Making Fiscal Adjustments Using Event Probability Forecasts in OECD Countries

Kevin Lee,[†] Kian Ong[††] and Kalvinder Shields[††*]

December 2019

Abstract

This paper describes an approach to making fiscal policy decisions based on probabilistic statements on the likely occurrence of events as specified in a rules-based framework for making fiscal adjustments. The event probability forecasts are obtained from a simple time series econometric model of the key variables influencing debt dynamics (interest rates, output and debt itself). The approach is applied to data for ten developed countries for 1956-2016 and the analysis demonstrates the importance of accommodating international linkages in forecasting, noting that failure to do so would have led to excessive fiscal cut-backs and austerity in recent years.

Keywords: Fiscal adjustment, Debt sustainability, Cross-country spillovers, Event probability forecasting.

JEL Classification: C22, C53, E63, H30, H68

---

[*] †University of Nottingham, UK; ††University of Melbourne, Australia. Paper prepared for the Special Issue of the *Economic Record* in honour of Mardi Dungey. This research was funded by the Australian Research Council under the Discovery-Projects scheme grant number DP140103029. Corresponding author: Kalvinder K. Shields, Department of Economics, University of Melbourne, Victoria, 3010, Australia. E-mail: k.shields@unimelb.edu.au, tel: 00 613-83443549, fax: 00 613-83446899. Version dated December 2019.

1 Introduction

The Global Financial Crisis (GFC) and the subsequent European Sovereign Debt crisis highlighted the importance of cross-country interdependencies in financial markets, the interplay between countries' monetary and fiscal policies and their crucial roles in determining macroeconomic performance. Applied macro-econometric analysis has been central to understanding these phenomena, and to formulating appropriate policies, and Mardi Dungey's work in the area has been extremely important. In particular, work on financial contagion - exemplified by Dungey and Martin (2007), and Dungey et al (2005, 2018a, 2018b) - has been key to recognising the importance and complexity of international linkages in financial markets. And innovations in the field of time series econometrics, including careful work on the identification of policy shocks in the context of a VAR models – exemplified by Dungey and Pagan's (2000, 2009) VAR models of the Australian economy, Dungey and Fry's (2009) study of the identification of fiscal and monetary policy shocks and Dungey and Osborn's (2014) modelling of cross-country interactions – have helped identify and understand the impact and ultimate multiplier effects of policy shocks as they are propagated across national borders and over time.

This paper takes up these themes and considers the role of international linkages in time series models used to determine a country's fiscal policy stance. Specifically, the paper explains how estimated macroeconomic models can be used to generate event probability forecasts to inform fiscal policy decisions. The approach has a number of useful features. *First*, the use of a macroeconomic model means that the decisions take proper account of the future dynamics of the key fiscal and macroeconomic variables (in contrast to the scenario analysis underlying many informal indicators used to motivate fiscal adjustments). *Second*, the estimation of event probability forecasts means that the decisions can be based on a nuanced description of the objectives of the fiscal adjustments and can take into account the uncertainties surrounding the future outcomes. And *third*, having grounded the analysis in a specific decision-making context, the selection of an appropriate model can be based on an *economic evaluation* of its forecast performance; i.e. judging whether a model's forecasts are useful in the specific decision-making context rather than focusing on standard statistical measures of forecast performance. This forms the criterion with which to judge the contribution of international linkages in making fiscal adjustments therefore, arguing that these linkages should be taken into account if they improve decision-making.

There has been considerable recent interest in modelling in the context of fiscal policy decisions. The GFC resulted in substantial increases in the levels of public debt and fiscal austerity in many countries, and a number of national fiscal councils and supranational bodies were established to provide independent evaluation and assessment of policy makers' announced fiscal plans.[1] The assessments are concerned with a country's ability to undertake fiscal policy without compromising potential lenders' willingness to hold the country's sovereign debt,[2] and typically focus on the likely level of a country's debt-to-GDP over the near future and its ability to service its debt as reflected in the real interest rate relative to the output growth rate. Judgements on the appropriate size and direction of fiscal adjustments are typically based on scenario analyses and forecasts of various degrees of sophistication. At one extreme, a variety of indicators have been proposed for use in making fiscal adjustments, expressed in terms of the feasibility of achieving a specified debt target, for example, or in terms of the adjustment required to achieve a specified debt target say.[3] The indicators provide a useful signal on the country's ability to undertake fiscal policy but often focus on just one aspect of the decision-making context and provide only imprecise advice on the direction and size of policy adjustments. At the other extreme, there are formal rules-based frameworks in which the forecasts of a range of macroeconomic variables are used to motivate relatively detailed fiscal adjustments. The framework rules can be very complicated, based on the conjunction of a set of inter-related events that can involve a large number of macroeconomic variables. The benefit of this approach, relative to the use of simple indicators, depends on the extent to which the forecasting model matches the complexity of the decision-rules in terms of

---

[1] See Baldwin and Giavazzi (2015) and Debrun and Kinda (2014) for discussion.

