepsilon 1 \varepsilon 2} & \sigma_{\eta 1 \varepsilon 1} & \sigma_{\eta 2 \varepsilon 2} \\ \sigma_{\varepsilon 1 \varepsilon 2} & \sigma^2_{\varepsilon 2} & \sigma_{\eta 1 \epsilon 2} & \sigma_{\eta 2 \epsilon 2} \\ \sigma_{\eta 1 \varepsilon 1} & \sigma_{\eta 1 \epsilon 2} & \sigma^2_{\eta 1} & \sigma_{\eta 1 \eta 2} \\ \sigma_{\eta 2 \varepsilon 2} & \sigma_{\eta 2 \epsilon 2} & \sigma_{\eta 1 \eta 2} & \sigma^2_{\eta 2} \end{bmatrix} \tag{A.5}$$

B Appendix

The structural vector autoregression (SVAR) model used to inform the priors is based on Cover et al. (2003). The reader is referred to their paper for more details of the model, but the main insight of their work is to use a basic AD-AS framework to identify a two-variable SVAR model, with output (y_t) and price (p_t) level, for demand and supply shocks. In practice, the two variables enter the SVAR in first differences to make them stationary.

In Cover et al. (2003), the starting point is the relationship between the reduced form residuals (e_{yt}, e_{pt}) and structural residuals (ε_t, η_t):

$$\begin{bmatrix} e_{yt} \\ e_{pt} \end{bmatrix} = \begin{bmatrix} \frac{1}{1+\alpha} & \frac{\alpha}{1+\alpha} \\ \frac{-1}{1+\alpha} & \frac{1}{1+\alpha} \end{bmatrix} \begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix} \tag{B.1}$$

where ε_t and η_t denote the serially uncorrelated structural aggregate supply and demand shocks, respectively, and α is the slope of the aggregate supply curve.

It then follows that:

$$\begin{aligned} & \begin{bmatrix} var(e_y) & cov(e_y, e_p) \\ cov(e_y, e_p) & var(e_p) \end{bmatrix} \\ = & \begin{bmatrix} \frac{1}{1+\alpha} & \frac{\alpha}{1+\alpha} \\ \frac{-1}{1+\alpha} & \frac{1}{1+\alpha} \end{bmatrix} \begin{bmatrix} \sigma_\varepsilon^2 & \sigma_{\varepsilon\eta} \\ \sigma_{\varepsilon\eta} & \sigma_\eta^2 \end{bmatrix} \begin{bmatrix} \frac{1}{1+\alpha} & \frac{-1}{1+\alpha} \\ \frac{\alpha}{1+\alpha} & \frac{1}{1+\alpha} \end{bmatrix} \end{aligned} \tag{B.2}$$

The original Blanchard-Quah (1989) decomposition is a special case of (B.2), with $\sigma_{\varepsilon\eta} = 0$, as well as forcing the variability of the demand and supply shocks to be equal (=1). However, by modelling the direction of causality between demand and supply shocks, as well as assuming the structure of the simple AD-AS model, one can identify the model even when allowing for $\sigma_{\varepsilon\eta} \neq 0$. There are then two cases considered; one in which the causal ordering runs from supply to demand, and another one in which the causal ordering runs from demand to supply.

In the first case (SVAR1), a demand shock is composed of a pure aggregate demand shock (v_t) and a change in aggregate demand that is caused by the aggregate supply shock ($\rho\varepsilon_t$):

$$\eta_t = \rho\varepsilon_t + v_t \tag{B.3}$$

In the second case (SVAR2), a supply shock is composed of a pure aggregate supply shock (δ_t) and a change in aggregate supply caused by aggregate demand ($\gamma\eta_t$):

$$\varepsilon_t = \gamma\eta_t + \delta_t \tag{B.4}$$

The first case is similar to the Blanchard-Quah model. In this model, all the variation in output due to common shifts of the AD and AS curves are attributed to the structural supply shock, and hence, demand shocks have no long-term effect on output. In this case, combining equations (B.2) and (B.3) yields the following equation that allows for identifying the structural shocks using the reduced-form VAR parameters on the LHS of the equation, and the AD-AS equations:

$$\begin{bmatrix} var(e_y) & cov(e_y, e_p) \\ cov(e_y, e_p) & var(e_p) \end{bmatrix} \tag{B.5}$$
$$= \begin{bmatrix} \frac{1}{1+\alpha} & \frac{\alpha}{1+\alpha} \\ \frac{-1}{1+\alpha} & \frac{1}{1+\alpha} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \rho & 1 \end{bmatrix} \begin{bmatrix} \sigma_\varepsilon^2 & 0 \\ 0 & \sigma_v^2 \end{bmatrix} \begin{bmatrix} 1 & \rho \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{1+\alpha} & \frac{-1}{1+\alpha} \\ \frac{\alpha}{1+\alpha} & \frac{1}{1+\alpha} \end{bmatrix}$$

In similar fashion, the identification in the second case is based on the following structure:

$$\begin{bmatrix} var(e_y) & cov(e_y, e_p) \\ cov(e_y, e_p) & var(e_p) \end{bmatrix} \tag{B.6}$$
$$= \begin{bmatrix} \frac{1}{1+\alpha} & \frac{\alpha}{1+\alpha} \\ \frac{-1}{1+\alpha} & \frac{1}{1+\alpha} \end{bmatrix} \begin{bmatrix} 1 & \gamma \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \sigma_\delta^2 & 0 \\ 0 & \sigma_\eta^2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \gamma & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{1+\alpha} & \frac{-1}{1+\alpha} \\ \frac{\alpha}{1+\alpha} & \frac{1}{1+\alpha} \end{bmatrix}$$

We use the results of a SVAR model with data on UK GDP and CPI prices (in first differences) from 1991 to 2018 to inform the prior on the relative size of the standard deviation of the trend and the cycle components (σ_η and σ_ϵ, respectively) of our UC models. More specifically, we use the historical decomposition of GDP into demand and supply shocks, and then use the standard deviation of innovations in these shocks over time. Figure 13 shows the dynamics for the two cases. For SVAR1, where the variation of the supply shock is relatively smaller, the ratio of the standard deviations is 0.86, and for SVAR2, it is 1.46. These are the values we use for the relative size of the standard deviation of the trend and the cycle components in our UC models, and we report results for both of these cases in the main text.

It is worth noting that in the SVAR model we use GDP (rather than productivity directly), as this is the specification used in the original model by Cover et al. (2003), and it also has the desirable feature of not using the same data for both the priors and the estimation of the UC models.

