BANK OF ENGLAND

External MPC Unit

Discussion Paper No. 38

Estimation of short dynamic panels in the presence of cross-sectional dependence and dynamic heterogeneity

Robert Gilhooly, Martin Weale and Tomasz Wieladek

December 2012

This document is written by the External MPC Unit of the Bank of England

BANK OF ENGLAND

External MPC Unit

Discussion Paper No. 38

Estimation of short dynamic panels in the presence of cross-sectional dependence and dynamic heterogeneity

Robert Gilhooly,[1] Martin Weale[2] and Tomasz Wieladek[3]

Abstract

We propose a Bayesian approach to dynamic panel estimation in the presence of cross-sectional dependence and dynamic heterogeneity which is suitable for inference in short panels, unlike alternative estimators. Monte Carlo simulations indicate that our estimator produces less bias, and a lower root mean squared error, than existing estimators. The method is illustrated by estimating a panel VAR on sector level data for labour productivity and hours worked growth for Canada, Germany, France, Italy, the UK and the US from 1992 Q1 to 2011 Q3. We use historical decompositions to examine the determinants of recent output growth in each country. This exercise demonstrates that failure to take cross-sectional dependence into account leads to highly misleading results.

Keywords: Bayesian dynamic panel estimator, dynamic heterogeneity, cross-sectional dependence, labour productivity.

JEL classification: C11, C31, C33.

(1) External MPC Unit, Bank of England. Email: robert.gilhooly@bankofengland.co.uk
(2) External MPC Unit, Bank of England. Email: martin.weale@bankofengland.co.uk
(3) External MPC Unit, Bank of England. Email: tomasz.wieladek@bankofengland.co.uk

Information on the External MPC Unit Discussion Papers can be found at
www.bankofengland.co.uk/publications/Pages/externalmpcpapers/default.aspx

External MPC Unit, Bank of England, Threadneedle Street, London, EC2R 8AH

ISSN 1748-6203 (on-line)

Introduction

Currently available dynamic panel estimators, which allow for both dynamic heterogeneity[1] and cross-sectional dependence, are either not feasible (Phillips and Sul, 2003) or biased (Pesaran, 2006) in short[2] panels. This paper develops a Bayesian approach to estimating such models. Monte Carlo simulations suggest that our estimator is substantially less biased than the currently available alternatives for short panels. Our approach is illustrated by estimating a multivariate version, a panel VAR, on growth of labour productivity and hours worked for five sectors, which aggregate to the whole private sector, for Canada, Germany, France, Italy, the UK and the US (the G6) from 1992Q1 to 2011Q3. We then identify, using long-run restrictions, five distinct permanent labour productivity shocks. Based on these structural VAR estimates, we investigate whether recent output movements, defined as the sum of movements in labour productivity and hours worked, have been driven by permanent productivity or temporary demand shocks, an issue of substantial interest in economic policy circles.

An important maintained assumption in applied panel data studies is the independence of individual units in the cross section. As first noted by Stephan (1934), this is unlikely to hold in economic applications. This issue, commonly referred to as 'cross-sectional dependence', has been the subject of a rapidly-growing academic literature in recent years. There are now several standard ways to address this problem[3]: In the case of strongly exogenous regressors and $\mathbf{T}$ (the number of time-series observations) greater than $\mathbf{C}$ (the number of cross-section observations) the Seemingly Unrelated Regression (SUR) approach, first introduced by Zellner (1962), can address this issue. When $\mathbf{C}>\mathbf{T}$, this estimator becomes infeasible, but Robertson and Symons (2007) show that the introduction of a factor-structure on the residuals and estimation by maximum likelihood techniques is a feasible solution in that case. An alternative approach, proposed by Coakley, Fuertes and Smith (2002), uses principal components to

[1] Those dynamic panel estimators that allow for cross sectional dependence and are specifically designed for inference in short panels, such as the quasi maximum likelihood approach in Bai (2009) or the GMM approach in Sarafidis (2009), do not allow for dynamic heterogeneity.

[2] Throughout this paper, 'short' refers to the time-series dimension of the panel.

[3] See Sarafidis and Wansbeek (2012) for an extensive survey of the literature.

proxy for the unobserved common factors in the residual. But their estimator is consistent only when the explanatory variables are uncorrelated with the factors in the residuals. Pesaran (2006) proposes the common correlated effects (CCE) estimator, which relaxes this assumption and can be implemented by including the cross-sectional means of the dependent and independent variables into the regression.

Where the regression equation includes dynamic terms, additional complications, related to the degree of heterogeneity in the lagged dependent variable coefficients, may arise. One option is to pool the data, assuming identical autoregressive dynamics across all units.[4] But in the presence of cross-sectional heterogeneity in the lagged dependent variable coefficients, Pesaran and Smith (1995) show that pooling will result in asymptotically inconsistent estimates. They propose the mean group estimator, which involves the estimation of the regression equation country-by-country, as a solution to this problem. The practical implementation of this approach therefore requires a **T** that is sufficiently large to ensure unbiased coefficients in each cross section. Indeed, in the case of small **T**, Hsiao, Pesaran and Tahmiscioglu (1999) recommend a Bayesian approach to address dynamic heterogeneity bias for single-equation dynamic panel models. But none of these estimators addresses the issue of cross-sectional dependence. To address both dynamic heterogeneity and cross-sectional dependence, Phillips and Sul (2003) propose an FGLS-SUR estimator. When **C**>**T**, their estimator is infeasible and here Pesaran (2006) recommends the common correlated effects mean group (CCEMG) estimator. As in the case of the regular mean group estimator, small sample bias means that this approach is unlikely to work when **T** is small. Most recently, Bai (2009) and Sarafidis (2009) propose dynamic panel estimators that allow for cross-sectional dependence and are specifically designed for short panels, but do not allow for dynamic heterogeneity.

Our main contribution to this literature is the introduction of a dynamic panel estimator that allows for both dynamic heterogeneity and cross-sectional dependence for short panels. We follow the Bayesian approach proposed in Hsiao, Pesaran and

[4]See Goodhart and Hoffman (2008), Lzzetzki, Mendoza and Vegh (2010) or Towbin and Weber (2010) for an application of this approach in the panel VAR context.

Tahmiscioglu (1999) and Jarocinski (2010) and introduce unobservable factors in the error terms of this model to allow for cross-sectional dependence. This follows the idea of Bai (2009) who also treats the unobserved factor as an estimable parameter in short panels. Our estimator is likely to prove useful for applied research, as recent years have seen a proliferation of applied work with panel VARs. In such circumstances **T** is typically small compared to the number of coefficients, with both dynamic heterogeneity and cross-sectional dependence likely. In macroeconomics, panel VARs have been used to examine fiscal multipliers (Llzzetzki, Mendoza and Vegh, 2011; Corsetti, Meier and Mueller, 2012), the transmission of monetary policy (Goodhart and Hoffman, 2008; Jarocinski, 2010, Calza et al, 2011) and external shocks (Broda, 2004; Canova, 2005; Radatz, 2010) to macroeconomic aggregates across countries. In microeconomics, researchers have used this approach to examine the dynamics of earnings and hours worked among workers, (Vidangos, 2009) and financial development and firm behaviour (Love and Zicchino, 2006). Theoretical contributions include Holtz-Eakin, Newey and Rosen (1988) who develop a GMM estimator for stationary panel VARs with short panels. Similarly, Hsiao, Pesaran and Binder (2005) propose GMM and quasi maximum likelihood estimators for short panels which can be applied when the panel VAR is non-stationary or includes cointegrating relationships. But none of these studies allows for both dynamic heterogeneity and cross-sectional dependence. As future applications of the proposed estimator are therefore likely to involve the estimation of panel VARs, our discussion is focused on the VAR version, treating the single equation model as a special case.

We employ Monte Carlo simulations to investigate the bias of our proposed estimator, relative to the alternatives, in short panels. As an application of our technique, we estimate a VAR version of our model on sectoral labour productivity and hours growth data for the G-6 and identify sector-specific permanent productivity shocks using long-run restrictions. With the help of historical decompositions we then study whether output growth at sector level is driven predominantly by temporary demand or permanent labour productivity shocks in the aftermath of the global financial crisis.

Our Monte Carlo experiments suggest that our proposed estimator has substantially smaller bias and lower root mean squared error than the CCEMG or CCEP estimators proposed in Pesaran (2006), particularly when **T**=5, even when the cross section is relatively small (**C**=20). More importantly, our application suggests that taking cross-sectional dependence into account is important when examining the determinants of output growth weakness in the G6. For instance, failure to take cross-sectional dependence into account would lead a researcher to conclude that temporary demand shocks have been the most important drivers of recent weak UK output growth. Once cross-sectional dependence is accounted for, on the other hand, our model shows that permanent productivity shocks are just as important. Given this stark difference in results and policy implications, future applied work should therefore not ignore these issues and there might be some merit in a re-examination of past panel VAR research.

The remainder of the paper is set up in the following way: Section II describes our empirical model and the Gibbs sampling approach used to estimate it. Section III undertakes the Monte Carlo study, comparing alternative dynamic panel estimators to ours. Section IV presents the empirical application. Section V offers concluding remarks.

2. Model

In this paper we propose a new approach to estimating dynamic panel data models with heterogeneous coefficients and cross-sectional dependence in short panels. The most likely application of this type of estimator is panel VAR work, since that is typically the case when the number of coefficients is large with respect to the number of time-series observations, meaning that the effective number of time-series observations is small. To ease implementation among applied researchers, we therefore choose to describe, and derive the Gibbs sampler for, our model as a VAR.

In practical terms, we follow the approach proposed in Hsiao, Pesaran and Tahmiscioglu (1999) and Jarocinski (2010) and use the hierarchical linear model with exchangeable prior in the formulation of Gelman et al (2003).[5] In the panel VAR context, the idea underlying this model is that all cross-sectional units share a common mean. This

[5] See Lindley and Smith (1972) for the first discussion of exchangeable prior in linear regression models.

is similar to the Litterman (1986) prior assumption popular in economic forecasting, but rather than shrinking towards a random walk, we are shrinking towards a common mean, with the degree of shrinkage determined by the data. Other work that uses the exchangeable prior in panel VAR applications includes Canova (2005) and Cicarelli and Rebucci (2004), but they do not infer the degree of shrinkage from the data.