[2] See the discussion of Kose et al. (2017).

[3] See, for example, Buiter (1985), Blanchard et al. (1990), Horne (1991), Wyplosz (2007), Cottarelli and Vinals (2007), Polito and Wickens (2012), the European Commission (2014) and the Office of Budget Responsibility (2015) for discussion of the use of fiscal indicators.

variable-coverage and its ability to capture the underlying events.

The approach described in the paper is based on a simple time series econometric model of the key variables influencing debt dynamics (interest rates, output and debt itself) and probabilistic statements of the likely occurrence of pre-defined future events. Despite the simplicity of the time series model, the events can be complicated and of the type found in rules-based frameworks. The approach has the advantage that it (i) acknowledges that the future values of the key variables are time-varying and determined jointly and endogenously; (ii) take into account the uncertainties surrounding the future values of the key variables; and (iii) can accommodate any number of the events relevant to fiscal adjustment in a single coherent framework. A potentially important element in the construction of the countries' event probability forecasts is the role played by cross-country interactions because there are potentially many common, global factors - including those underlying financial contagion - driving the key variables influencing debt dynamics in each country and many international feedbacks that propagate the effects of changes experienced in one country over time and across borders.[4] For this reason, we estimate our probability forecasts using both nation-specific vector-autoregressive (VAR) models and Global Vector-Autoregressive (GVAR) and Global Vector Error-Correction (GVECM) models which can accommodate the effects of international linkages on forecasts of the key variables and hence on the forecast probability of the events motivating fiscal adjustments. We then judge the usefulness of the extra complexity involved in the global models according to standard statistical criteria and according to economic criteria derived from the impact on the adjustment decisions made.[5]

We illustrate our approach using data for ten OECD countries over the period 1956-2016 and describe an analysis of fiscal adjustments from the perspective of a country adopting the formal rules-based framework set out in the European Union's Stability and Growth Pact (SGP). The SGP aims at establishing fiscal sustainability with reference to two nominal objectives; namely, limiting the deficit ratio to be no more than

---

[4] See, for example, Popov and Van Horen (2015) or Becker and Ivashina (2018).

[5] See, for example, Granger and Pesaran (2000) and Granger and Machina (2006) who emphasise the importance of judging the economic value of a model's forecast, concentrating on the usefulness of the models in a specific decision-making context, rather than on its statistical performance.

3% of GDP and the debt ratio to be no more than 60% of GDP. Countries define their medium-term budgetary objectives as a means of maintaining or achieving these objectives over a reasonable time frame taking into account country circumstances. But there is a commonly-agreed position on the adjustment path that should be taken towards these medium-term objectives through annual fiscal adjustments as summarised in Table 1 reproduced from the EU's (2016) technical paper on the implementation of the SGP. The cells of the table define thirteen potential fiscal adjustments (ranging from making no fiscal adjustment to reducing debt by 1.00 percentage point of GDP in the year). Each cell relates to a conjunction of events defined by the country's output gap (i.e. its output level relative to potential), its growth rate, its debt-to-GDP level and the level of its sustainablility risk (which depends in turn on expected future debt-to-GDP levels). Our approach uses relatively simple VAR, GVAR or GVECM models to forecast the probability of each of these events occurring, using the specified adjustments in the cells as a direct measure of the arguments of the government's objective function, and making recommendations on fiscal adjustments to maximise these objectives. We find that it is indeed important to consider international linkages in these decisions and note that - had the countries followed these rules - failure to acknowledge cross-country interactions in the modelling would have led to excessive fiscal cut-backs and austerity in these countries over the years since the GFC.

The rest of the paper is organised as follows: Section 2 describes the way in which models inform fiscal policy decisions and explains our proposed approach based on event probability forecasting; Section 3 describes our modelling exercise, explaining the GVAR and GVECM framework, describing the preferred forecasting model and presenting our analysis for the OECD10 over 1991-2016; and Section 4 concludes.

2 Modelling and Making Fiscal Policy Decisions

To focus discussion, and to introduce notation, we note that the willingness of lenders to hold debt is often approached through a 'fiscal accounting' analysis starting from the description of debt dynamics:

$$\widetilde{B}_t = (1 + R_t)\widetilde{B}_{t-1} - \widetilde{S}_t, \tag{1}$$

where $\widetilde{B}_t$ is nominal stock of outstanding debt at the end of period t, $\widetilde{S}_t$ is the nominal primary surplus, reflecting the taxes minus the current value of spending on goods and services and transfers during period t, and R_t is the nominal interest rate on the debt during the period. To accommodate the fact that economies grow over time, the dynamic budget constraint is usually written in terms of ratios to GDP:

$$\frac{B_t}{Y_t} = \left(\frac{1 + R_t - \pi_t}{1 + g_t}\right)\frac{B_{t-1}}{Y_{t-1}} - \frac{S_t}{Y_t},$$

or equivalently,

$$\begin{aligned} b_t &= \frac{1 + r_t}{1 + g_t} b_{t-1} - s_t \\ &\approx (1 + r_t - g_t)b_{t-1} - s_t \end{aligned} \tag{2}$$

where B_t, S_t and Y_t denote debt, primary deficit and output in constant prices respectively, b_t is the debt-to-GDP ratio, s_t is the primary surplus-to-GDP ratio, $r_t = R_t - \pi_t$ is the real rate of interest, with π_t denoting inflation, and $g_t = \log(Y_t) - \log(Y_{t-1})$ is output growth.

Writing (2) at $t+h$ and substituting recursively back to time t, the debt-to-GDP ratio in time $t + h$ is equal to the value of the ratio at t compounded according to the growth-adjusted real interest rates holding over the period ($r^g_{t+k} = r_{t+k} - g_{t+k}$, $k = 1, .., h$) plus the sum of the successive primary surpluses compounded in the same way. Equivalently, writing $q_{t,h} = \Pi^h_{j=1}(1 + r^g_{t+j})^{-1}$, the recursive substitution provides

$$q_{t,h} b_{t+h} = b_t - \sum_{k=1}^{h} q_{t,k} s_{t+k} \tag{3}$$

so that the discounted value of the debt ratio at $t + h$ is equal to the current debt ratio plus the sum of the discounted future primary surpluses. The most widely-used notion of a sustainable fiscal policy is a set of policy rules and inherited debt that precludes governments indefinitely paying off past loans with new loans so that

$$\lim_{h\to\infty} E\left[q_{t,h}b_{t+h}\middle|\quad_t\right]=0 \tag{4}$$

where $E[.|\quad_t]$ is the expectations operator and $\quad_t$ denotes the information available at time t.

The expression in (4) illustrates some of the important elements influencing fiscal policy sustainability. But its focus on the infinite horizon means it provides little practical guidance for policy-makers making fiscal adjustments with short- to medium-term interests. Instead, national fiscal councils and international agencies tend to take a more pragmatic approach to fiscal policy decision-making based on modelling of varying degrees of sophistication. For example, Blanchard et al. (1990) focus on the use of two 'tax gap' measures, $\widehat{\tau}_t-\tau_t$ which show the difference between the current tax rate τ_t and a target rate $\widehat{\tau}_t$ that would, taking spending and transfers as given, leave the present value of the debt-to-GDP ratio unchanged over the short- or medium run (i.e. considered over a one- or five-year horizon).[6] The fact that the indicators take the projected path of government spending and receipts as given, either from published plans or as a hypothesised benchmark, illustrates a common weakness in the simple indicator-based approach to fiscal decision-making. This is because agents' willingness to lend depends on their preferences and beliefs on how future policy will react *endogenously* to future circumstances. An assessment of the debt position should take into account the feedbacks between debt and surplus decisions and interest rate and output growth paths, either through a structural model that explicitly describes the interplay between consumer preferences and government policy (as set out in fiscal rules and debt management strategies) or through the forecasts from an empirical model that fully captures the interdependencies observed in the data.[7]

---

[6] They also consider a forty-year horizon in which the long-term implications of an ageing population, large-scale environmental projects and so on play a more influential role.

[7] The structural approach is taken in the illustrations of Bohn (2005) or Horne (1991) for example,

A second weakness of the use of fiscal indicators - even when based on forecasts - is their reliance on simple point forecasts. The tax gaps described above are easy to interpret, since a positive gap means taxes have to be increased or spending reduced to maintain the debt ratio at current levels, but they do not provide an obvious metric against which to judge the seriousness of the debt position or on which fiscal adjustments can be based.[8] Moreover, while the calculation of the target rate makes use of point forecasts of interest rates and output growth over the decision horizon, no account is taken of the uncertainties surrounding these forecasts in the gap measure and this also makes it more difficult to judge the urgency and size of any required fiscal adjustment.[9]

A third weakness in the standard approaches to determining fiscal adjustments is the focus on relatively simple scenarios or forecasts in decision rules; for example, considering the level of the primary surplus that would ensure unchanged debt-to-GDP. In practice, the feedback rules for spending, taxes and transfers, and the nature of the shocks impacting on these magnitudes, can be quite distinct. This means there can be enormous variation in the (endogenously-determined) time paths of output, interest rates and debt, depending on the shocks experienced and the set of government fiscal policies and debt management strategies followed, all of which are consistent with the behaviour of optimising lenders and debt sustainability.[10] In reality, then, the decision-rules involved in making fiscal adjustments are likely to depend on a complicated conjuncture of events - as exemplified by the events specified in the fiscal rules of Table 1 - which cannot be captured by the simple linear combinations of the point forecasts of relevant variables typically considered in the literature.[11]

---

while the latter approach is exempified by Polito and Wickens (2012) say.