C Appendix

This appendix details the Bayesian Markov Chain Monte Carlo (MCMC) Gibbs sampling algorithm used in the bivariate versions of the model. It draws heavily on the original algorithm as introduced by Grant and Chan (2017a); but also highlights the ways in which we have modified the original model. For brevity and ease of disposition, we only describe the 2M version of the bivariate model (i.e., models C-2M-US, C-2M-Gprod and C-2M-trade in Table 2 in the main text). The basic bivariate version of the model is nested in this and is a simple extension of the univariate model described in Grant and Chan (2017a), so we refer the reader to this reference for more details.

We begin by setting up the error terms and the covariance matrix of the bivariate model, assuming jointly normal errors (see equations (5) to (8) and definitions in the main text) in the following way:

$$\begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \eta_{1t} \\ \eta_{2t} \end{pmatrix} \sim N\left(\mathbf{0}, \begin{pmatrix} \sigma^2_{\varepsilon 1} & \sigma_{\varepsilon 1 \varepsilon 2} & \sigma_{\eta 1 \varepsilon 1} & \sigma_{\eta 2 \varepsilon 2} \\ \sigma_{\varepsilon 1 \varepsilon 2} & \sigma^2_{\varepsilon 2} & \sigma_{\eta 1 \epsilon 2} & \sigma_{\eta 2 \epsilon 2} \\ \sigma_{\eta 1 \varepsilon 1} & \sigma_{\eta 1 \epsilon 2} & \sigma^2_{\eta 1} & \sigma_{\eta 1 \eta 2} \\ \sigma_{\eta 2 \varepsilon 2} & \sigma_{\eta 2 \epsilon 2} & \sigma_{\eta 1 \eta 2} & \sigma^2_{\eta 2} \end{pmatrix}\right) \tag{C.1}$$

where the subscript 1 refers to variable 1 (in our case, UK productivity) and 2 to variable 2 (US or global productivity, or UK trade openness), and the covariance matrix is expressed in terms of covariances between the different pairs of trend and cycle shocks. In practice, for the sampling algorithm described below, we will be parameterising standard deviations ($\sigma_{\epsilon 1}$, $\sigma_{\epsilon 2}$, $\sigma_{\eta 1}$ and $\sigma_{\eta 2}$) as well as the different correlation coefficients ρ_{x1x2} from the definition $\rho_{x1x2} = \sigma_{x1x2}/\sigma_{x1}\sigma_{x2}$.

Note that the covariance matrix Σ in (C.1) is symmetric and stacked so that the cycle shocks are located in the upper left quadrant and the trend components in the lower right quadrant of the matrix. We denote the different quadrants of the matrix as Σ_{11}, Σ_{21}, Σ_{22} and Σ_{12} in counter clock-wise order.

Next, it is useful to stack the observable and unobservable time series as follows:

$$\begin{aligned} \mathbf{y} &= (y_{11}, y_{21}, ..., y_{1T}, y_{2T})' \\ \boldsymbol{\tau} &= (\tau_{11}, \tau_{21}, ..., \tau_{1T}, \tau_{2T})' \\ \mathbf{c} &= (c_{11}, c_{21}, ..., c_{1T}, c_{2T})' \\ \mathbf{u}^{\tau} &= (\eta_{11}, \eta_{21}, ..., \eta_{1T}, \eta_{2T})' \\ \mathbf{u}^{c} &= (\epsilon_{11}, \epsilon_{21}, ..., \epsilon_{1T}, \epsilon_{2T})' \end{aligned}$$

Posterior draws can then be obtained by sequentally sampling from the following densities:

$$\begin{aligned} 1.\ p(\boldsymbol{\tau} &\mid \mathbf{y}, \boldsymbol{\phi}, \boldsymbol{\tau}_0, \boldsymbol{\tau}_{-1}, \Sigma) \\ 2.\ p(\boldsymbol{\phi} &\mid \mathbf{y}, \boldsymbol{\tau}, \boldsymbol{\tau}_0, \boldsymbol{\tau}_{-1}, \Sigma) \\ 3.\ p(\boldsymbol{\tau}_0, \boldsymbol{\tau}_{-1} &\mid \mathbf{y}, \boldsymbol{\tau}, \boldsymbol{\phi}, \Sigma) \\ 4.\ p(\Sigma &\mid \mathbf{y}, \boldsymbol{\tau}, \boldsymbol{\phi}, \boldsymbol{\tau}_0, \boldsymbol{\tau}_{-1},) \end{aligned}$$

In what follows, we describe the main features of these four steps, but refer the reader to the original paper by Grant and Chan (2017a) for more details.

In Step 1, the model is first rewritten with the stacked variables defined above:

$$
\begin{aligned}
\mathbf{y} &= \boldsymbol{\tau} + \mathbf{c} \\
H_\phi \mathbf{c} &= \mathbf{u}^c \\
H\boldsymbol{\tau} &= \widetilde{\boldsymbol{\alpha}} + \mathbf{u}^\tau
\end{aligned}
\tag{C.2}
$$

where $\widetilde{\boldsymbol{\alpha}} = (2\tau_{1,0} - \tau_{1,-1}, 2\tau_{2,0} - \tau_{2,-1}, -\tau_{1,0}, -\tau_{2,0}, 0, 0, ..., 0, 0)'$ ($\tau_{i,t}$ is observation at time t for variable i) and

$$
H = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & \cdots & 0 \\
-2 & 0 & 1 & 0 & 0 & 0 & \cdots & 0 \\
0 & -2 & 0 & 1 & 0 & 0 & \cdots & 0 \\
1 & 0 & -2 & 0 & 1 & 0 & \cdots & 0 \\
0 & 1 & 0 & -2 & 0 & 1 & \cdots & 0 \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\
0 & \cdots & 0 & 1 & 0 & -2 & 0 & 1
\end{bmatrix},
$$

$$
H_\phi = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & \cdots & 0 \\
-\phi_{11} & 0 & 1 & 0 & 0 & 0 & \cdots & 0 \\
0 & -\phi_{21} & 0 & 1 & 0 & 0 & \cdots & 0 \\
-\phi_{12} & 0 & -\phi_{11} & 0 & 1 & 0 & \cdots & 0 \\
0 & -\phi_{22} & 0 & -\phi_{21} & 0 & 1 & \cdots & 0 \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\
0 & \cdots & 0 & -\phi_{22} & 0 & -\phi_{21} & 0 & 1
\end{bmatrix}
$$

where ϕ_{ij} denotes the lag j for variable i. Both H and H_ϕ are square matrices with a unit determinant, and hence invertible. Given the parameters of the model ($\boldsymbol{\phi}$,$\mathbf{y}, \boldsymbol{\tau}, \boldsymbol{\tau}_0, \boldsymbol{\tau}_{-1}, \Sigma$), we have the following distribution for the unobservable variables of the model:

$$
\begin{pmatrix} \mathbf{c} \\ \boldsymbol{\tau} \end{pmatrix} \sim N\left(\begin{pmatrix} \mathbf{0} \\ \boldsymbol{\alpha} \end{pmatrix}, \begin{pmatrix} H_\phi^{-1}(I_T \otimes \Sigma_{11})H_\phi^{'-1} & H_\phi^{-1}(I_T \otimes \Sigma_{12})H^{'-1} \\ H^{'-1}(I_T \otimes \Sigma_{21})H_\phi^{-1} & H^{'-1}(I_T \otimes \Sigma_{22})H^{-1} \end{pmatrix} \right)
\tag{C.3}
$$

where I_T is an identity matrix of sample size T and $\boldsymbol{\alpha} = H^{-1}\widetilde{\alpha}$.