2.1 Model assumptions

The panel VAR[6] model we propose is the following:

$$\boldsymbol{Y_c} = \boldsymbol{X_c B_c} + \boldsymbol{E_c} \tag{1}$$

$$\boldsymbol{E_c} = \boldsymbol{F D_c} + \boldsymbol{U_c} \tag{2}$$

where $\boldsymbol{Y_c}$ is a matrix with $\boldsymbol{N}$ endogenous variables in the columns and time-series observations in the rows, in country $\boldsymbol{c}$, with the total number of countries $\boldsymbol{C}$. $\boldsymbol{X_c}$ contains the lags of the variables in $\boldsymbol{Y_c}$ and $\boldsymbol{B_c}$ is the array of associated coefficients. We assume that the corresponding matrix of VAR residuals $\boldsymbol{E_c}$ is made up of $\boldsymbol{M}$ unobservable factors, which are common across countries and are contained in the matrix $\boldsymbol{F}$. The matrix $\boldsymbol{D_c}$ is the matrix of factor loadings, allowing each factor to affect each equation differently. $\boldsymbol{U_c}$ is the matrix of the actual reduced form country-specific VAR innovations. This is assumed to be normally distributed with variance-covariance matrix $\boldsymbol{\Sigma_c}$. When $\mathbf{T}$ is small, estimates of $\boldsymbol{B_c}$ are likely to be imprecise and it may be more efficient to pool estimates across countries. In particular, we assume that the following prior for $\boldsymbol{B_c}$:

$$\boldsymbol{p}(\boldsymbol{B_c} \mid \boldsymbol{\overline{B}}, \boldsymbol{\Lambda_c}) = \boldsymbol{N}(\boldsymbol{\overline{B}}, \boldsymbol{\Lambda_c}) \tag{3}$$

where $\boldsymbol{\overline{B}}$ is the pooled mean across countries with the variance $\boldsymbol{\Lambda_c}$ determining the tightness of this prior. We follow Jarocinski (2010) and parameterize $\boldsymbol{\Lambda_c} = \boldsymbol{\lambda L_c}$. $\boldsymbol{\lambda}$ is treated as a hyper parameter and is estimated from the data, based on an inverse gamma

[6] The description of most the components of our proposed model closely follows the presentation of Jarocinski (2010). For brevity we just cover the bare necessities and focus most of our attention on our innovation, the factor structure in the residuals. See his work for more details on the remaining parts of the model.

distribution, while $\boldsymbol{L_c}$, as explained in detail below, is calibrated pre-estimation. The greater $\boldsymbol{\lambda}$ the larger the degree to which the country-specific coefficients are allowed to differ from the common mean. If $\boldsymbol{\lambda} \rightarrow \boldsymbol{\infty}$, this approach will lead to country-by-country estimates, while $\boldsymbol{\lambda} = \mathbf{0}$ implies pooling across all countries. The parameterisation of $\boldsymbol{\Lambda_c}$ in this manner has the econometrically convenient property that it is necessary only to estimate one hyper-parameter to determine the degree of heterogeneity in the coefficients. But there is of course one drawback: the coefficients in $\boldsymbol{B_c}$ may have different magnitudes. In specifying a single parameter that determines the degree of heterogeneity, there is therefore the risk that some coefficients are allowed to differ from the common mean by a small fraction of their own size, while others can differ by orders of magnitude. Following Jarocinski (2010) and an analogous procedure for the Litterman (1986) prior, $\boldsymbol{L_c}$ is a matrix of scaling factors used to address this problem. In particular, $\boldsymbol{L_c(k,n)} = \frac{\boldsymbol{\sigma_{cn}^2}}{\boldsymbol{\sigma_{ck}^2}}$, where $\boldsymbol{c}$ is the country, $\boldsymbol{n}$ the equation and $\boldsymbol{k}$ the number of the variable regardless of lag. $\boldsymbol{\sigma_{cn}^2}$ is the estimated variance of the residuals of a univariate auto-regression of the endogenous variable in equation $\boldsymbol{n}$, of the same order as the VAR, and is obtained pre-estimation. $\boldsymbol{\sigma_{ck}^2}$ is the corresponding variance for variable $\boldsymbol{k}$ and obtained in an identical manner. To the extent that unexpected movements in variables will reflect the difference in the size of VAR coefficients, scaling by this ratio of variances allows us to address this issue.

Jarocinski (2010) shows that based on these assumptions, the joint posterior of the model can be written as:

$$\prod_c |\boldsymbol{\Sigma_c}|^{\frac{T_c}{2}} \exp\left(-\frac{1}{2}\sum_c (\boldsymbol{y_c} - \widetilde{\boldsymbol{X}}_c\boldsymbol{\beta_c} - \widetilde{\boldsymbol{F}}\boldsymbol{d_c})'\left(\boldsymbol{\Sigma}_c^{-1}\otimes \boldsymbol{I}_{T_c}\right)(\boldsymbol{y_c} - \widetilde{\boldsymbol{X}}_c\boldsymbol{\beta_c} - \widetilde{\boldsymbol{F}}\boldsymbol{d_c})\right)$$

$$\boldsymbol{\lambda}^{-\frac{CNK}{2}}\boldsymbol{exp}(-\frac{\mathbf{1}}{\mathbf{2}}\sum_{\boldsymbol{c}}(\boldsymbol{\beta_c} - \overline{\boldsymbol{\beta}})'\boldsymbol{L}_c^{-1}\boldsymbol{\lambda}^{-\mathbf{1}}(\boldsymbol{\beta_c} - \overline{\boldsymbol{\beta}})\prod_c |\boldsymbol{\Sigma_c}|^{-\frac{N+1}{2}}\boldsymbol{\lambda}^{-\frac{v+2}{2}}\boldsymbol{exp}(-\frac{\mathbf{1}}{\mathbf{2}}\frac{\boldsymbol{s}}{\boldsymbol{\lambda}})$$

where $\widetilde{\boldsymbol{X}}_{\boldsymbol{c}} \equiv \boldsymbol{I}_{\boldsymbol{N}} \otimes \boldsymbol{X}_{\boldsymbol{c}}$, $\widetilde{\boldsymbol{F}} \equiv \boldsymbol{I}_{\boldsymbol{N}} \otimes \boldsymbol{F}$, $\boldsymbol{y}_{\boldsymbol{c}} \equiv \boldsymbol{vec}(\boldsymbol{Y}_{\boldsymbol{c}})$, $\boldsymbol{\beta}_{\boldsymbol{c}} \equiv \boldsymbol{vec}(\boldsymbol{B}_{\boldsymbol{c}})$, $\overline{\boldsymbol{\beta}} \equiv \boldsymbol{vec}(\overline{\boldsymbol{B}})$ and $\boldsymbol{d}_{\boldsymbol{c}} \equiv \boldsymbol{vec}(\boldsymbol{D}_{\boldsymbol{c}})$. Based on this posterior, it is easy to derive the conditional distributions for the Gibbs sampler of this model.

Before describing the Gibbs sampler, it is useful to lay out the assumptions regarding the factor model as well. The matrix $\boldsymbol{F}$ contains $\boldsymbol{M}$ factors. The idea of estimating factors to address cross-sectional dependence in short panels follows the approach in Bai (2009). We assume that these factors are independent with distribution $\boldsymbol{N}(\boldsymbol{0}, \boldsymbol{I}_{\boldsymbol{M}})$ at each point in time and that the VAR residuals $\boldsymbol{U}_{\boldsymbol{c}}$ are uncorrelated across countries, as the unobserved factors will absorb this cross-country correlation. Finally, it is assumed that $\boldsymbol{E}[\boldsymbol{U}_{\boldsymbol{c}}'\boldsymbol{F}] = \boldsymbol{0}$, the VAR residuals and the factors are orthogonal.

As with any factor model, there are issues of indeterminacy that need to be addressed ahead of estimation. First, there is a question of scale. One can multiply the matrix of factor loadings, $\boldsymbol{D}_{\boldsymbol{c}}$, by a constant d for all i, which gives $\widehat{\boldsymbol{D}_{\boldsymbol{c}}} = d\boldsymbol{D}_{\boldsymbol{c}}$. We can also divide the factor by d, which yields $\widehat{\boldsymbol{F}} = \frac{\boldsymbol{F}}{d}$. The scale of the model $\widehat{\boldsymbol{F}}\widehat{\boldsymbol{D}_{\boldsymbol{c}}}$ is thus observationally equivalent to the scale of the model $\boldsymbol{F}\boldsymbol{D}_{\boldsymbol{c}}$. In order to address this problem the scale of each factor is set to unity. Even then a choice remains as to the sign of $\boldsymbol{F}$. To identify the sign of the factors we restrict all of the factor loadings in one particular country to be positive. Finally, to identify multiple factors, additional assumptions may need to be made on the matrix of coefficients $\boldsymbol{D}_{\boldsymbol{c}}$.

2.2 The Gibbs Sampler

Under the assumptions laid out in the previous section, it is then easy to show that the model can be estimated by Gibbs sampling through iteratively drawing from the following distributions. The country-specific VAR coefficients $\boldsymbol{\beta}_{\boldsymbol{c}}$ are drawn from:

$$\boldsymbol{p}\big(\boldsymbol{\beta}_{\boldsymbol{c}} \mid \overline{\boldsymbol{\beta}}, \boldsymbol{F}, \boldsymbol{Y}_{\boldsymbol{c}}, \boldsymbol{\Lambda}_{\boldsymbol{c}}\big) = \boldsymbol{N}((\boldsymbol{G}_{\boldsymbol{c}})^{-1}(\boldsymbol{\Sigma}_{\boldsymbol{c}}^{-1} \otimes \boldsymbol{X}_{\boldsymbol{c}}')vec(\boldsymbol{Y}_{\boldsymbol{c}} - \boldsymbol{F}\boldsymbol{D}_{\boldsymbol{c}}) + \boldsymbol{\lambda}^{-1}\boldsymbol{L}_{\boldsymbol{c}}^{-1}\overline{\boldsymbol{\beta}}, (\boldsymbol{G}_{\boldsymbol{c}}{}^{-1})) \qquad \textbf{(4)}$$

where $\boldsymbol{G_c} = \boldsymbol{\Sigma}_c^{-1} \otimes \boldsymbol{X}_c' \boldsymbol{X_c} + \boldsymbol{\lambda^{-1} L_c^{-1}}$. $\overline{\boldsymbol{\beta}}$ is drawn from:

$$p\left(\overline{\boldsymbol{\beta}} \mid \boldsymbol{\beta}_c, \boldsymbol{\Lambda}_c \right) = \boldsymbol{N}((\boldsymbol{\lambda}^{-1} \textstyle\sum_c \boldsymbol{L}_c^{-1})^{-1} \boldsymbol{\lambda}^{-1} \textstyle\sum_c \boldsymbol{L}_c^{-1} \boldsymbol{\beta}_c, (\boldsymbol{\lambda}^{-1} \textstyle\sum_c \boldsymbol{L}_c^{-1})^{-1}) \tag{5}$$

$\boldsymbol{\lambda}$ is treated as a hyper parameter and drawn from the following inverse gamma 2 distribution:

$$p\left(\boldsymbol{\lambda} \mid \overline{\boldsymbol{\beta}}, \boldsymbol{\beta}_c, \boldsymbol{L}_c^{-1} \right) = \boldsymbol{IG}_2(\boldsymbol{s} + \textstyle\sum_c (\boldsymbol{\beta}_c - \overline{\boldsymbol{\beta}})' \boldsymbol{L}_c^{-1} (\boldsymbol{\beta}_c - \overline{\boldsymbol{\beta}}), \boldsymbol{CNK} + \boldsymbol{v}) \tag{6}$$

A completely non-informative prior with s and v set to 0 results in an improper posterior in this case. We therefore set both of the quantities to very small positive numbers, which is equivalent to assuming a weakly informative prior. But it is important to point out that $\boldsymbol{\lambda}$ is estimated from the total number of coefficients that this prior is applied to, namely the product of country (**C**), equations (**N**) and total number of coefficients in each equation (**K**). Given this large number of effective units, any weakly informative prior will be dominated by the data. Finally, the country-specific variance matrix of the residuals, $\boldsymbol{\Sigma_c}$, is drawn from an inverse-Wishart distribution:

$$p\left(\boldsymbol{\Sigma_c} \mid \boldsymbol{d_c}, \boldsymbol{\beta_c}, \widetilde{\boldsymbol{F}} \right) = \boldsymbol{IW}(\boldsymbol{U}_c' \boldsymbol{U}_c, \boldsymbol{T}_c) \tag{7}$$

where $\boldsymbol{U_c} = \boldsymbol{y_c} - \widetilde{\boldsymbol{X}}_c \boldsymbol{\beta_c} - \widetilde{\boldsymbol{F}} \boldsymbol{d_c}$ and $\boldsymbol{T_c}$ is the number of observations for each country.