[8] Blanchard et al (1990) note that a positive gap is potentially more worrying where initial taxes are alreay high, since there is less scope for government to appropriate resources, so that the scaled measure $\frac{\widehat{\tau}_t-\tau_t}{1-\tau_t}$ might provide a more appropriate indicator of sustainability than the gap itself.

[9] An exception to this statement is Celasun, Debrun, and Ostry (2006) who generate fan charts for primary surpluses in emerging economies as part of the IMF's debt sustainability assessment.

[10] This point is made very clearly in the context of the illustrative DSGE model of Chung and Leeper (2007).

[11] Polito and Wickens (2012) adopt a similar modelling strategy to ours but consider a relatively simple scenario, suggesting the use of a model-based indicator based on a linear combination of forecast variables and compares this with a target level of the debt-GDP that is constant over a given forecast horizon.

2.1 Making Fiscal Adjustments based on Event Probability Forecasts

The discussion emphasises that decisions on fiscal adjustments require a modelling approach which (i) can capture the simultaneity of the determination of debt and other macroeconomic magnitudes; (ii) conveys a sense of the seriousness of the debt position, and the associated response, given the uncertainties surrounding future outcomes; and (iii) acknowledges the wide range of possible future outcomes and provides a correspondingly sophisticated translation into fiscal policy recommendations. These features can all be captured through the use of forecasts of event probabilities based on a simple time-series model of the relevant variables.

To see this, write $\mathbf{x}_t = (b_t, r_t, y_t)'$, a vector containing the debt-to-GDP ratio, real interest rate and (log) output level. Assuming that the three series are difference-stationary, they can be characterised by a vector error-correction model

$$\Delta\mathbf{x}_t = \mathbf{\Gamma}_0 + \sum_{j=1}^{p} \mathbf{\Gamma}_j \Delta\mathbf{x}_{t-j} + \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{x}_{t-1} + \mathbf{u}_t \tag{5}$$

where $\mathbf{\Gamma}_0$, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}'$ are (3×1) vectors of parameters, $\mathbf{\Gamma}_j$, $j = 1, .., p$ are (3×3) matrices of parameters and the $\mathbf{u}_t$ is a (3×1) vectors of shocks with variance-covariance matrix $\mathbf{\Sigma}$. The model of (5) can be estimated to capture the endogenous interplay between debt, real interest rates and output observed in the data, including any equilibrating pressures from any long-run relations that exist between the levels of the series. Of course, the model can also be written in a VAR($p+1$) form

$$\mathbf{x}_t = \mathbf{C}_0 + \sum_{j=1}^{p+1} \mathbf{C}_j \mathbf{x}_{t-j} + \mathbf{u}_t \tag{6}$$

where $\mathbf{C}_0 = \mathbf{\Gamma}_0$, $\mathbf{C}_1 = \mathbf{I} + \boldsymbol{\alpha}\boldsymbol{\beta}' + \mathbf{\Gamma}_1$ and $\mathbf{C}_j = -\mathbf{\Gamma}_j + \mathbf{\Gamma}_{j+1}$, $j = 2, .., p+1$ which, as explained below, is very convenient for simulation purposes.

Using $\mathbf{X}_{t_1,t_2} = (\ \mathbf{x}_{t1}, ..., \mathbf{x}_{t2})'$ to denote the values of $\mathbf{x}_t$ observed between $t = t_1, ..., t_2$, a general event relating to the variables in $\mathbf{x}_t$ at $T+1$ and over the forecast horizon up to $T+H$ is defined by F_T: $\{\ \phi_l(\mathbf{X}_{T+1,T+H}) < a_l$ for $l = 1, 2, .., L\ \}$ or, equivalently, F_T: $\{\ \boldsymbol{\phi}(\mathbf{X}_{T+1,T+H}) < \mathbf{a}\ \}$ where $\boldsymbol{\phi}(.) = (\phi_1(.),\ \phi_2(.),\ ..\phi_L(.))'$ and $\mathbf{a} = (a_1,...,a_L)'$ are $L \times 1$ vectors. A general event is therefore based on the simultaneous occurrence of L (possibly interdependent) individual events $\phi_l(.)$. These individual events can themselves involve complicated non-linear functions of the variables and can involve the variables dated at a variety of different forecast horizons. For example, in Table 1, thirteen general events are described, each defined according to the simultaneous occurrence of various individual events relating to the country's output gap, its growth rate, and its current and expected future debt-to-GDP levels. But if, for example, potential output is defined as a smoothed function of recent and expected future outputs obtained using the Hodrick-Prescott filter say, then even an individual event relating to the output gap will be defined by a complicated function of observed and forecast values of past and future output levels.