It can then be shown that (see e.g. Kroese and Chan (2014) and Grant and Chan (2017a) and (2017b) τ is sampled from the following distribution:

$$(\boldsymbol{\tau}|\mathbf{y}, \boldsymbol{\phi}, \boldsymbol{\tau}_0, \boldsymbol{\tau}_{-1}, \Sigma) \sim N(\widetilde{\boldsymbol{\tau}}, K_\tau^{-1}) \tag{C.4}$$

where

$$K_\tau = H'(I_T \otimes (\Sigma_{22}^{-1} * I_2))H + B'(I_T \otimes (\Sigma_{11} - ((\Sigma_{21}/\Sigma_{22})\Sigma_{12})^{-1} I_2)B \tag{C.5}$$

and

$$\begin{aligned} \widetilde{\boldsymbol{\tau}} &= K_\tau^{-1}((H'(I_T \otimes (\Sigma_{22}^{-1} * I_2))H)\alpha \\ &\quad + B'(I_T \otimes (\Sigma_{11} - ((\Sigma_{21}/\Sigma_{22})\Sigma_{12})^{-1} I_2)(H_\phi \mathbf{y} - a))) \end{aligned} \tag{C.6}$$

with

$$a = -I_T \otimes (\Sigma_{21}/\Sigma_{22})H\alpha \text{ and } B = H_\phi + I_T \otimes (\Sigma_{21}/\Sigma_{22})H$$

where / indicates right-division of a matrix.

Next, in Step 2, we sample $\boldsymbol{\phi}$ and start by noting that the joint distribution for $\mathbf{u}^c$ and $\boldsymbol{\tau}$ is the following:

$$\begin{pmatrix} \mathbf{u}^c \\ \boldsymbol{\tau} \end{pmatrix} \sim N\left(\begin{pmatrix} \mathbf{0} \\ \boldsymbol{\alpha} \end{pmatrix}, \begin{pmatrix} I_T \otimes \Sigma_{11} & (I_T \otimes \Sigma_{12})H'^{-1} \\ (I_T \otimes \Sigma_{21})H^{-1} & H'^{-1}(I_T \otimes \Sigma_{22})H^{-1} \end{pmatrix} \right) \tag{C.7}$$

It is helpful to write the process for the cycle in the following form:

$$\mathbf{c} = X_\phi \boldsymbol{\phi} + \mathbf{u}^c \tag{C.8}$$

where X_ϕ is a $(2Tx4)$ matrix of lagged values of c_t. Then, we have the following distribution for sampling $\boldsymbol{\phi}$:

$$(\boldsymbol{\phi}|\mathbf{y}, \boldsymbol{\tau}, \boldsymbol{\tau}_0, \boldsymbol{\tau}_{-1}, \Sigma) \sim N(\widehat{\boldsymbol{\phi}}, K_\phi^{-1}) \tag{C.9}$$

where

$$K_\phi = V_\phi^{-1} + X_\phi'(I_T \otimes (\Sigma_{11} - ((\Sigma_{21}/\Sigma_{22})\Sigma_{12})^{-1} I_2)X_\phi \tag{C.10}$$

and

$$\begin{aligned} \widehat{\boldsymbol{\phi}} &= K_\phi^{-1}(V_\phi^{-1}\boldsymbol{\phi}_0 + X_\phi'(I_T \otimes (\Sigma_{11} - ((\Sigma_{21}/\Sigma_{22})\Sigma_{12})^{-1})\mathbf{c} \\ &\quad - (I_T \otimes (\Sigma_{21}/\Sigma_{22}))H(\boldsymbol{\tau} - \boldsymbol{\alpha})) \end{aligned} \tag{C.11}$$

with prior $\boldsymbol{\phi} \sim N(\boldsymbol{\phi}_0, V_\phi)$. A draw is created by the acceptance-rejection method, i.e., sampling from $N(\widehat{\boldsymbol{\phi}}, K_\phi^{-1})$ until $\boldsymbol{\phi} \in \mathbb{R}$.

Next, in Step 3, we sample jointly for τ_0 and τ_{-1} (where both of these are $(2x1)$ vectors of pre-sample values for the two variables) by first writing $\boldsymbol{\alpha} = X_\delta \boldsymbol{\delta}$, where $\boldsymbol{\delta}$ is a $(4x1)$ vector of all the pre-sample values $(\tau_{1,0}, \tau_{2,0}, \tau_{1,-1}, \tau_{2,-1})$ and

$$X_\delta = \begin{bmatrix} 2 & 0 & -1 & 0 \\ 0 & 2 & 0 & -1 \\ 3 & 0 & -2 & 0 \\ 0 & 3 & 0 & -2 \\ 4 & 0 & -3 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ T+1 & 0 & -T & 0 \\ 0 & T+1 & 0 & -T \end{bmatrix} \tag{C.12}$$

Then we have:

$$(\boldsymbol{\tau}_0, \boldsymbol{\tau}_{-1} | \mathbf{y}, \boldsymbol{\tau}, \boldsymbol{\phi}, \Sigma) \sim N(\widehat{\boldsymbol{\delta}}, K_\delta^{-1}) \tag{C.13}$$

where

$$K_\delta = V_\delta^{-1} + X_\delta' H' (I_T \otimes (\Sigma_{22} - ((\Sigma_{12}/\Sigma_{11})\Sigma_{12})^{-1} I_2) H X_\delta \tag{C.14}$$

with $V_\delta = diag(V_\tau, V_\tau)$ (the prior for $\boldsymbol{\tau}_0, \boldsymbol{\tau}_{-1} \sim N(\boldsymbol{\tau}_{00}, V_\tau)$) and

$$\begin{aligned} \widehat{\boldsymbol{\delta}} &= K_\delta^{-1}(V_\delta^{-1}\boldsymbol{\delta}_0 + X_\delta' H' (I_T \otimes (\Sigma_{22} - ((\Sigma_{12}/\Sigma_{11})\Sigma_{12})^{-1} I_2) \\ &\quad * H(\boldsymbol{\tau} - H^{-1}(I_T \otimes (\Sigma_{12}/\Sigma_{11}))\mathbf{u}^c)) \end{aligned} \tag{C.15}$$

with $\boldsymbol{\delta}_0 = (\tau_{1,00}, \tau_{2,00}, \tau_{1,00}, \tau_{2,00})'$.