As in Lopes and West (2004), each individual factor, $\boldsymbol{f^i}$, can be drawn from:

$$p(\boldsymbol{f^i} \mid \boldsymbol{Y}_c, \boldsymbol{B}_c, \boldsymbol{K}_i, \boldsymbol{\Sigma}_I) = \boldsymbol{N}(\left(\boldsymbol{1} + \boldsymbol{K}_i' \boldsymbol{\Sigma}_i^{-1} \boldsymbol{K}_i\right)^{-1} \boldsymbol{K}_i' \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{Y}_c^i - \boldsymbol{X}_c^i \boldsymbol{B}_c^i), (\boldsymbol{1} + \boldsymbol{K}_i' \boldsymbol{\Sigma}_i^{-1} \boldsymbol{K}_i)^{-1}) \tag{8}$$

where $\boldsymbol{K_i}$ is an $\boldsymbol{Cx1}$ vector the associated factor loadings, made up from the elements in $\boldsymbol{d_c}$. $\boldsymbol{\Sigma_i}$ is an $\boldsymbol{CxC}$ diagonal matrix of variances associated with equation **n** in country **c** that the factor loads on. The coefficients associated with the factors are drawn from:

$$p(\, \boldsymbol{d_c} \mid \boldsymbol{F}, \boldsymbol{Y_c}, \boldsymbol{B_c}, \boldsymbol{\Sigma_I}\,) = \boldsymbol{N}((\boldsymbol{F}\,'\boldsymbol{F})^{-1}\boldsymbol{F}'(\boldsymbol{Y_c} - \boldsymbol{X_c}\boldsymbol{B_c}), (\boldsymbol{F}\,'\boldsymbol{F})^{-1}) \tag{9}$$

3. A Monte Carlo Study

In this section we undertake a Monte Carlo study to compare the small sample properties of our proposed estimator with those suggested by previous work. We focus on a special case of the model described in section 2, where **N**, the number of equations is one. Previous studies that propose estimators which allow for both cross-sectional dependence and dynamic heterogeneity include Phillips and Sul (2003) and Pesaran (2006). The estimator in Phillips and Sul (2003) is applicable only when **T** >**C** and hence infeasible in our situation. The only other alternative estimator that accounts for both dynamic heterogeneity and cross-sectional dependence is the common correlated effects mean group **(CCEMG)** estimator proposed by Pesaran (2006), though he notes that his proposed estimator works better when T > 30. He also proposes the common correlated effects pooled **(CCEP)** estimator, though that does not allow for dynamic heterogeneity. We compare the performance of these two to our proposed estimator. As in Phillips and Sul (2003), we assume the following data generating process:

$$\boldsymbol{y_{i,t}} = \boldsymbol{\gamma_i}\boldsymbol{y_{i,t-1}} + \boldsymbol{u_{i,t}}$$

$$\boldsymbol{u_{i,t}} = \boldsymbol{\theta_i}\boldsymbol{f_t} + \boldsymbol{\varepsilon_{i,t}}$$

In our simulation, we consider the following cases:

Case I (Homogeneity and Low Cross sectional dependence). In this simulation, $\boldsymbol{\gamma_i}$=$\boldsymbol{\gamma} = \mathbf{.6}$ ∀ $\boldsymbol{i}$ and $\boldsymbol{\theta_i} \sim \boldsymbol{N}(\mathbf{.2}, \mathbf{.1})$.

Case II (Homogeneity and High Cross sectional dependence). In this simulation, $\boldsymbol{\gamma_i}$=$\boldsymbol{\gamma} = \mathbf{.6}$ ∀ $\boldsymbol{i}$ and $\boldsymbol{\theta_i} \sim \boldsymbol{N}(\mathbf{1}, \sqrt{\mathbf{.2}})$.

Case III (Heterogeneity and Low Cross sectional dependence). In this simulation, $\boldsymbol{\gamma_i} \sim \boldsymbol{N}(\mathbf{.6}, .\sqrt{\mathbf{.2}})$. ∀ $\boldsymbol{i}$ and $\boldsymbol{\theta_i} \sim \boldsymbol{N}(\mathbf{.2}, \sqrt{\mathbf{.1}})$.

Case IV (Heterogeneity and High Cross sectional dependence). In this simulation, $\boldsymbol{\gamma_i} \sim \boldsymbol{N}(\mathbf{.6}, .\sqrt{\mathbf{.2}})$. ∀ $\boldsymbol{i}$ and $\boldsymbol{\theta_i} \sim \boldsymbol{N}(\mathbf{1}, \sqrt{\mathbf{.1}})$.

Table 1: Estimates of γ

	CCEP	CCEMG	Bayes	CCEP	CCEMG	Bayes
		Case I			**Case II**	
T=5,C=20	.45	.33	.58	.45	.34	.57
T=10,C=20	.53	.45	.59	.54	.45	.56
T=20,C=20	.57	.51	.57	.56	.51	.55
T=5,C=50	.47	.36	.64	.46	.36	.65
T=10,C=50	.54	.44	.57	.53	.44	.54
T=20,C=50	.57	.52	.59	.57	.52	.56
		Case III			**Case IV**	
T=5,C=20	.40	.25	.54	.42	.24	.56
T=10,C=20	.49	.32	.48	.51	.34	.55
T=20,C=20	.56	.40	.46	.57	.39	.54
T=5,C=50	.45	.26	.57	.43	.25	.63
T=10,C=50	.53	.33	.48	.52	.33	.55
T=20,C=50	.59	.39	.47	.59	.39	.54

The numbers reported in the table above are the average of the estimate of $\boldsymbol{\gamma}$ with each method across all 250 replications.

In all experiments, $\boldsymbol{\varepsilon_{i,t}} \sim \boldsymbol{NID(0,1)}$, $\boldsymbol{f_t} \sim \boldsymbol{NID(0,1)}$. We repeat these experiments for T=5, 10, 20, C=20 and 50 to compare the small sample performance of these estimators. The number of replications for the Gibbs sampler is chosen to be 10000, with 5000 as burn in, retaining every tenth draw for inference.[7] The total number of replications for our

[7] The number of replications was chosen for computational reasons. It might at times take longer for the Gibbs sampling chain to converge, which is why our Bayesian estimates should be treated as lower bounds, in the sense that it may be possible to obtain estimates closer to the true values with more iterations.

experiment is 250. The resulting estimates of γ are shown in Table 1. Cleary, estimates of γ that are close to .6, the true value of the data generating process, suggest small bias. When T=5, the Bayes estimate is always closest to the true value, typically with a bias of less than 10%. For T > 10, the other estimators become relatively less biased and in some cases are better than the Bayes estimator. Nevertheless, our estimator produces the least biased coefficients in most circumstances.

Table 2: RMSE of γ

	CCEP	CCEMG	Bayes	CCEP	CCEMG	Bayes
		Case I			**Case II**	
T=5,C=20	.25	.34	.16	.24	.34	.17
T=10,C=20	.11	.18	.074	.10	.18	.087
T=20,C=20	.059	.098	.058	.062	.10	.065
T=5,C=50	.218	.30	.076	.22	.29	.086
T=10,C=50	.093	.17	.056	.09	.18	.074
T=20,C=50	.047	.09	.03	.04	.089	.05
		Case III			**Case IV**	
T=5,C=20	.32	.40	.18	.30	.42	.20
T=10,C=20	.18	.29	.18	.17	.28	.16
T=20,C=20	.13	.22	.16	.126	.22	.125
T=5,C=50	.26	.37	.13	.27	.38	.13
T=10,C=50	.14	.28	.14	.15	.28	.105
T=20,C=50	.097	.223	.15	.0873	.218	.092

The numbers reported in the table above are the average of the estimate of γ with each method across all 250 replications.

Table 2 presents the root mean squared error of γ across the 250 iterations. When **T** is small, our proposed estimator typically has a much smaller root mean square error than the alternatives. This again suggests that our approach is preferable to the other estimators in small **T** applications.

Out of the three estimators investigated in this section, the CCEMG estimator appears to show the largest bias, in particular when the number of time-series is small. This stems from the fact that implementation involves the estimation of individual regressions for each cross section. Such estimates are likely to be subject to severe small sample bias when the panel is short. But the bias does become much smaller when T=20. This clearly illustrates the principle that the effective length of the panel is determined by the number of time-series observations relative to the number of coefficients, rather than the absolute number of time-series observations. The CCEMG is therefore likely to suffer from the same problem in panel VAR models, where the number of relative time-series observations is small. Our estimator therefore seems to be clearly the preferred alternative to estimating such models.

4. Examining the determinants of recent G6 output growth outturns

The recovery in output growth, across the G6, following the 'Great Recession' seems to have been weaker than recoveries from past recessions. One side of the debate argues that this is the result of a shock to the supply capacity of the economy, while others maintain that weak demand is the culprit underlying weak output growth. Whether weak output growth is driven by the former or later has important implications for monetary policy. If low output growth is mostly driven by demand (supply), additional monetary stimulus to increase output, is likely to generate little (substantial) inflationary pressure. With all the caveats typically associated with such exercises, we therefore estimate the model proposed in section 2 on growth in labour productivity and hours worked for five sectors, which together make up the private sector macroeconomic aggregates, for the G6. Following Gali (1999), we then identify sector level permanent productivity shocks based on the assumption that in the long run, shocks to hours worked cannot affect labour productivity. Since the sum of labour productivity and hours growth is output growth, we can then decompose output growth into the contributions of permanent labour productivity and demand shocks since the start of global financial crisis in 2008Q3.

This section starts with a description of the data. We then proceed to the description of our model and the identification schemes we use. The last subsection discusses our results based on a model with and without cross-sectional dependence.

4.1 Data

Descriptive statistics summarising the final series are shown in table 2 below. These are presented for the period from 1992Q2 to 2007Q2, so as to offer a picture of the six economies before the start of the recent economic crisis. In all countries, output grew the fastest in the information and communication sector, consistent with the presence of an 'information revolution' in these countries during this period. There is clearly a lot of heterogeneity in sectoral labour productivity and output growth rates between sectors and countries. This suggests that taking this into account could be important when attempting to disentangle the various sources underlying business cycle fluctuations. For more information on the data see Gilhooly, Weale and Wieladek (2012).