The probability density function (pdf) $p(\mathbf{X}_{1,T+H}\ ;\ \boldsymbol{\theta})$ describes the probability of obtaining specified values of the observed and forecasted data in $\mathbf{x}_t$ over the estimation and forecast horizons, $t = 1, ..., T$ and $t = T+1, ..., T+H$ respectively, based on a given model indexed by the parameters $\boldsymbol{\theta} = (\boldsymbol{\Gamma}_0,..,\boldsymbol{\Gamma}_p,\ \boldsymbol{\alpha},\ \boldsymbol{\beta},\ \boldsymbol{\Sigma})$. The pdf can be decomposed into the product of the conditional distributions of the successive observations on $\mathbf{x}_t$, so that it can be written in terms of the density forecast and sample likelihood

$$p(\mathbf{X}_{1,T+H};\ \boldsymbol{\theta}) = p(\mathbf{X}_{T+1,T+H} \mid \mathbf{X}_{1,T};\ \boldsymbol{\theta}) \prod_{j=1}^{T} p(\mathbf{x}_j \mid \mathbf{X}_{1,j-1}\ ;\ \boldsymbol{\theta}), \tag{7}$$

The forecast probability of a particular general event F_T occurring, conditional on information available at T, is given by

$$\begin{aligned} \pi_{F_T}(a,\ H,\ \boldsymbol{\phi}(.),\ \boldsymbol{\theta}) &= \Pr\left[\boldsymbol{\phi}(\mathbf{X}_{T+1,T+H}) < \mathbf{a} \mid \mathbf{X}_{1,T}\ ;\ \boldsymbol{\theta}\right] \\ &= \int_{F_T} p(\mathbf{X}_{T+1,T+H} \mid \mathbf{X}_{1,T};\ \boldsymbol{\theta})\ d\mathbf{X}_{T+1,T+H} \end{aligned} \tag{8}$$

integrating the density forecast $p(\mathbf{X}_{T+1,T+H} \mid \mathbf{X}_{1,T};\ \boldsymbol{\theta})$ over the past and future values of debt, interest rates and output that define the general event F_T. If there are N mutually exclusive events $\{F_{1T}, ..., F_{NT}\}$ that cover all possible outcomes, then $\sum_{i=1}^{N} \pi_{F_{iT}} = 1$.

In practice, these statistics are readily calculated using simulation methods based on the estimated versions of (5) and (6). For example, abstracting from parameter uncertainty but using time-T estimates of the model parameters $\widehat{\boldsymbol{\theta}}_T$, one can make SH random draws from a $N(0, \widehat{\boldsymbol{\Sigma}}_T)$ distribution and use these with the estimated model to generate S replications of the future values of the series, denoted $\widehat{\mathbf{x}}^{(s)}_{T+h}$ for $h = 1, ..., H$ and $s = 1, ..., S$. These simulated future series give directly the real-time forecast joint pdf's of the debt-to-GDP, real interest rate and output series over the relevant forecast horizon; i.e. the estimated density forecasts $\widehat{f}(\mathbf{X}_{T+1,T+H} \mid \mathbf{X}_{1,T};\ \widehat{\boldsymbol{\theta}}_T)$. Simply counting the number of times the event F_{iT} occurs in these simulations, expressed as a fraction of S, also provides the forecast of the probability that the event will occur, $\widehat{\pi}_{F_{iT}}\left(a,\ H,\ \boldsymbol{\phi}(.),\ \widehat{\boldsymbol{\theta}}_T\right)$.[12]

In the context of our fiscal policy decisions, we have argued that we are likely to require a nuanced approach to fiscal adjustments with changes in fiscal policy carefully calibrated to reflect the occurrence of one of a number of possible general events. The fiscal rules of Table 1, for example, describe seven alternative fiscal adjustments as being appropriate depending on which one of thirteen general events occurs. The rules place a scalar value on the events, in terms of the size of the fiscal adjustment, which can be used to define a loss function for the government describing the costs incurred when forecasting events incorrectly. So, if the fiscal adjustment warranted by the event that actually occurs - F_T - is denoted by f_T , and the fiscal adjustment recommended by the model when event F_{iT} is forecast to occur is denoted by $\widehat{f_{iT}}$, for $i = 1, ..., N$, then two examples of possible loss functions are given by

$$\nu(f_T, \widehat{f_{iT}}) = (f_T - \widehat{f_{iT}})^2 \tag{9}$$

$$\text{or} \quad \nu(f_T, \widehat{f_{iT}}) = \left[(f_T - \widehat{f_{iT}})^2 \times I(f_T > \widehat{f_{iT}})\right] + \lambda \left[(f_T - \widehat{f_{iT}})^2 \times I(f_T < \widehat{f_{iT}})\right] \tag{10}$$

for $i = 1, ..., N$, where $I(.)$ represents an indictor function taking the value 1 if its argument is true and zero otherwise. In (9), the government cares equally about making an adjustment that is too large or too small while there is an asymmetry in (10), with the