Finally, in Step 4, we sample for Σ by first denoting $\mathbf{u}^\tau = H\boldsymbol{\tau} - \widetilde{\boldsymbol{\alpha}}$, stacking the error terms as $\mathbf{u} = [\mathbf{u}_1^c, \mathbf{u}_2^c, \mathbf{u}_1^\tau, \mathbf{u}_2^\tau]$ and the prior of Σ as Σ_0 and then drawing from the inverse Wishart distribution:

$$\Sigma \sim W^{-1}(\Sigma_0 + \mathbf{u}'\mathbf{u}, df) \tag{C.16}$$

Integrated likelihood

For model comparison, integrated likelihood functions are calculated for the bivariate models with the methods introduced by Chan and Grant (2016) and Grant and Chan (2017a). Efficient band matrix routines can be used to evaluate the likelihoods. The reader is referred to the original paper for further details and derivation of the likelihood function. Here, we state the log-likelihood function, which takes the following form:

$$
\begin{aligned}
L &= \frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln|\Sigma_{22}| + \sum(\ln(diag(c\Omega_{11}))) \\
&\quad - \sum(\ln(diag(cK_\tau))) - \frac{1}{2}(((\mathbf{y}_1 - \Omega_{11}^{-1}\mathbf{d})'\Omega_{11}(\mathbf{y}_1 - \Omega_{11}^{-1}\mathbf{d}) \\
&\quad + \boldsymbol{\alpha}'H'(I_T \otimes (\Sigma_{22}^{-1}I_2))H\boldsymbol{\alpha} - (cK_\tau\widehat{\boldsymbol{\tau}})'cK_\tau\widehat{\boldsymbol{\tau}})
\end{aligned}
\tag{C.17}
$$

where

$$
\begin{aligned}
\Omega &= H_\phi'(I_T \otimes (\Sigma_{21}/\Sigma_{22}))H_\phi \\
\mathbf{d} &= P\Omega(H_\phi^{-1}a) - \Omega_{12}\mathbf{y}_2
\end{aligned}
\tag{C.18}
$$

and Ω_{11} denotes a matrix of every even row and column of Ω, Ω_{12} denotes a matrix of every even row and odd column of Ω, $c\Omega_{11}$ is upper triangular Cholesky factor of Ω and cK_τ is lower triangular Cholesky factor of K_τ. Furthermore, P is a $(Tx2T)$ sparse matrix of ones in every $(t, 2t-1)$ cell and zeros elsewhere and $\mathbf{y}_1$ and $\mathbf{y}_2$ denote data for UK and the 2nd variable, respectively.

D Charts

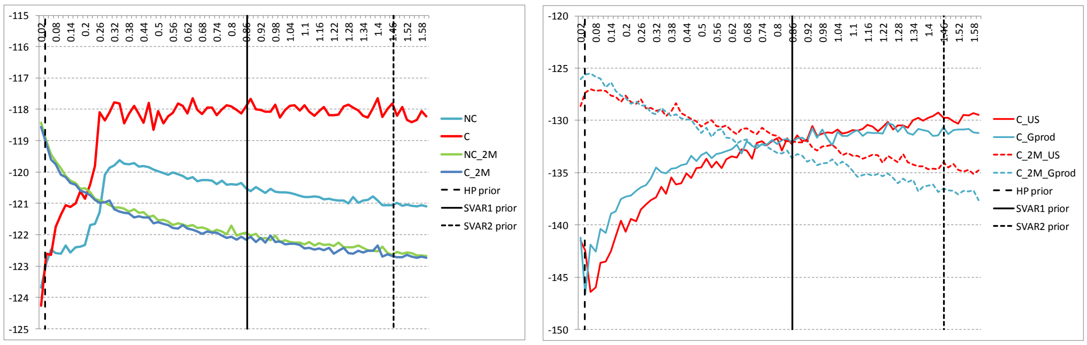

Figure 1: Marginal data likelihood - grid search

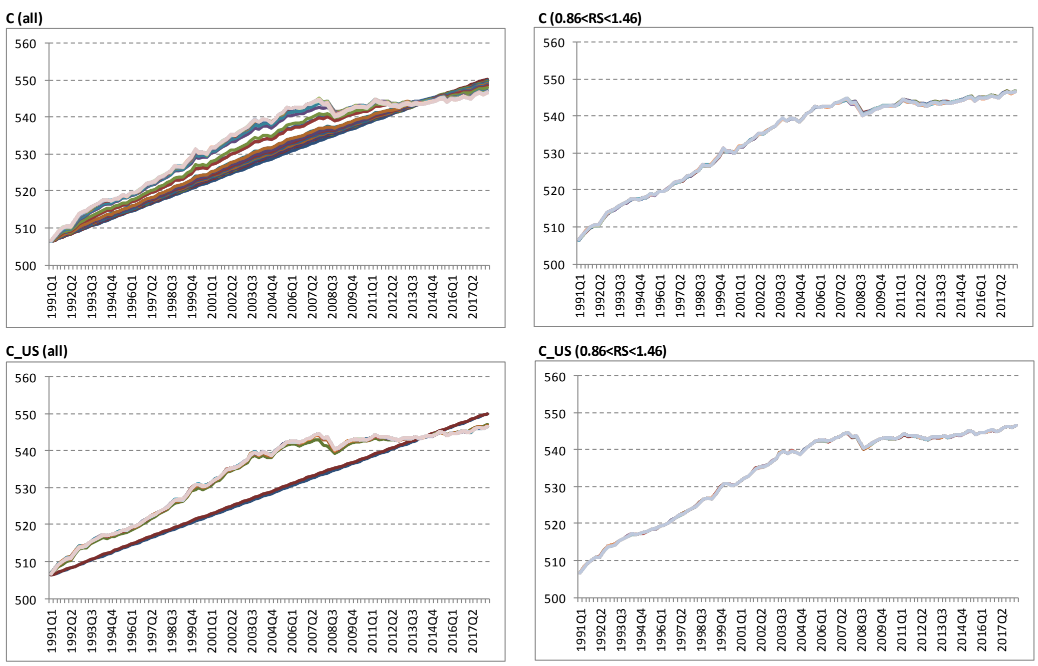

Figure 2: Trend value with different RS (selected models)