Table 2: Data Summary, Annualised Growth Rates 1992Q2 to 2007Q2

	Output		Hours		Productivity	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
	B-E : Industry					
Canada	2.0	5.0	-0.1	5.4	2.1	5.2
France	1.9	3.1	-1.9	1.7	3.7	2.7
Germany	1.0	6.5	-2.4	5.3	3.5	7.1
Italy	1.2	4.7	-0.5	3.5	1.7	5.1
United Kingdom	0.6	3.3	-3.1	4.1	3.8	4.3
United States	2.9	5.4	-1.3	3.7	4.2	4.6
	F : Construction					
Canada	1.9	7.3	1.9	9.4	0.1	8.6
France	0.4	4.2	-0.2	3.4	0.6	2.9
Germany	-1.2	11.8	-1.2	10.8	0.1	9.7
Italy	0.9	6.4	1.8	7.6	-0.9	8.5
United Kingdom	1.1	6.0	-0.7	6.9	1.8	7.1
United States	0.6	5.6	2.0	6.1	-1.4	6.0
	G-I : Services					
Canada	3.2	4.5	1.3	3.1	1.9	4.5
France	2.4	2.7	0.3	1.6	2.1	2.7
Germany	1.6	5.1	-0.3	2.5	1.9	5.8
Italy	1.6	3.7	0.2	4.4	1.5	5.9
United Kingdom	2.4	3.5	0.3	3.1	2.2	4.1
United States	4.2	3.7	0.7	2.4	3.5	3.8
	J : Information and communication					
Canada	5.4	3.9	1.9	7.6	3.6	8.1
France	4.7	3.9	1.4	2.5	3.3	3.9
Germany	4.5	10.6	1.1	3.8	3.5	11.4
Italy	5.3	9.1	2.4	4.6	2.9	9.6
United Kingdom	8.4	10.4	-1.1	7.1	9.5	10.5
United States	5.8	5.9	0.7	5.6	5.1	8.2
	K : Financial, insurance and Real Estate activities					
Canada	3.2	2.0	1.3	4.9	1.9	4.9
France	1.9	2.1	-0.1	1.9	2.0	3.2
Germany	2.2	4.2	-0.6	2.2	2.3	4.5
Italy	1.2	3.6	0.6	2.4	0.7	4.15
United Kingdom	4.5	4.0	1.4	2.6	3.1	5.3
United States	3.5	3.0	1.6	2.6	1.9	3.5

4.2 Estimation and model specification

Substituting **(2)** into **(1)**, the reduced-form VAR model we estimate is:

$$\boldsymbol{Y_{c,t}} = \sum_{l=1}^{L} \boldsymbol{A_{c,l}} \boldsymbol{Y_{c,t-l}} + \boldsymbol{D_c} \boldsymbol{F_t} + \boldsymbol{u_t} \qquad \boldsymbol{u_t} \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{\Sigma_c}) \qquad \textbf{(10)}$$

where $\boldsymbol{Y_{c,t}} = [\Delta \boldsymbol{lp_{1,c,t}} \dots \Delta \boldsymbol{lp_{s,c,t}} \dots \Delta \boldsymbol{lp_{5,c,t}}\ \Delta \boldsymbol{hours_{1,c,t}} \dots \Delta \boldsymbol{hours_{s,c,t}} \dots \Delta \boldsymbol{hours_{5,c,t}}]$[8] is a 10x1 vector of labour productivity and hours growth in each of the five sectors *s*. $\boldsymbol{u_t}$ is a normally distributed vector of reduced-form shocks with covariance matrix $\boldsymbol{\Sigma_c}$. $\boldsymbol{F_t}$ is a vector of unobserved factors to proxy for cross-sectional dependence. The number of equations **N** is ten, with the number of countries, $\boldsymbol{C}$, six with seventy-six as the number of time-series observations. The chosen lag length is two. With this number of lags the number of coefficients in each equation will be twenty, which means that the effective **T** (relative to the number of coefficients) is small. Every version of equation **(10)** presented in this paper is estimated by replicating the Gibbs sampler in section 2.2 200,000 times, discarding the first 100,000 as burn-in, and retaining every fiftieth draw to reduce autocorrelation among the draws.

What remains to be specified is the number of factors and their interaction with each VAR equation. As the explicit purpose of the factors is to account for cross-sectional dependence, we first test for the presence of this phenomenon in each equation. Previous work suggested several tests. With **C** > **T**, Breusch and Pagan (1980) proposed an LM-test, based on the pair-wise correlations of the residual of the model. Pesaran (2004) shows that the LM-test is inconsistent for large **C** and proposes the CD-test statistic instead, which he shows has good power and size properties in a variety of situations, such as in the presence of dynamic heterogeneity and multiple structural breaks. If pair-wise correlations are both positive and negative and can offset each other, the CD test statistic will lack power. This could arise in a situation when the mean factor loadings are zero. In this case, Pesaran, Ullah and Yamagata (2008) propose the bias-adjusted LM test instead. Their Monte Carlo simulations suggest that this test has generally good power

[8] Note that all of series were time-demeaned pre-estimation and hence our model does not feature country-specific constant terms.

properties, but they weaken slightly in the case of weakly exogenous regressors and dynamic heterogeneity. For this reason, we use both tests together to determine the appropriate number of factors and which equations they should enter, taking into account their relative strengths and weaknesses.

Table 3 (4) presents the results for the CD (bias-adjusted LM) test. Both tests are carried out on the residuals[9] of the estimated models referred to in the tables. Asymptotically both test statistics have a $N(0,1)$ distribution. An absolute value of greater than 1.96 (1.6) thus implies the rejection of the null hypothesis of cross-sectional independence at the 5% (10%) level. The CD test statistic for the model without any factors in row two of table 3 allows us to reject the null hypothesis of cross-sectional independence in equations 1, 3 and 6. The bias adjusted LM test on the other hand only weakly confirms cross-sectional independence in equations 1 and 6, but not in equation 3. We therefore proceed and re-estimate our model with factors in equations 1 and 6. The CD (bias-adjusted LM) test statistic for this model is presented in row three of table 3 (4). Now both tests also reject cross-sectional independence in equations 3, 7, 8 and 9. It may be surprising that additional factors are needed now. This stems from the fact that in our model, only one shrinkage parameter determines the degree of heterogeneity for the whole system of equations. The inclusion of factors in equation 1 and 6 will therefore not only affect the degree of shrinkage (pooling towards the common mean) in equations 1 and 6, but in all equations. In our case the estimated degree of heterogeneity increases, which may give more scope for finding cross-sectional dependence in the remaining equations. For equation 10, the LM test suggests the presence of cross-sectional dependence, while the CD test does not. A close inspection of the pair-wise correlations in equation 10 suggests that some large correlations are equal in magnitude, but different in sign, which is exactly the case when the CD statistic looses power, meaning that results from this test need to be interpreted with care. We therefore re-estimate the model with factors in equations 1,3,6,7,8,9 and 10. Even then both tests still reject the null hypothesis for equation 6. To address this remaining cross-sectional dependence, we include an

[9] Residuals have been constructed based on the estimated median coefficients of the model.

additional crisis factor in equation 6, which takes the value of zero pre 2008Q2 and loads on the UK and Italy, restricting the UK loading on the first factor to 0.[10] The last row of table 4 shows that in this case, all cross-sectional dependence appears to be removed. The final specification that we use in the structural VAR analysis below therefore includes one factor each in equations 1, 3,7,8,9 and 10 and two factors in equation 6.

Table 3 – CD test statistic for different models

Model / Equation	1	2	3	4	5	6	7	8	9	10
No Factors	3.20	.52	2.84	.22	.15	2.78	1.36	1.10	1.71	-.29
Factors in EQ 1 and EQ6	-1.32	.84	4.91	.39	.99	-4.36	4.21	4.74	5.60	.69
Factors in EQ1, EQ3, EQ6, EQ7, EQ8, EQ9,EQ10	-1.21	.79	-1.20	.34	.97	-4.15	-1.12	-2.51	-1.75	.63
Factors in EQ1, EQ3, EQ6, EQ7, EQ8, EQ9,EQ10 and additional factor in EQ6	-1.2	.9	-1.20	.6	1.0	-1.1	-.9	-1.95	-1.63	.67

[10] These choices were made based on the highest individual pair-wise correlations and the associated model residuals across countries in equation 6. Furthermore, this correlation seemed particular strong from 2008Q2 onwards.

Table 4 – Biased adjusted LM test statistic for different models

Model / Equation	1	2	3	4	5	6	7	8	9	10
No Factors	1.61	-.12	.75	.49	.10	1.6	-.04	1.26	.05	1.24
Factors in EQ 1 and EQ6	.76	-.25	2.1	.66	.56	3.1	2.06	4.68	2.74	1.71
Factors in EQ1, EQ3, EQ6, EQ7, EQ8, EQ9,EQ10	.67	-.27	.02	.67	.56	2.8	.257	.845	-.114	.0054
Factors in EQ1, EQ3, EQ6, EQ7, EQ8, EQ9,EQ10 and additional factor in EQ6	.92	-.28	.10	1.0	.7	1.5	.194	.39	.21	.19

4.3. Identification

There is now a long economic literature on identifying macroeconomic shocks using long-run restrictions, highlighting both the advantages and caveats of this approach. Blanchard and Quah (1989) first introduced this method for systems of equations, while Watson and Shapiro (1988) proposed a single equation approach to identify shocks via long-run restrictions in VAR systems. Most recent work examines the United States and follows Gali (1999) in identifying permanent technology shocks using a VAR specification which typically consists of growth in both labour productivity and hours worked. It has been pointed out that long-run restrictions may provide only weak identification information (Faust and Leeper, 1997) and that existing samples may be too short to implement such restrictions (Erceg et al, 2005). Similarly, different types of technology shocks may have different impacts on *per capita* hours worked; Fisher (2006) proposes the investment-specific technology shock, as a separate important business cycle determinant of hours.[11]

Some of these criticisms can be naturally addressed within our framework. In particular, we are interested in identifying sector level productivity shocks, from five different economic sectors. Given the potential for spillovers across sectors, what previous studies label as an 'investment-specific' technology shock, could just be a shock originating in a particular sector. Even if this is not the case, the comprehensive robustness examination exercise presented in Canova, Lopez-Salido and Michelacci (2010) finds that in the US, Gali-type neutral technology shocks explain a much greater fraction of output growth, the variable we ultimately seek to explain, than 'investment-specific' technology shocks. We therefore follow the approach in Gali (1999) and attempt to identify what previous work labels 'neutral' technology shocks at the sector level.