---

[12] The methods, including those that accommodate model uncertainty and parameter uncertainty as well as the stochastic uncertainty considered here, are described in detail in Garratt et al. (2003).

government caring more about making an adjustment that is too large than one that is too small if λ takes a value greater than 1. The government's decision problem, mapping the observed debt-to-GDP, interest rate and output outcomes to a recommended fiscal adjustment, via the specified events and associated fiscal rules, is to recommend a fiscal adjustment given by

$$f_T^* = \min_{f_T \in \{F_{1T},\ldots,F_{NT}\}} \left\{ \sum_{i=1}^{N} \nu(f_T,\ \widehat{f_{iT}})\, \widehat{\pi}_{F_{iT}} \right\}, \tag{11}$$

choosing the fiscal adjustment that minimises the expected average loss.

In practice, the decision problem of (11) can again be approached in a straightforward way through simulation. Here, for each of the S simulated futures based on our estimated model and used to generate the probability forecasts described above, we identify the general event that is forecast to occur and the associated fiscal adjustment specified by the fiscal rules. Comparison with each of the possible warranted adjustments according to (9) or (10), say, then provides measures of the losses that would be incurred for any outcome under that simulated future. The recommended adjustment is the outcome that minimises the losses when summed (or averaged) over all S simulations. Further, if we repeat this exercise over an out-of-sample evaluation period $T = T_1, \ldots, T_M$, then an economically-meaningful measure of the forecast performance of the model is given by the statistic

$$\overline{\Psi}_{T_1 T_M} = \frac{1}{M+1} \sum_{\tau=1}^{T_M} \nu(f_\tau,\ f_\tau^*), \tag{12}$$

which provides a direct measure of the loss incurred by using the model in this decision-making context. Similar statistics can be calculated for any other model, based on the associated recommended fiscal adjustments, and these provide the basis of a comparison of the forecast performance of the models on economic grounds.[13]

---

[13] See Pesaran and Skouras (2000) for further discussion.

3 Fiscal Adjustments in the OECD10, 1956-2016

In this section, we investigate the usefulness of our approach to making fiscal adjustments using data on public debt, real interest rates and output over the last fifty years in ten OECD countries: namely, Belgium, Canada, Denmark, France, Germany, Ireland, Italy, Netherlands, the UK and the US. Details of the definitions and sources of the data are provided in the Data Appendix. In the first instance, we use the data from 1956-1991 to estimate our models and then roll forward through the sample, producing out-of-sample forecasts and real-time recommendations on fiscal adjustments over the period 1991-2016.

Much of the early work on investigating fiscal policy and debt dynamics focused on the unit root properties of the variables involved and, given that our own analysis is based on a time-series model of debt and related macro variables, we also take care to consider this aspect of the data in what follows. Perhaps more unusually, we also pay attention to the cross-country interactions that exist between the series. This could be important on a purely practical level since there are common cross-country patterns in the data that suggest that forecasting performance might be enhanced when these are taken into account. But there are also economically–meaningful interactions - through interest rate parity, cross-country growth convergence, international capital flows, financial contagion and so on - that suggest that theory-based long-run relations might be present which, if omitted from the analysis, would render any forecasting model misspecified. These observations motivate an analysis of cointegrating relations between the series in different countries and the use of GVAR and GVECM models in addition to county-specific VARs. It will also be useful to examine the forecast performance of the GVAR and GVECM models - relative to that of nation-specific models- both in terms of the standard statistical criteria and, following the earlier discussion of (12), in this particular context of making fiscal adjustments.

3.1 Data Overview

The raw data for the analysis are provided in Figures 1 and 2 and illustrate both the heterogeneity in countries' experiences over the last fifty years and the extent to which they are also subject to important common global influences. For example, there are substantial differences in the level of the debt ratio across countries, averaging at 42% across the sample period in Germany but at 92% in Belgium, for example. And while there has been considerable volatility in most countries, with standard deviations typically in excess of 20% and at 32% in Italy, others have been more stable; e.g. Canada and US at just 14% and 15% respectively). On the other hand, nearly all countries experienced a gradual fall in their debt ratios during the seventies, coinciding with the common experience of high inflation and low real interest rates; most countries saw their ratios rise from the early eighties through to the mid-nineties; and all saw their ratios rise rapidly following the GFC in 2008.