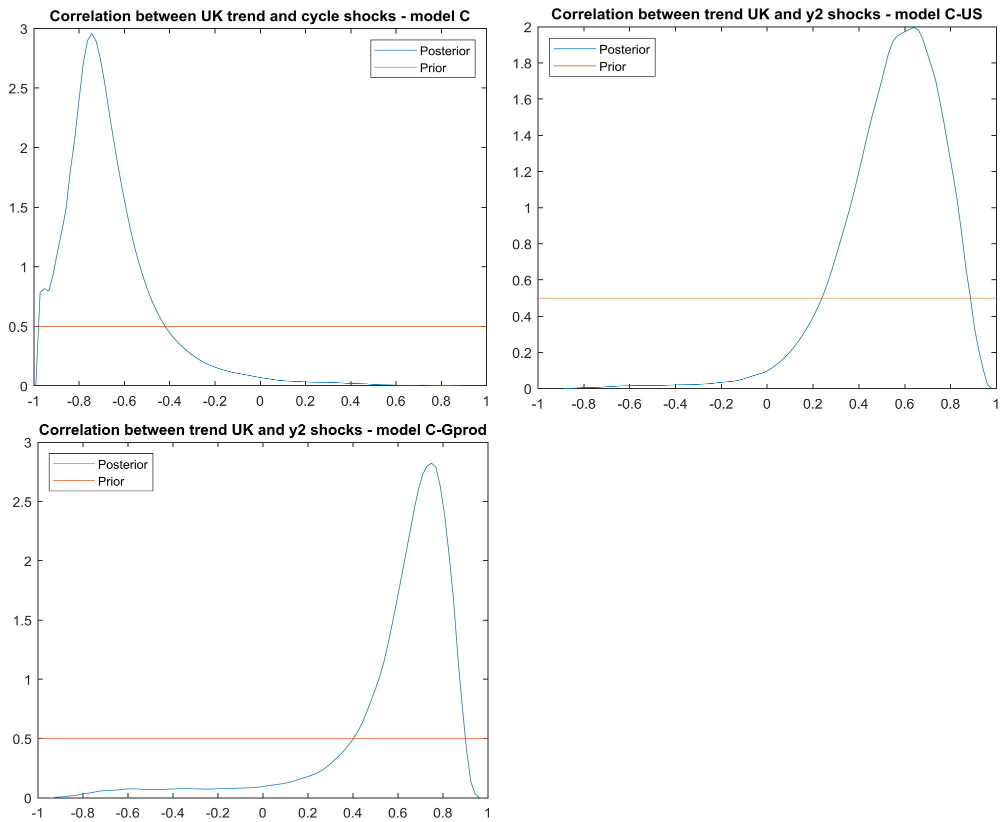

Figure 3: Posterior distributions for correlation coefficients between different trend and cycle shocks

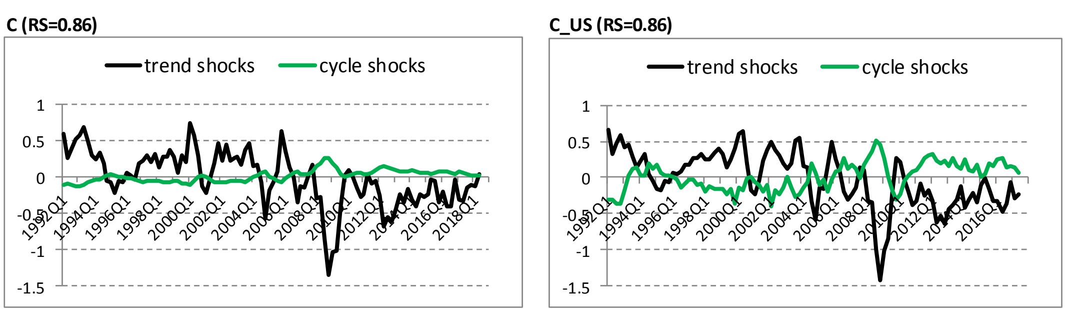

Figure 4: Trend and cycle shocks as percent of productivity for selected models, 4-quarter moving averages

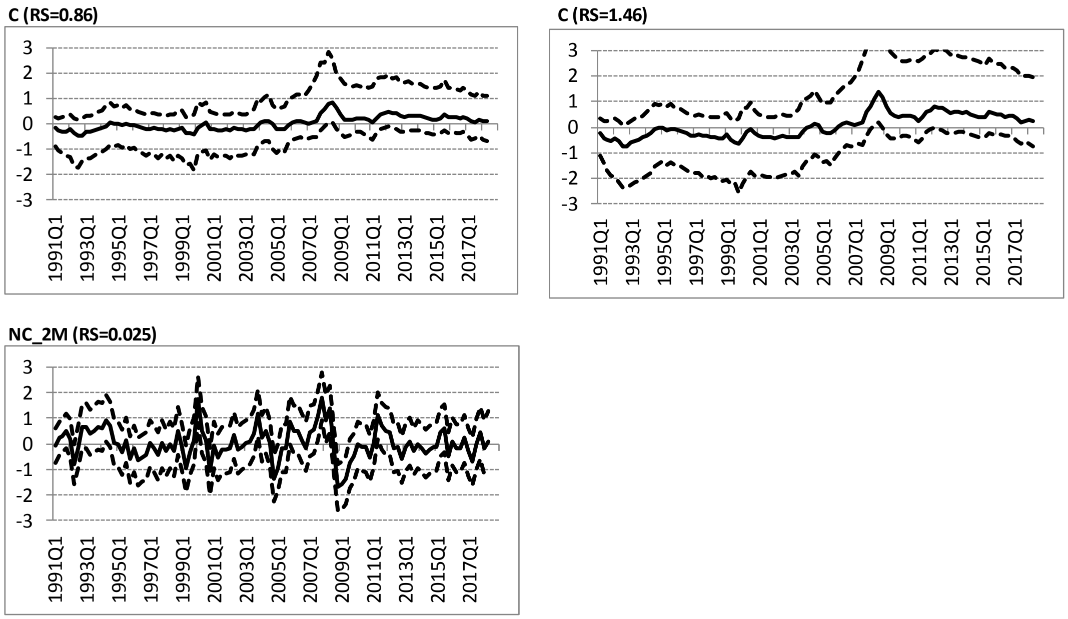

Figure 5: Cyclical productivity components as percent of productivity, univariate models, 90 percent confidence intervals