The challenge for structural VAR models is to disentangle orthogonal, structural economic shocks, $\boldsymbol{\varepsilon_t}$, from the correlated reduced form shocks $\boldsymbol{u_t}$. This is typically achieved with the help of a matrix $\boldsymbol{C_0}$, such that $\boldsymbol{C_0 u_t} = \boldsymbol{\varepsilon_t}$. As discussed below, we recover $\boldsymbol{C_0}$ either with long-run only or a combination of long-run and short-run, restrictions. Estimates of **(10)** can be used to obtain the reduced-form vector moving average:

$$\boldsymbol{Y_{c,t}} = (\boldsymbol{I} - \textstyle\sum_{l=1}^{L} \boldsymbol{A_{c,l}})^{-1}\boldsymbol{u_t} + (\boldsymbol{I} - \textstyle\sum_{l=1}^{L} \boldsymbol{A_{c,l}})^{-1}\boldsymbol{D_c F_t} = \boldsymbol{R}(\boldsymbol{L})\boldsymbol{u_t} + \boldsymbol{R}(\boldsymbol{L})\boldsymbol{D_c F_t} \qquad \textbf{(11)}$$

From **(11)**, it is easy to see that $\boldsymbol{R}(\boldsymbol{L})\boldsymbol{C_0^{-1}} = \boldsymbol{C}(\boldsymbol{L})$, where $\boldsymbol{R}(\boldsymbol{L})$ ($\boldsymbol{C}(\boldsymbol{L})$) is the implied matrix of long-run reduced form (structural) multipliers. By imposing restrictions on $\boldsymbol{C}(\boldsymbol{L})$, it is therefore possible to use long-run restrictions to recover $\boldsymbol{C_0}$ and therefore identify the structural shocks. To understand our proposed identification restrictions better, it is useful to divide the matrix of structural coefficients $\boldsymbol{C_0}$ and long-run multipliers $\boldsymbol{C}(\boldsymbol{L})$ into four quadrants:

[11] A related point, made by Fernald (2007) is that breaks hours worked data may be responsible for some of the puzzles typically investigated in these studies that examine the United States. It is unclear however, to which extent this critique applies to our work, which takes a cross-country approach.

$$C_0 = \begin{bmatrix} C_{0,11} & C_{0,12} \\ C_{0,21} & C_{0,22} \end{bmatrix} \qquad C(L) = \begin{bmatrix} C(L)_{11} & C(L)_{12} \\ C(L)_{21} & C(L)_{22} \end{bmatrix} \tag{12}$$

Identification with long-run restrictions only requires imposing $\frac{N(N-1)}{2}$ restrictions on $C(L)$, leaving C_0 unrestricted. Our main identification assumption, that only permanent productivity shocks can affect labour productivity in the long run, meaning that $C(L)_{12}$ is a matrix of zeros, provides $\frac{N^2}{4}$ restrictions. The remaining $\frac{N(N-2)}{4}$ restrictions can be imposed by requiring both $C(L)_{11}$ and $C(L)_{22}$ to be lower triangular matrices. Conveniently, these restrictions can be imposed by taking the lower triangular choleski decomposition of $C(L)C(L)' \equiv R(L)\Sigma_c R(L)'$, which can be calculated from observable reduced form quantities. This will be referred to as identification scheme I for the rest of the paper.

The above is of course not the only possible identification scheme. To ensure that our results are robust to this particular choice, we also experiment with a second identification scheme, following the IV approach to structural VAR identification of Robertson and Pagan (1998). With this approach, which allows for the combination of short and long-run restrictions, we assume that $C_{0,11}$ has ones on the diagonal and that $C(L)_{11}$ is a diagonal matrix. This allows us to interpret the shocks as sector-specific permanent productivity shocks. More importantly, this assumption implies that, since $C(L)C_0 = R(L)$, we can recover each row x of $C_{0,11}$ and $C_{0,12}$ through $C_{0,11}(x, N/2) = R(L)_{11}(x, N/2)/R(L)_{11}(x,x)$ and $C_{0,12}(x, N/2) = R(L)_{12}(x, N/2)/R(L)_{11}(x,x)$ (See Fry and Pagan, 2005). To identify additionally the temporary demand shocks, it is necessary to impose some short-run restrictions. We require $C_{0,21}$ to be a diagonal matrix of ones, which means that permanent productivity shocks spill over to other sectors only with a lag. With $C_{0,11}$ and $C_{0,12}$ in hand, it is possible to obtain the first $\frac{N}{2}$ elements of ε_t. These will be by definition uncorrelated with the remaining elements of ε_t, but correlated with u_t. The coefficients of $C_{0,22}$ are therefore estimated with the first $\frac{N}{2}$ elements of ε_t as instruments for the corresponding elements of u_t. This identification scheme is referred to as identification scheme II throughout.

Equipped with $\boldsymbol{C_0}$, we can decompose each time series in the VAR into the contribution from each structural shock. That is, we can determine to what extent labour productivity growth, hours growth and output growth (defined as the sum of the former and the latter) are driven by the structural shocks or the common factors. To see how this done in practice, consider that based on **(10)**, abstracting from the lag structure for simplicity, the data at any point in time can be expressed as:

$$\boldsymbol{Y_{c,t} = A_c^{t-t0} Y_{c,t0} + \sum_{i=0}^{t-t0} A_c^i\, C_0 \varepsilon_{t-i} + \sum_{i=0}^{t-t0} A_c^i\, D_c F_{t-i}} \qquad \textbf{(13)}$$

where $\boldsymbol{t0}$ is the base period. This expression can also be used to understand the contributions of the identified shocks to each series in $\boldsymbol{Y_{c,t}}$. For instance, to calculate the contribution of a given structural shock, set that shock to zero, generate the implied path for all the time series and then subtract it from the actual data. This is how we compute the contribution of each identified shock. The contributions of the global factors are then obtained as the difference between the sum of the contributions from all of the identified shocks and the corresponding data series.[12] For this type of exercise, the choice of baseline period (i.e. the time after which shocks arrive) is particularly important, as people are assumed to know everything up until that time. Since we are interested in the impact of permanent labour productivity and transitory demand shocks following the onset of the 'Great Recession', we choose 2008Q2 as our baseline period.

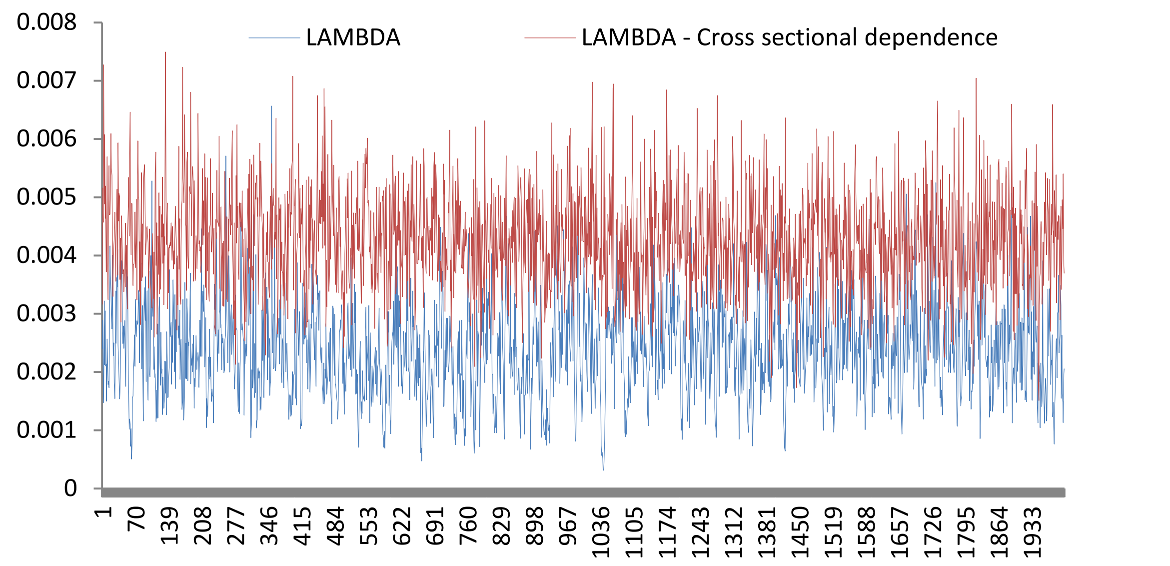
4.4. Results

In this section we first show the reduced form results from our model and then proceed to a discussion of the identified VAR results and their policy implications. One interesting statistic is the distribution of $\boldsymbol{\lambda}$, the degree of heterogeneity in the data. Figure 1 shows these statistics for the model estimated with and without controlling for cross-sectional dependence. The mean of draws of $\boldsymbol{\lambda}$ of the former model, at .0024, is roughly half as large as that of the latter at .0043. This is an intuitive result: ignoring the presence of a common factor in the residuals may make data appear more similar across countries

[12] In models without a common factor, temporary demand shocks are calculated as the difference between the path implied by the permanent productivity shocks and the data.

than they actually are, resulting in an artificially smaller degree of dynamic heterogeneity. In terms of magnitudes, it is useful to compare these numbers to the degree of shrinkage that is typically imposed in the Litterman (1986) prior. The former is typically set to between .1 - .2. The comparative statistic from our model is $\sqrt{\lambda}$, whose mean is between .048 and .065. This suggests that there is a substantial degree of heterogeneity in our data and pooling could thus lead to substantial bias.

Figure 1: Retained draws of λ with and without cross-sectional dependence



In what follows, we show the results for the historical decompositions. The models' estimated contributions, as well as the data, are weighted by the relative sizes of the sectors concerned. Results for each individual sector can be found in appendix A. Figure 2 shows historical decompositions for output growth from the model estimated without cross-sectional dependence.[13] For the UK, the recent weakness in output growth appears to be mostly due weakness in demand as opposed to a sequence of permanent labour productivity shocks. The 'Great Recession' in Canada is mostly driven by weak demand, while labour productivity explains a substantial fraction of the evolution of

[13] Note that output growth is by definition the difference between labour productivity and hours growth. Though our model is the estimated on the former two quantities, historical decompositions are available for all three variables.

output growth in 2008 and 2009 in Germany and the UK. Figure 3 repeats this exercise in the model where cross-sectional dependence is accounted for. The previous conclusions are now clearly overturned. Both permanent labour productivity and temporary demand shocks now contribute roughly equal amounts to recent (2010 and 2011) weak output growth in the UK. Furthermore, the common factors now appear to be the most important determinant of output growth during the Global Financial Crisis in both Germany and Canada. This clearly illustrates that ignoring the issue of cross-sectional dependence can lead to substantial bias, resulting in incorrect research conclusions. Figures 4-5 repeat this exercise with identification scheme II. The results are very similar, suggesting that they are independent of the identification scheme.

Figure 2: Historical decomposition for output growth from model estimated on sector data – Identification scheme I