3.1.1 Unit root properties

We test the stationarity of the variables using the Cross-section Augmented Dickey-Fuller (CADF) unit root tests of Pesaran (2007) and the Cross-Section IPS (CIPS) panel unit root test described in Pesaran, Smith, and Yamagata (2013). These procedures exploit the multi-country dimension of our data to provide more powerful tests of the unit root hypothesis. The CADF tests are carried out on a variable separately for each country but take into account any unobserved common factors that might affect the variable across countries by including the cross-country average in the underlying ADF regression. The CIPS test considers whether there is a unit root in the variable in all countries and is based on the average value of the unit root tests across the countries.

The test results are presented in Table 2a and are quite clear in the case of the debt-to-GDP ratio and output. Here there is little evidence with which to reject the unit root hypothesis in the level of the series in any of our ten countries but very strong evidence to reject a unit root in the differenced series. It is clear that these series are integrated of order one - I(1) - variables therefore. The findings are a little more mixed for the real interest rate. There is again strong evidence with which to reject the unit root hypothesis in the differenced series but the hypothesis is also rejected in the levels series in five of our countries and in the relatively powerful CIPS test. Our view on this is that the balance of evidence is for a unit root in the real interest rate too and we treated the series in this way in what follows. But we remained aware of the ambiguity of this result and the implications for the subsequent analysis.

3.1.2 Cointegrating properties

The results of Table 2b investigate the possibility of cointegration between the domestic and foreign variables. The starred 'foreign' measure of a variable for country i, for example, is constructed as a simple cross-section average of the variable in the other countries - e.g. $b_{it}^{*} = \frac{1}{9}\sum_{j \neq i} b_{jt}$. For the debt-to-GDP series, the statistics of Table 2b show that, despite the apparent similarity in patterns in the movements of debt ratios across countries, there is almost no evidence of cointegrating relations across countries. This means that, if a country-specific shock causes the debt ratio to rise in that country alone, then the effects of the increase and the gap that arises with other countries, will persist indefinitely. The evidence on real interest rates and output is more mixed however. Looking across our ten countries, both series show some evidence with which to reject the null of no cointegration when using Johansen's 'trace statistic' test which imposes no structure on the form of any cointegrating relation. However, the evidence for cointegration only survives for the real interest rate when the test is based on a ADF test in which the cointegrating vector of $(1,\ 1)'$ is imposed. This restricted test is more appropriate if we think the cointegration is based on cross-country arbitrage in investment opportunities (for interest rates) and growth convergence (for output) and so incorporates more economic structure. Taken together, we interpret these results as providing little support for cross-country cointegration in outputs but support for the presence of equilibrating relations in real interest rates across countries. The latter are therefore included in one of our forecasting models discussed below.

3.2 The Forecasting Models

We turn now to the alternative models of debt, real interest rates and output which we might consider as candidates for producing the forecasts on which our fiscal space measures can be based. Clearly forecast accuracy is an important criterion for choosing our model but this can depend many factors, including the trade-off between using many variables (to avoid omitting any that are genuinely part of the data-generating process) and restricting the set of variables to minimise the error in estimating model parameters. With this trade-off in mind, we considered three models: a model based on nation-specific vector-autoregressions explaining the growths of the three series of interest (VAR); a Global Vector-Autoregressive (GVAR) which extends the VAR to allow for feedbacks between growths across countries; and a Global Vector Error-Correction (GVECM) model which allows for further cross-country feedbacks through the influence of the equilibrating relations between the levels of domestic and foreign real interest rates.[14]

The GVECM can be described by extending the error correction model of (5) for country $i = 1, .., 10$ to write

$$\Delta\mathbf{x}_{it} = \mathbf{\Gamma}_{i0} + \mathbf{\Gamma}^*_{i0}\Delta\mathbf{x}^*_{it} + \sum\nolimits_{j=1}^{p}\mathbf{\Gamma}_{ij}\Delta\mathbf{x}_{t-j} + \sum\nolimits_{j=1}^{p}\mathbf{\Gamma}^*_{ij}\Delta\mathbf{x}^*_{t-j} + \boldsymbol{\alpha}_i\boldsymbol{\beta}'_i(\mathbf{x}'_{i,t-1},\ \mathbf{x}^{*\prime}_{i,t-1})' + \mathbf{u}_{it}, \qquad i = 1, ..., 10 \tag{13}$$

so that country i's debt growth, real interest rate growth and output growth depend on the corresponding 'starred'-variables abroad contemporaneously, on lagged growth at home and abroad, and on equilibrating relations between the levels of the domestic and foreign variables $\boldsymbol{\beta}'_i(\mathbf{x}'_{i,t-1},\ \mathbf{x}^{*\prime}_{i,t-1})$. Now, writing $\mathbf{z}_t = (\mathbf{x}'_{1t},\ \mathbf{x}'_{2t},\ ...,\mathbf{x}'_{nt})'$ to obtain the 30×1 vector containing all the variables of interest, and noting that the foreign starred-variables can be written in terms of weighted sum of variables in $\mathbf{z}_t$ (i.e. $\mathbf{x}^*_{it} = \mathbf{W}_i\mathbf{z}_t$), the country-specific equations in (13) can be stacked to write

$$\Delta\mathbf{z}_t = \mathbf{A}_0 + \mathbf{B}^*_0\Delta\mathbf{z}_t + \sum\nolimits_{j=1}^{p}(\mathbf{A}_j + \mathbf{B}_j)\Delta\mathbf{z}_{t-j} + \mathbf{\Pi}\mathbf{z}_{t-1} + \mathbf{e}_t, \tag{14}$$