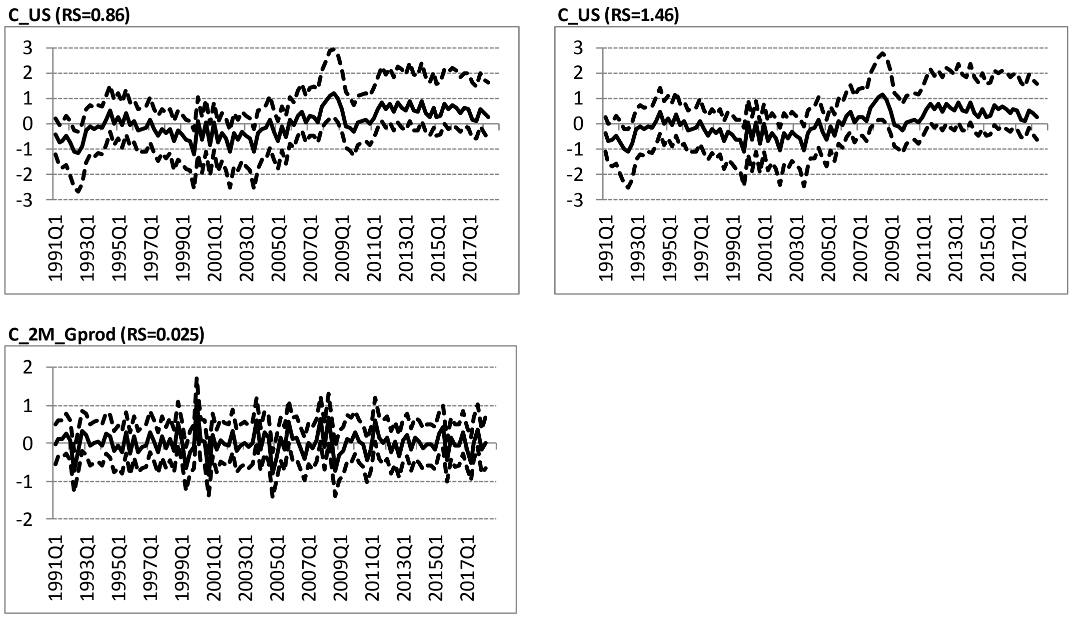

Figure 6: Cyclical productivity components as percent of productivity, bivariate models, 90 percent confidence intervals

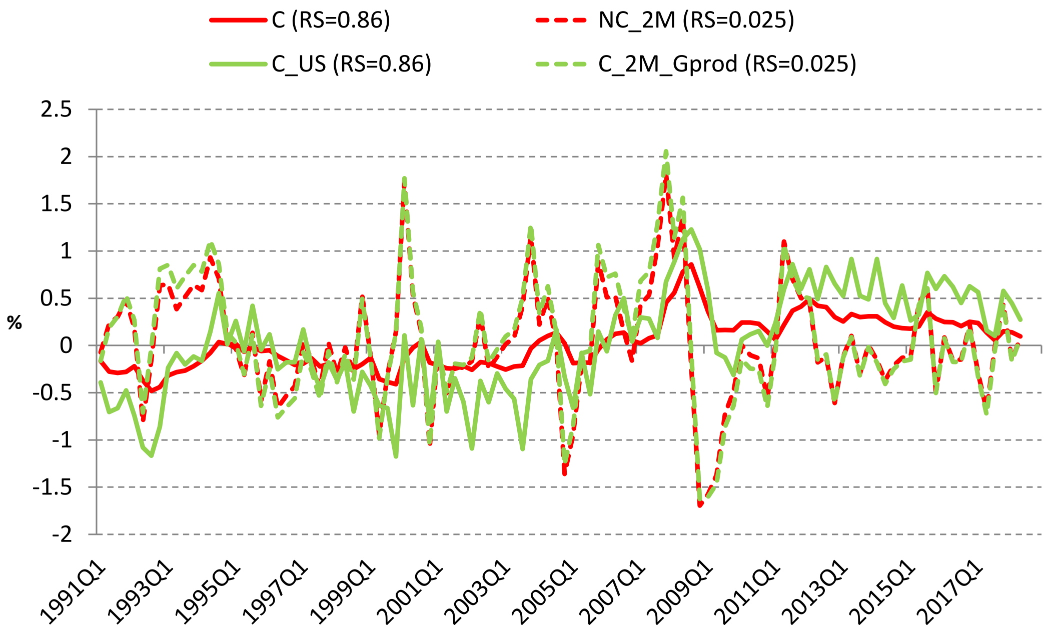

Figure 7: Cyclical productivity component in selected models

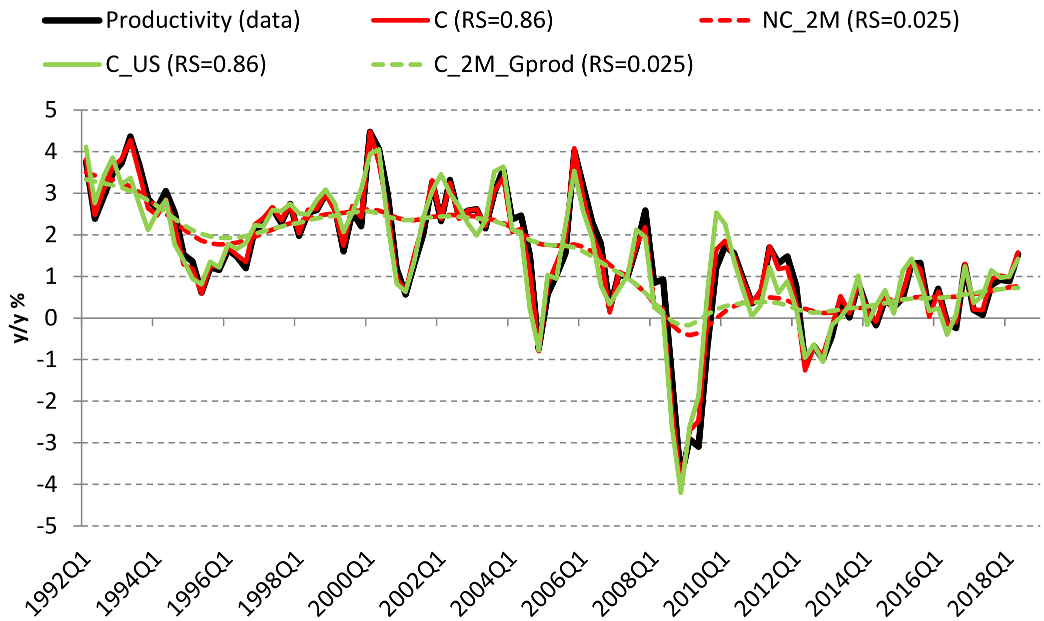

Figure 8: Trend productivity growth in selected models

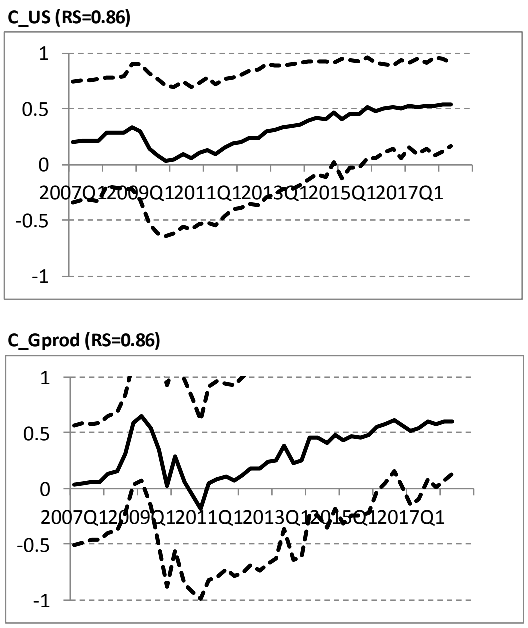

Figure 9: Rolling end-point estimates of trend correlations in selected bivariate models, 90 percent confidence intervals