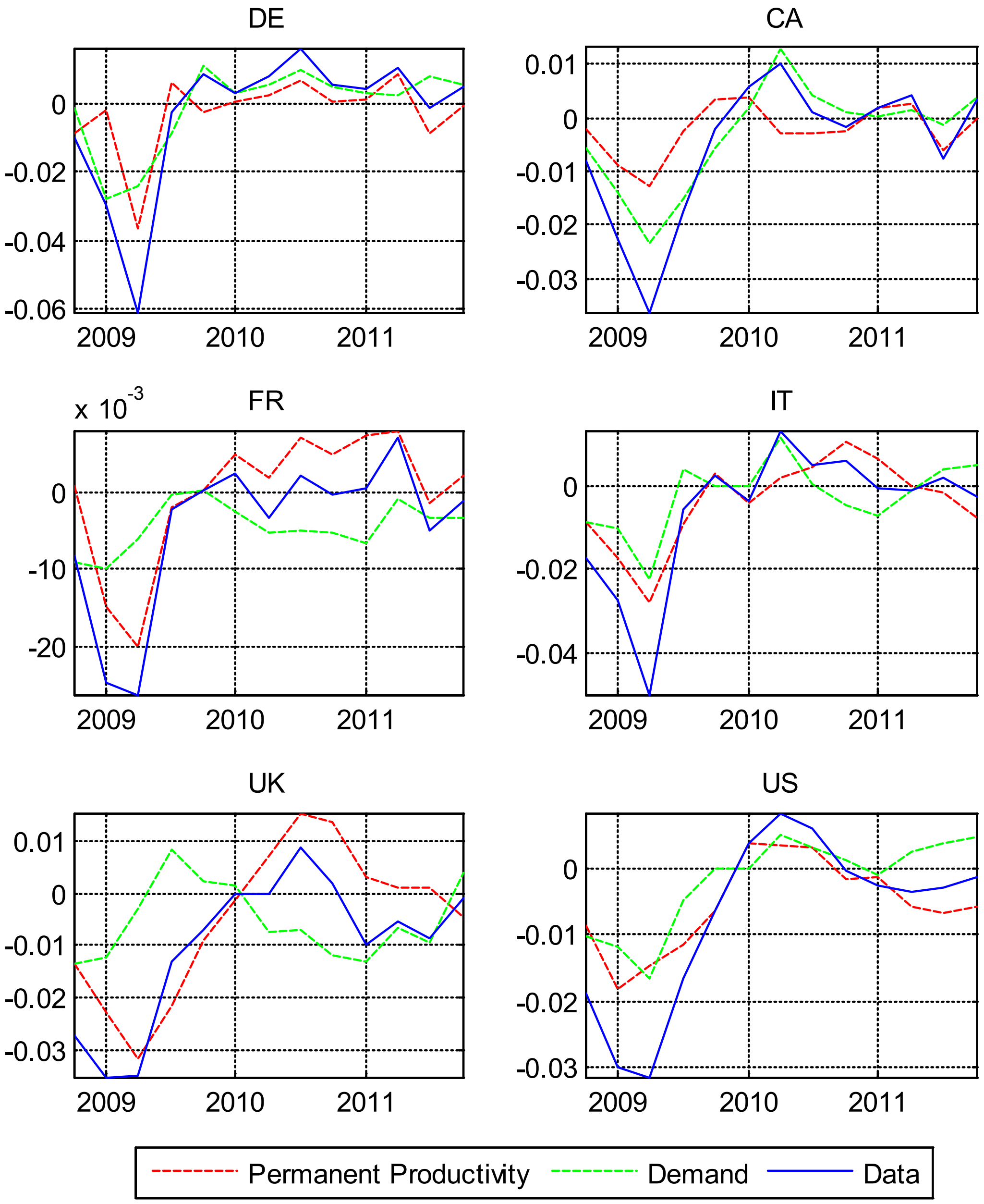


Figure 3: Historical decomposition for output growth from model estimated on sector data allowing for cross-sectional dependence – Identification scheme I

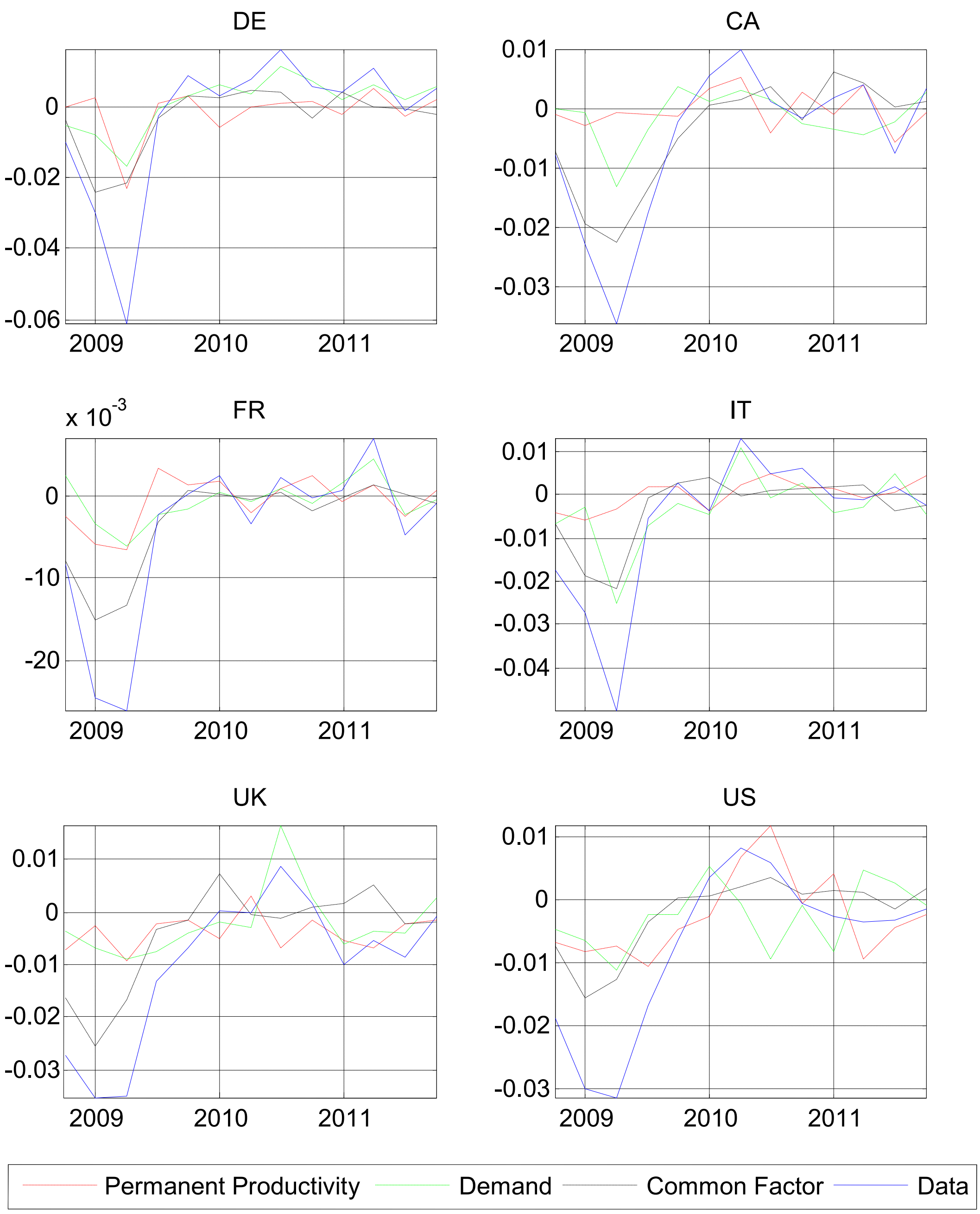


Figure 4: Historical decomposition for output growth from model estimated on sector data – Identification scheme (II)

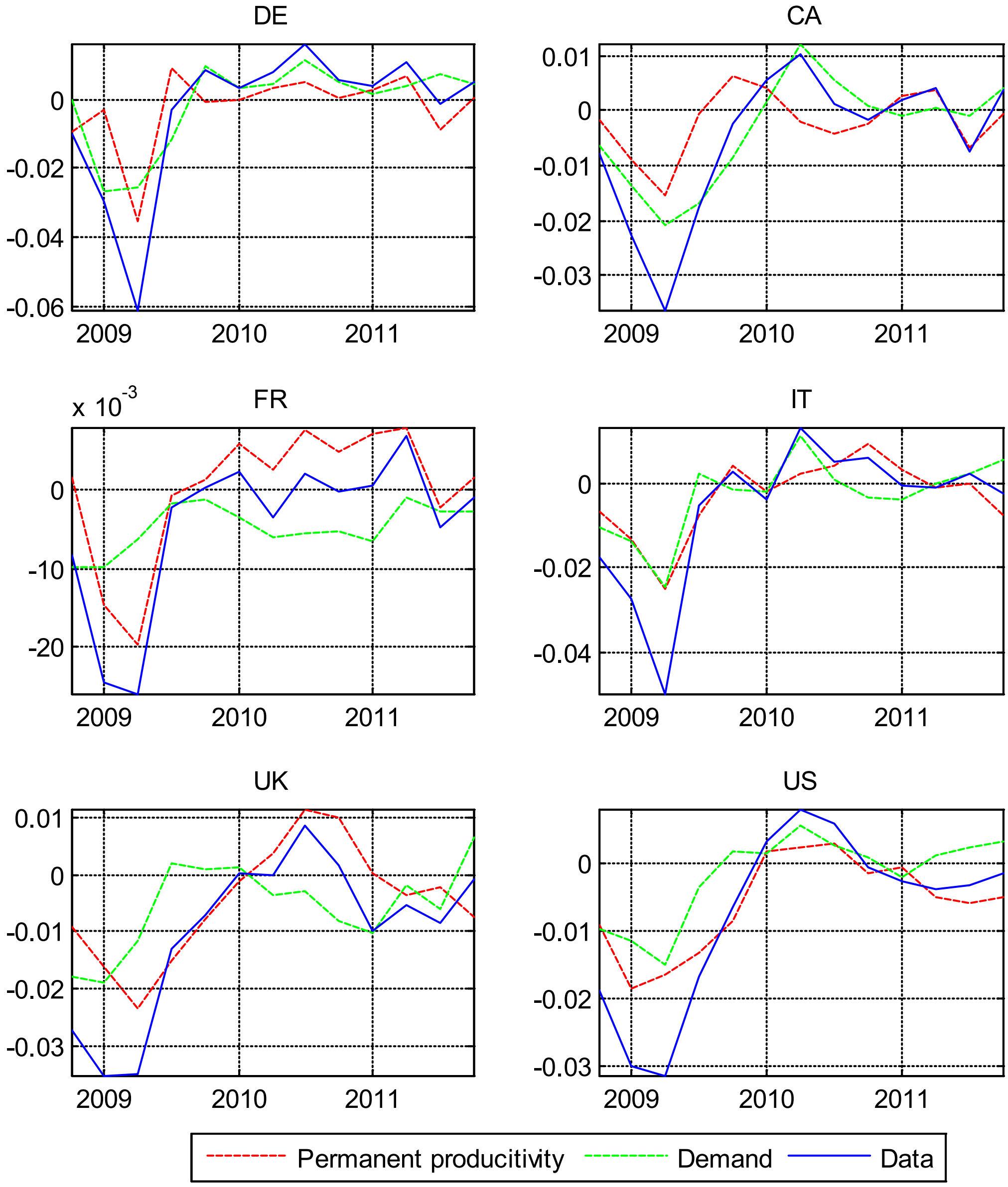
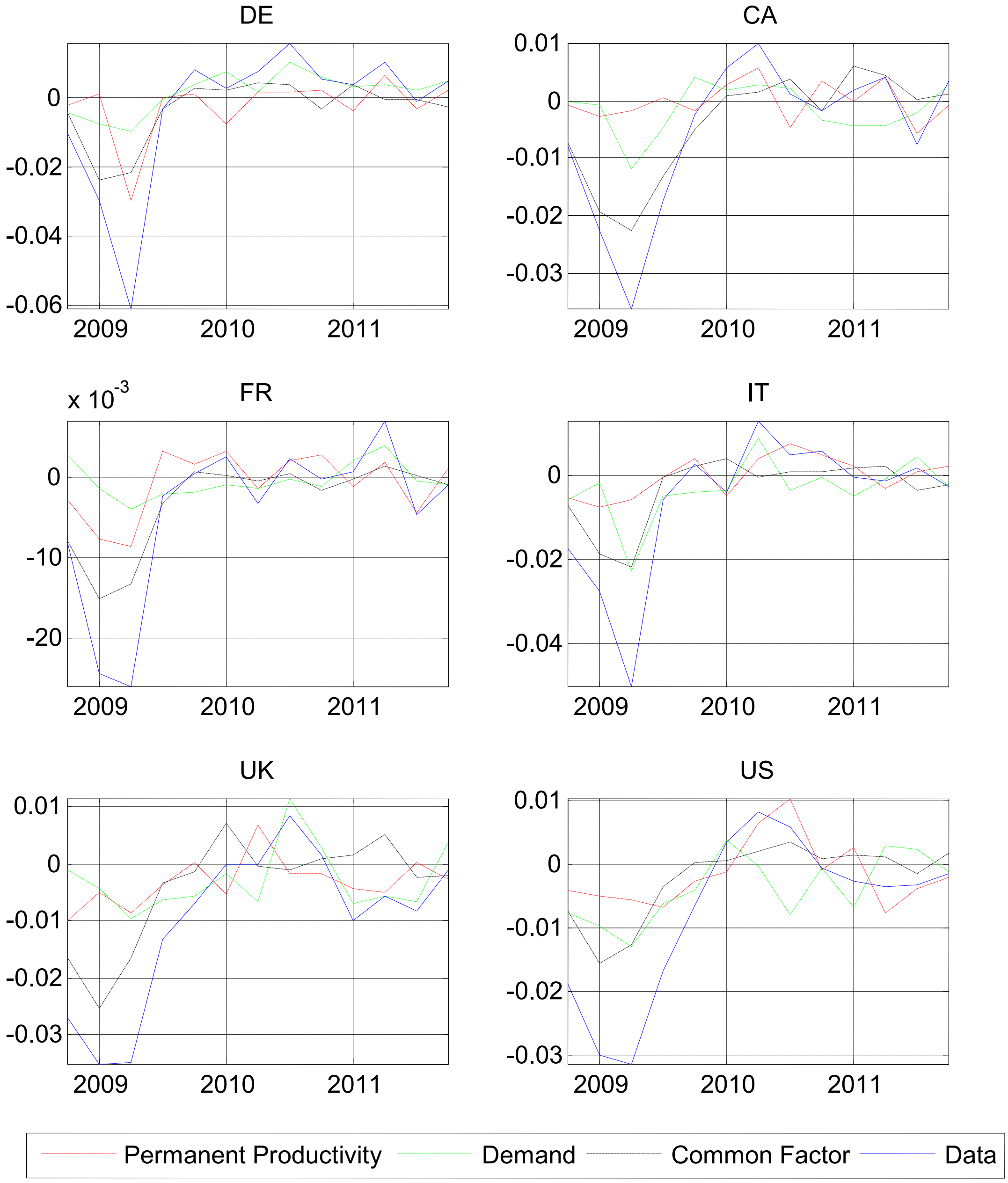


Figure 5: Historical decomposition for output growth from model estimated on sector data allowing for cross-sectional dependence – Identification scheme (II)



5. Conclusion

Currently available dynamic panel estimators that allow for both dynamic heterogeneity and cross-sectional dependence are not suitable for inference in short panels. Those estimators that address cross-sectional dependence and are designed specifically for short panels, as in Bai (2009) and Saradifis (2010), do not allow for dynamic heterogeneity. While those that allow for cross-sectional dependence and dynamic heterogeneity, as in Phillips and Sul (2003) and Pesaran (2006), are not suitable for inference in short panels. We propose a Bayesian estimator, which allows both cross-sectional dependence and dynamic heterogeneity and is feasible in short panels, to fill this gap. We use Monte Carlo simulations to compare the small sample properties of our estimators to those proposed in previous work and illustrate our technique by estimating a panel VAR on sectoral data on growth in labour productivity and hours worked for the G6.

Our Monte Carlo simulations show that our proposed estimator produces estimates substantially less biased than either the common correlated effects pooled or common correlated effects mean group estimator, when the panel is short. In our application, the estimated panel VAR contains ten endogenous variables and two lags with seventy-six time series observations for each country, leaving few time series observations per estimated coefficient. Our analysis suggests that it is important to take both dynamic heterogeneity and cross-sectional dependence in our application into account. Once estimated, we identify permanent productivity and temporary demand shocks using long-run restrictions, and investigate their contribution to recent output growth outturns in the G6 by means of a historical decomposition. The results show that taking cross-sectional dependence into account is important. For instance, ignoring cross-sectional dependence would lead a researcher to believe mistakenly that recent weak UK output growth was entirely due to demand rather than permanent productivity shocks.

Recent applied econometric work has relied on panel VARs for inference about important areas of economic policy, such as for instance the transmission of external

(Broda, 2004; Radatz, 2007), monetary (Goodhart and Hoffman, 2008) and fiscal policy (Llzzetzki, Mendoza and Vegh, 2011; Corsetti, Meier and Mueller, 2012) shocks to the real economy. While most of these studies are based on short panels, none of them make a serious effort to address both cross-sectional dependence and dynamic heterogeneity. This is not surprising, as previous estimators that address both issues are not suitable for inference in short panels. Our proposed estimator fills this gap and we show that ignorance of cross-sectional dependence in our estimated panel VAR model leads to serious econometric bias. A useful future exercise would be therefore to re-examine past results with our new estimator.

References

Bai, J (2009), 'Likelihood approach to small T dynamic panel data models with interactive effects', *Columbia University Mimeo.*

Binder, M, Hsiao, C and Pesaran, H (2005), 'Estimation and Inference in Short Panel Vector Autoregressions with Unit Roots and Cointegration', *Econometric Theory*, 21, pp. 795-837.

Blanchard, O J and Quah, D (1989), 'The Dynamic Effects of Aggregate Demand and Supply Disturbances', *American Economic Review,* Vol. 79(4), pages 655-73.

Breusch, T S and Pagan, A R (1980), 'The Lagrange Multiplier Test and Its Applications to Model Specification in Econometrics', *Review of Economic Studies,* Vol. 47(1), pages 239-53.

Broda, C (2004), 'Terms of trade and exchange rate regimes in developing countries', *Journal of International Economics,* Vol. 63(1), pages 31-58.

Canova, F (2005), 'The transmission of US shocks to Latin America', *Journal of Applied Econometrics*, Vol. 20(2), pages 229-251.

Canova, F, Lopez-Salido, D and Michelacci, C (2010), 'The effects of technology shocks on hours and output: a robustness analysis', *Journal of Applied Econometrics,* Vol. 25(5), pages 755-773.

Ciccarelli, M and Rebucci, A (2002), 'The Transmission Mechanism of European Monetary Policy: Is There Heterogeneity? Is it Changing over Time?', *IMF WP 02/54.*

Coakley, J, Fuertes, AM and Smith, R (2002), 'A Principal Components Approach to Cross-Section Dependence in Panels', *paper presented at10th International Conference on Panel Data*, Berlin, Germany, B5-3.

Corsetti, G, Meier, A and Müller, G (2012), 'What Determines Government Spending Multipliers?', *CEPR Discussion Papers 9010.*

Erceg, C J, Guerrieri, L and Gust, C (2005), 'Can Long-Run Restrictions Identify Technology Shocks?', *Journal of the European Economic Association*, Vol. 3(6), pages 1237-1278.

Faust, J and Leeper, E M (1997), 'When Do Long-Run Identifying Restrictions Give Reliable Result?', *Journal of Business & Economic Statistics,* Vol. 15(3), pages 345-53.

Fernald, J G (2007), 'Trend breaks, long-run restrictions, and contractionary technology improvements', *Journal of Monetary Economics*, Vol. 54(8), pages 2467-2485.

Fisher, J D M (2006), 'The Dynamic Effects of Neutral and Investment-Specific Technology Shocks', *Journal of Political Economy*, Vol. 114(3), pages 413-451.

Frees, E W (1995), 'Assessing cross-sectional correlation in panel data', *Journal of Econometrics,* Vol. 69(2), pages 393-414.

Fry, R and Pagan, A (2005), 'Some Issues In Using Vars For Macroeconometric Research', *CAMA Working Papers 2005-19.*

Gali, J (1999), 'Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?', *American Economic Review*, Vol. 89(1), pages 249-271.

Gelman, A, Carlin, J, Stern, H and Rubin, D (2003), *Bayesian Data Analysis (*2nd edn*),* Chapman & Hall/CRC: London.

Gilhooly, B, Weale, M and Wieladek, T (2012), 'Disaggregating the international business cycle', *External MPC Unit Discussion Paper No. 37.*

Goodhart, C and Hofmann, B (2008), 'House prices, money, credit, and the macroeconomy', *Oxford Review of Economic Policy*, Vol. 24(1), pages 206-209.

Holtz-Eakin, D, Newey, W and Rosen, H S (1988), 'Estimating Vector Autoregressions with Panel Data', *Econometrica*, vol. 56(6), pages 1371-95, November.

Hsiao, C, Pesaran, H and Tahmiscioglu, K (1999), 'Bayes Estimation of Short-Run Coefficients in Dynamic Panel Data Models', *Analysis of Panels and Limited Dependent Variables: A Volume in Honour of G S Maddala,* CU Press, Chapter 11, pages 268-296.

Jarocinski, M (2010), 'Responses to monetary policy shocks in the east and the west of Europe: a comparison', *Journal of Applied Econometrics,* Vol. 25(5), pages 833-868.

Lindley, D and Smith, A (1972), 'Bayes estimates for the linear model'*, Journal of the Royal Statistical Society,* B 34 (1): 1-41.

Litterman, R B (1986), 'Forecasting with Bayesian Vector Autoregressions - Five Years of Experience', *Journal of Business & Economic Statistics,* Vol. 4(1), pages 25-38.

Love, I and Zicchino, L (2006), 'Financial development and dynamic investmenr: Evidence from panel VAR', *Quarterly Review of Economics,* Vol. 46(2), pages 190-210.

Lzzetzki, E, Mendoza, E G and Vegh, C A (2010), 'How Big (Small?) are Fiscal Multipliers?', *NBER Working Paper 16479.*

Nickell, S J (1981), 'Biases in Dynamic Models with Fixed Effects', *Econometrica,* Vol. 49(6), pages 1417-26.

Pagan, A R and Robertson, J C (1998), 'Structural Models of The Liquidity Effect', *The Review of Economics and Statistics*, Vol. 80(2), pages 202-217.

Pesaran, H, and Smith, R (1995), 'Estimating Long-Run Relationships from Dynamic Heterogeneous Panels', *Journal of Econometrics,* Vol. 68, pages 79-113.

Pesaran, H (2004), 'General Diagnostic tests for Cross Section Dependence in Panels,' *Cambridge Working Papers in Economics 0435.*

Pesaran, H (2006), 'Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure', *Econometrica,* Vol. 74(4), pages 967-1012.

Pesaran, H, Ullah, A and Yamagata, T (2008), 'A bias-adjusted LM test of error cross-section independence', *Journal of Econometrics,* Vol. 11(1), pages 105-127.

Phillips, P C B and Sul, D (2003), 'Dynamic panel estimation and homogeneity testing under cross section dependence', *Econometrics Journal,* Vol. 6(1), pages 217-259.

Raddatz, C (2007), 'Are external shocks responsible for the instability of output in low-income countries?', *Journal of Development Economics,* Vol. 84(1), pages 155-187.

Robertson, D and Symons, J (2007), 'Maximum likelihood factor analysis with rank-deficient sample covariance matrices', *Journal of Multivariate Analysis,* Vol. 98(4), pages 813-828.

Sarafidis, V (2009), 'GMM Estimation of short dynamic panel data models with error cross-sectional dependence', *MPRA Paper 25176.*

Sarafidis, V and Robertson, D (2009), 'On the impact of error cross-sectional dependence in short dynamic panel estimation', *Econometrics Journal,* Vol. 12(1), pages 62-81.

Sarafidis, V and Wansbeek, T (2010), 'Cross-sectional dependence in panel data analysis', Econometric Reviews, Taylor and Francis Journals, vol. 31(5), pages 483-531, September.

Stephan, F (1934), 'Sampling Errors and Interpretation of Social Data Ordered in Time and Space', Journal of the American Statistical Association, Vol. 29, pages 165-166.