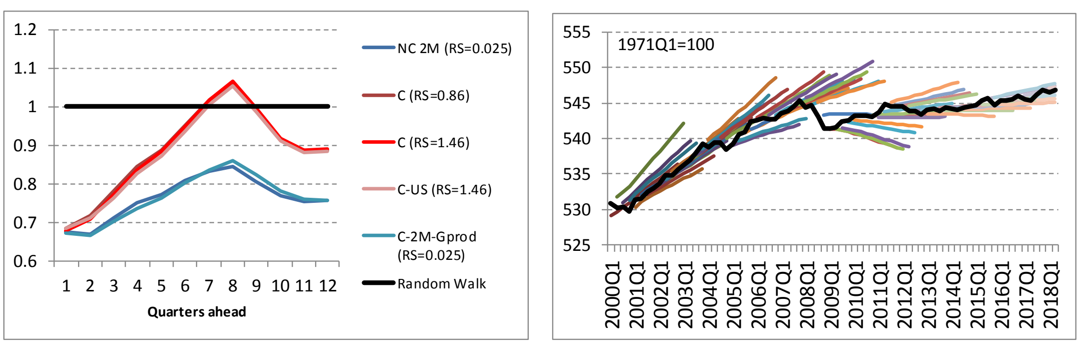

Figure 10: Forecast results. Theil U statistics (LHS) and pseudo-RT forecast for NC-2M model (RHS)

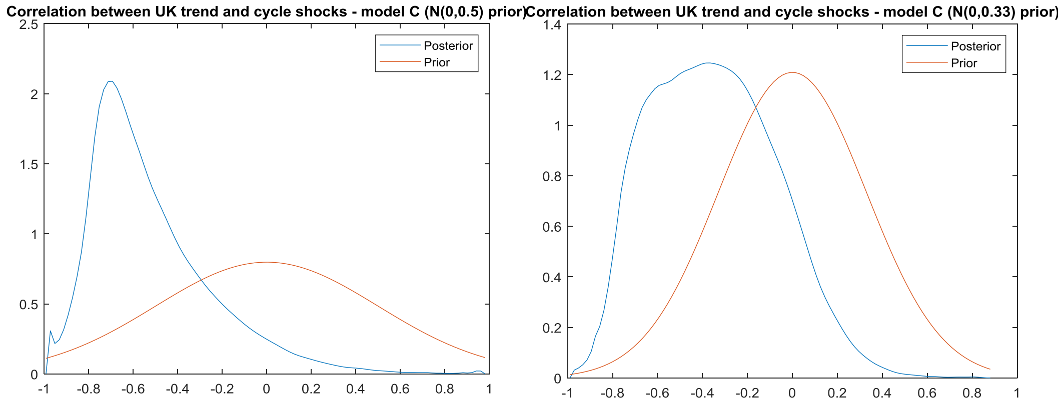

Figure 11: Normal priors for rho (C model)

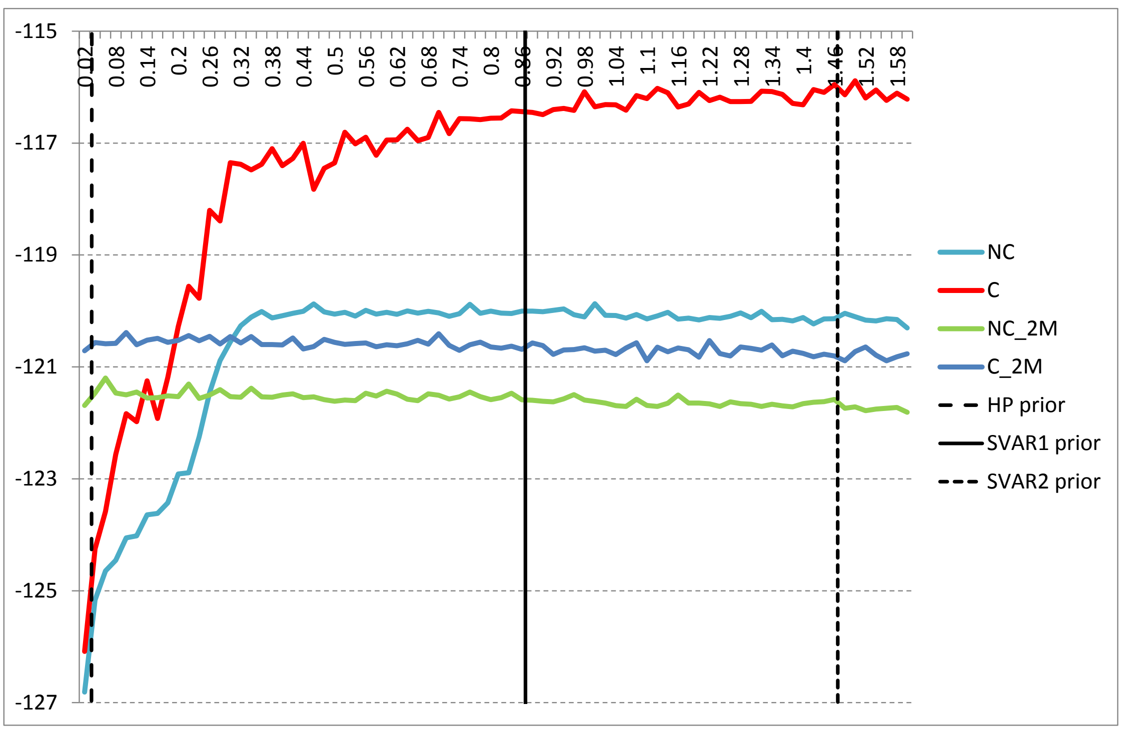

Figure 12: Marginal data likelihood for univariate models - grid search (normal priors)

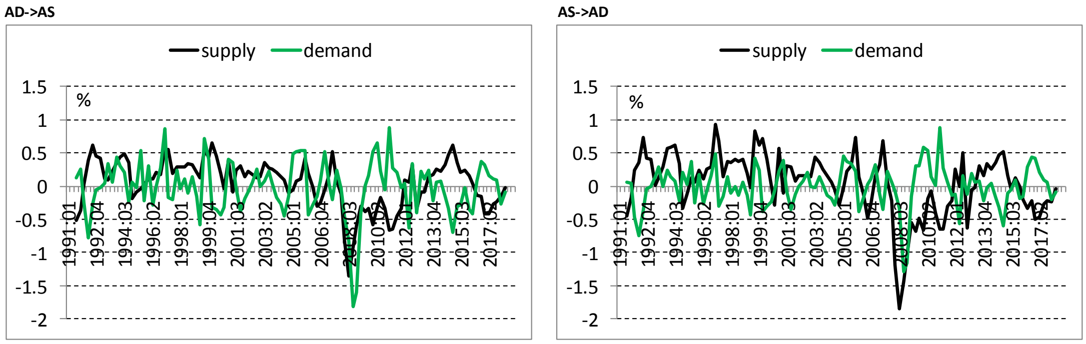

Figure 13: SVAR demand and supply shocks

References

[1] Andrews, D., C. Criscuolo and P. Gal (2016), "The best versus the rest: the global productivity slowdown, divergence across firms and the role of public policy", OECD Productivity Working Papers No. 5

[2] Barnett, A., S. Batten, A. Chiu, J. Franklin and M Sebastiá-Barriel (2014a), "The UK productivity puzzle", Bank of England Quarterly Bulletin 2014Q2

[3] Barnett, A., A. Chiu, J. Franklin and M Sebastiá-Barriel (2014b), "The productivity puzzle: a firm-level investigation into employment behaviour and resource allocation over the crisis", Bank of England Working Paper No. 495

[4] Bartelsmann, E., J. Haltiwanger, S. Scarpetta (2013), "Cross-Country Differences in Productivity: The Role of Allocation and Selection", American Economic Review, 103(1), pp. 305-334

[5] Bartelsmann, E., J. Haskel and R. Martin (2008), "Distance to Which Frontier? Evidence on Productivity Convergence from International Firm-level Data", CEPR Discussion Paper No. 7032

[6] Basistha, A. (2007), "Trend-cycle correlation, drift break and the estimation of trend and cycle in Canadian GDP", Canadian Journal of Economics, 40(2), pp. 584-606

[7] Bergeaud, A., G. Cette and R. Lecat (2016a, "Productivity trends from 1890 to 2012 in advanced countries", Review of Income and Wealth, 62, pp. 420-444

[8] Bergeaud, A., G. Cette and R. Lecat (2016b), "The role of production factor quality and technology diffusion in 20th century productivity growth", Banque de France Working Paper No. 588

[9] Blanchard, O. and D. Quah (1989), "The Dynamic Effects of Aggregate Demand and Supply Disturbances", American Economic Review, 79, pp. 655-673

[10] Chan, J.C.C. and E. Eisenstat (2015), "Marginal Likelihood Estimation with the Cross-Entropy Method", Econometric Reviews, 34, pp. 256-285

[11] Chan, J.C.C. and A.L. Grant (2016), "Fast Computation of the Deviance Information Criterion for Latent Variable Models", Computational Statistics and Data Analysis, 100, pp. 847–59

[12] Chan, J.C.C. and I. Jeliazkov (2009), "Efficient Simulation and Integrated Likelihood Estimation in State Space Models", International Journal of Mathematical Modelling and Numerical Optimisation, 1, pp. 101–20

[13] Clark, P. K. (1987), "The Cyclical Component of U.S. Economic Activity", The Quarterly Journal of Economics, 102(4), pp. 797-814

[14] Cover, J., W. Enders and C.J. Hueng (2003), "Using the Aggregate Demand-Aggregate Supply Model to Identify Structural Demand-Side and Supply-Side Shocks: Results Using a Bivariate VAR", Journal of Money, Credit and Banking, 38(3), pp. 777-790

[15] Goodridge, P., J. Haskel and G. Wallis (2014), "The UK Productivity Puzzle is a TFP Puzzle: Current Data and Future Predictions", Imperial College Discussion Paper 2014/9

[16] Grant, A.L. and J.C.C. Chan (2017a), "A Bayesian Model Comparison for Trend-Cycle Decompositions of Output", Journal of Money, Credit and Banking, 49(2-3), pp. 525-552

[17] Grant, A.L. and J.C.C. Chan (2017b), "Reconciling output gaps: Unobserved components model and Hodrick-Prescott filter", Journal of Economic Dynamics and Control, pp. 525-552

[18] Kamber, G., J. Morley and B. Wong (2018), "Intuitive and reliable estimates of the output gap from a Beveridge-Nelson filter", The Review of Economics and Statistics, Vol 100(3), pp. 550-566

[19] Kim, C.-J. and J. Kim (2013), "The 'Pile-up Problem' in Trend-Cycle Decomposition of Real GDP: Classical and Bayesian Perspectives", MPRA Paper 51118, University Library of Munich, Germany

[20] Kroese, D.P. and J.C.C. Chan (2014), "Statistical Modeling and Computation", New York, NY:Springer

[21] Lewis, K. F. and F. Vazquez-Grande (2017), "Measuring the Natural Rate of Interest: Alternative Specifications, Finance and Economics Discussion Series 2017-059. Washington: Board of Governors of the Federal Reserve System

[22] Mitra, S. and T.M. Sinclair (2012), "Output fluctuations in the G-7: an unobserved components approach", Macroeconomic Dynamics, 16, pp. 396-422

[23] Morley, J. (2007), "The Slow Adjustment of Aggregate Consumption to Permanent Income", Journal of Money, Credit and Banking, 39, pp. 615-638

[24] Morley, J., C. Nelson and E. Zivot (2003), "Why are the Beveridge-Nelson and unobserved-components decompositions of GDP so different?", The Review of Economics and Statistics, 85(2), pp. 235-243

[25] Oulton, N. and M Sebastiá-Barriel (2017), "Effects of financial crises on productivity, capital and output", Review of Income and Wealth, 63(1), February 2017

[26] Riley, R., C. Rosazza-Bondibene and G. Young (2015), "The UK productivity puzzle 2008-13: evidence from British businesses", Bank of England Working Paper No. 531

[27] Sinclair, T.M. (2009), "The Relationhips between Permanent and Transitory Movements in U.S. Output and the Unemployment Rate", Journal of Money, Credit and Banking, 41, pp. 529-542

[28] Stock, J. H. and M.W. Watson (1998), “Median Unbiased Estimation of Coefficient Variance in a Time-Varying Parameter Model”, Journal of the American Statistical Association, 93, pp. 349-358

[29] Trenkler, C. and E. Weber (2016), "On the identification of multivariate correlated unobserved components models", Economic Letters, 138, pp. 15-18

[30] Weber, E. (2011), "Analyzing U.S. Output and the Great Moderation by Simultaneous Unobserved Components", Journal of Money, Credit and Banking, 43, pp. 1579-1597