Vidangos, I (2009), 'Fluctuations in Individual Labor Income: A Panel VAR Analysis', *Board of Governors (U.S.) Finance and Economics Discussion Series 2009-09.*

Zellner, A (1962), 'An efficient method of estimating seemingly unrelated regression equations and tests for aggregation bias', *Journal of the American Statistical Association* Vol. 57, pages 348–368.

Appendix A

Figure A2: Historical decomposition from model estimated on sector data – Identification scheme I

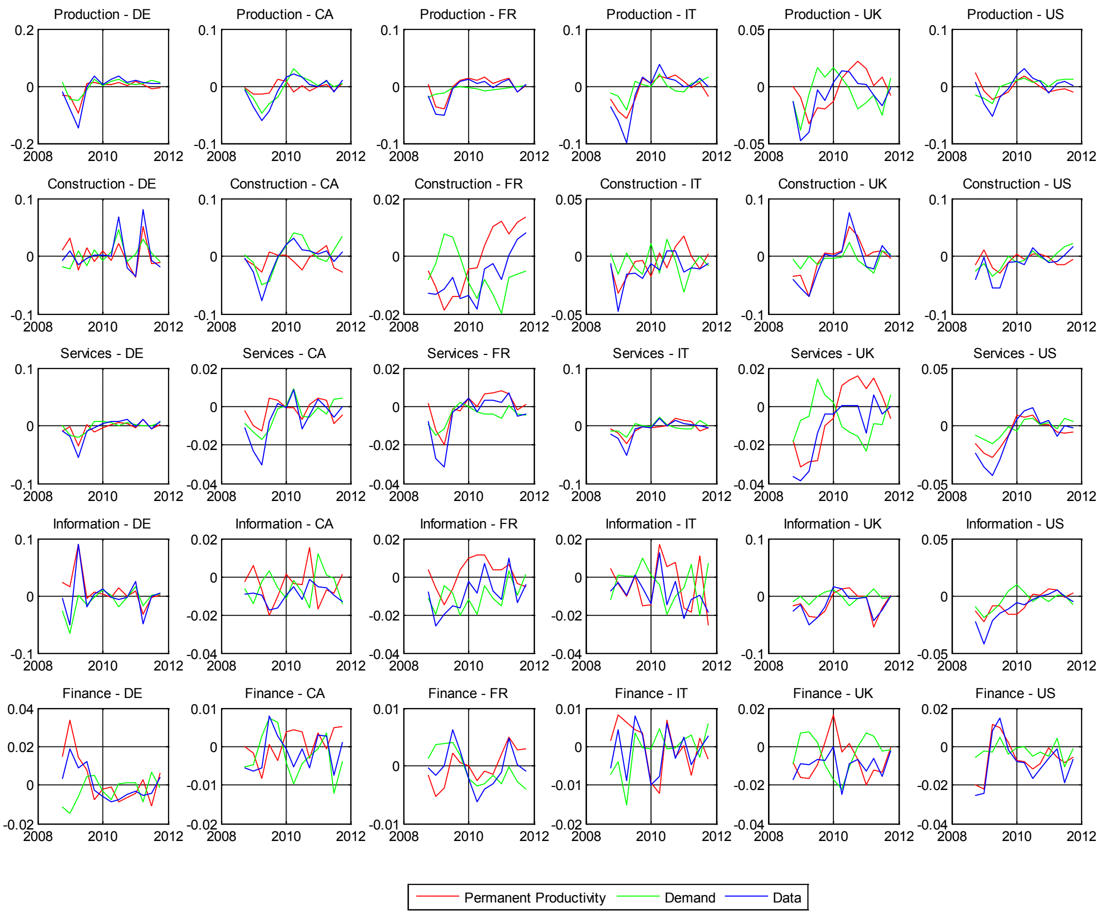


Figure A3: Historical decomposition from model estimated on sector data allowing for cross-sectional dependence – Identification scheme I

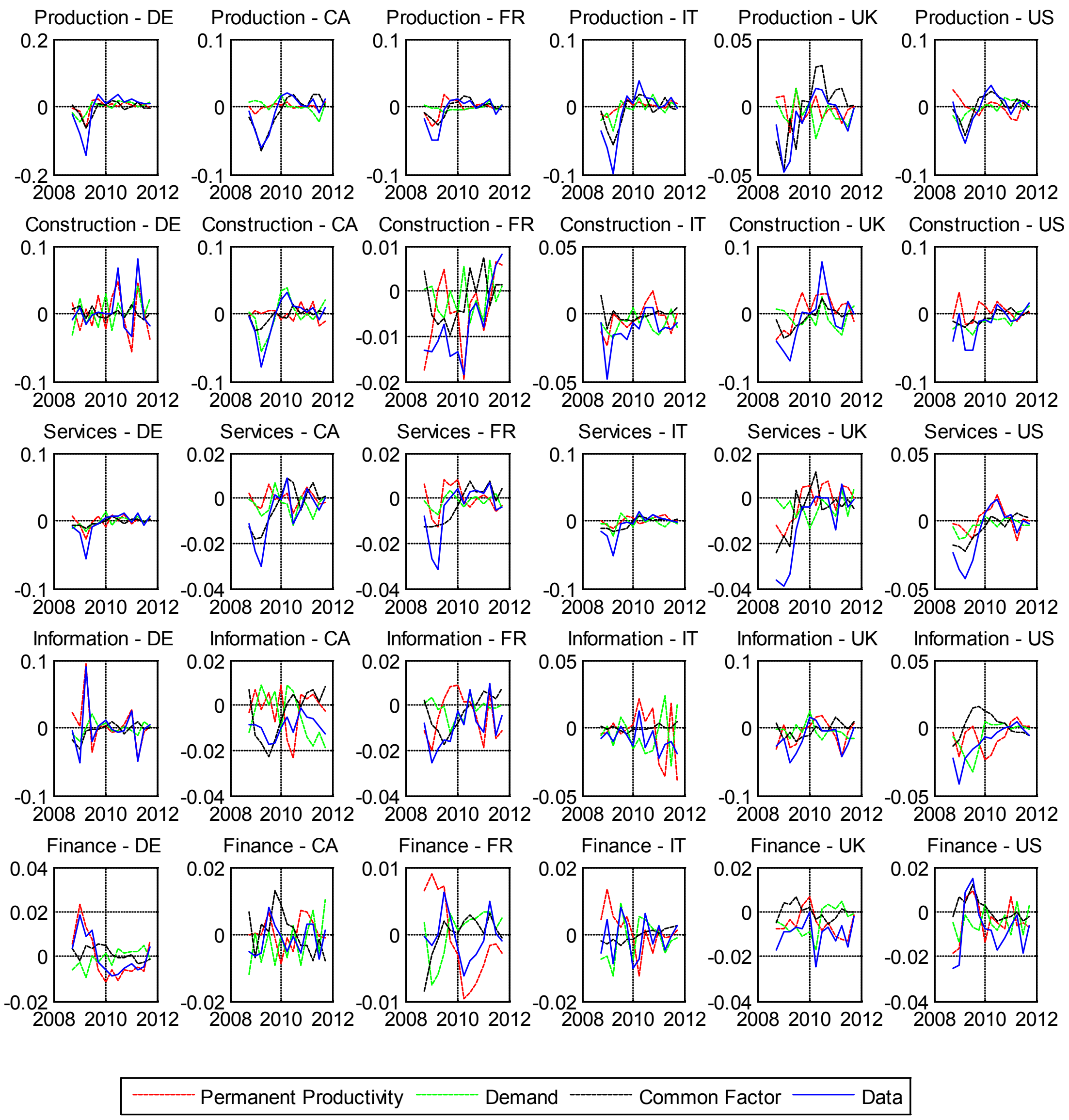


Figure A4: Historical decomposition from model estimated on sector data – Identification scheme (II)

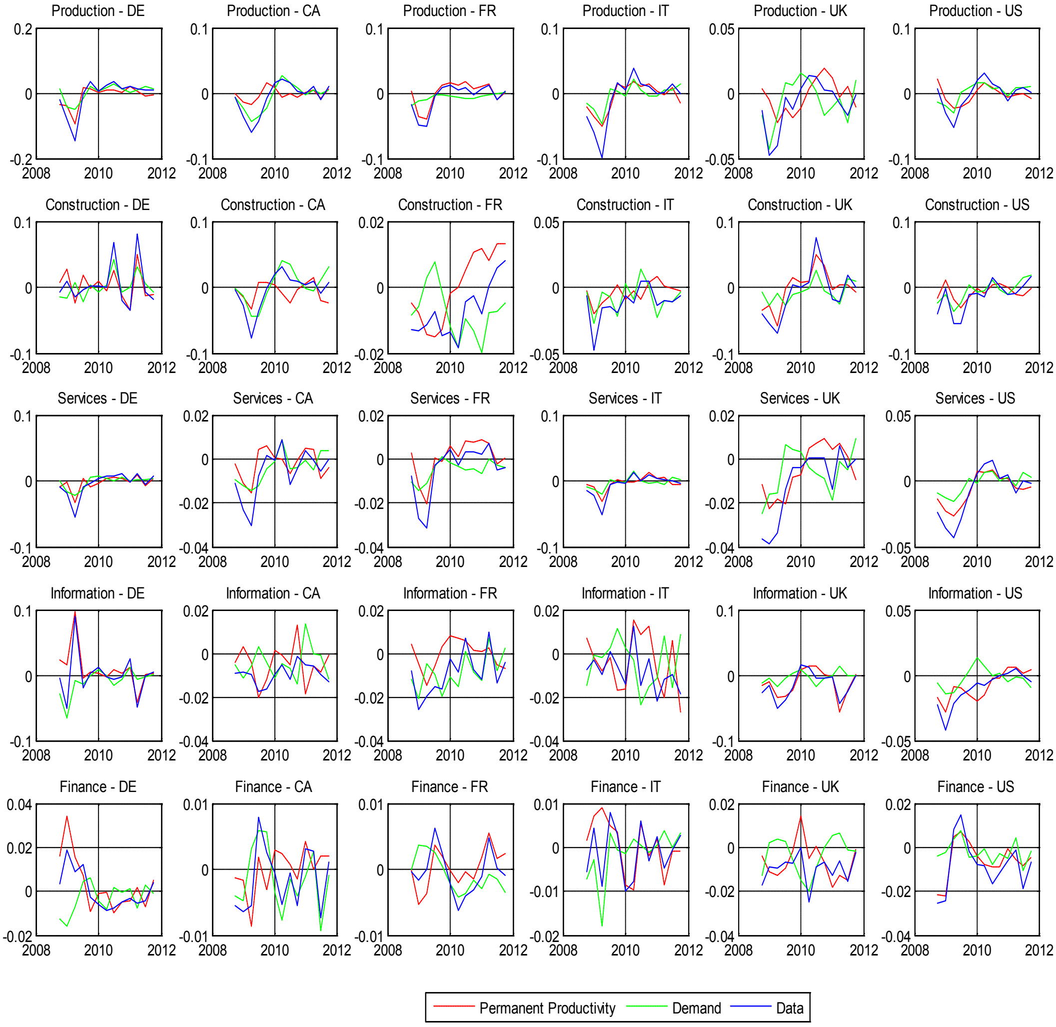


Figure A5: Historical decomposition from model estimated on sector data allowing for cross-sectional dependence – Identification scheme (II)