---

[14] See Pesaran et al. (2004) and Garratt et al. (2012) for more detailed discussions of the GVAR and GVECM models.

where the $\mathbf{A}$'s, $\mathbf{B}$'s and $\mathbf{\Pi}$ are based on the $\mathbf{\Gamma}^*$'s, $\boldsymbol{\alpha}$'s, $\boldsymbol{\beta}$'s and $\mathbf{W}$'s,[15] and hence

$$\mathbf{z}_t = \mathbf{C}_0 + \sum_{j=1}^{p+1} \mathbf{C}_j \mathbf{z}_{t-j} + \boldsymbol{\varepsilon}_t, \tag{15}$$

where $(\mathbf{I} - \mathbf{B}_0^*)^{-1} = \mathbf{G}$, $\mathbf{C}_0 = \mathbf{G}\mathbf{A}_0$, $\mathbf{C}_1 = \mathbf{G}(\mathbf{I} + \mathbf{A}_1 + \mathbf{B}_1 + \mathbf{\Pi})$, etc., and $\boldsymbol{\varepsilon}_t = \mathbf{G}\mathbf{e}_t$.[16]

The model in (15) is of the same form as (6) so can be used to generate event probability forecasts relating to the variables in many countries or to variables in just one country exactly as described above. Of course, in the latter case where the focus is on the forecasts of variables from a single country, the GVECM still takes into account the cross-country feedbacks from growths and the effects of any equilibrating cross-country relations in levels in generating the single-country forecasts. Both the GVAR model and the VAR model are nested within the GVECM: GVAR is obtained by setting $\boldsymbol{\alpha}_i = 0$ $\forall i$ so that there are no equilibrating feedbacks from the levels; and the VAR model is obtained by also setting the $\mathbf{\Gamma}_{ij}^* = 0$ $\forall i, j$ so that no starred-variables are included in the countries' underlying autoregressions. GVECM is able to capture all the potentially important influences that might influence forecasts, while GVAR and VAR are not, but GVECM is also the most highly parameterised and so could be the most vulnerable to the sort of parameter uncertainty that damages forecast performance.

Before using the models in forecasting, we also undertake a specification search to deal with over-paramerisation[17] and the effects of extreme observations which may have a disproportionate effect on the modelling exercise. To deal with potential over-parameterisation in the GVAR and GVECM models, we restrict $\mathbf{\Gamma}_{i0}^*$ and $\mathbf{\Gamma}_{ij}^*$ in (13) to be diagonal (so that the cross-country feedbacks to each variable in $\mathbf{x}_{it}$ are only from the corresponding foreign variable). We also implement a procedure in each rolling window whereby variables with coefficients whose t-values are less than one in absolute terms are dropped. Variables are

---

[15] For example, if equal weights are used in constructing the starred variables, then $\mathbf{W}_i = \frac{1}{n}(\mathbf{I}_3, \mathbf{I}_3, ..., \mathbf{I}_3)$ and defining $\mathbf{G}_{i0}$ by the $3 \times n$ matrix $\frac{1}{n}(\mathbf{\Gamma}_{i0}^*, \mathbf{\Gamma}_{i0}^*, ..., \mathbf{\Gamma}_{i0}^*)$, $\mathbf{\Gamma}_{i0}^* \Delta \mathbf{x}_{it}^* = \mathbf{G}_{i0} \Delta \mathbf{z}_t$ and $\mathbf{B}_0^* = (\mathbf{G}_{10}, \mathbf{G}_{20}, ..., \mathbf{G}_{n0})'$.

[16] The variance-covariance matrix of $\mathbf{e}_t$, $\mathbf{\Sigma}$, can be assumed to be block diagonal, with 3×3 matrices on the diagonal. Then the variance-covariance of $\boldsymbol{\varepsilon}_t$ is $\mathbf{G}\mathbf{\Sigma}\mathbf{G}'$.

[17] In practice, we consider a lag length of $p = 1$ in the GVECM, justified on the basis of lag length reduction tests and tests of residual serial correlation. Even restricting attention to versions where cointegrating relations take the form $(1, -1)'$, the forecasts based on an unrestricted GVECM are based on 30 equations each involving up to 13 parameters therefore.

dropped one-at-a-time, starting with the longest lags, and tests conducted at each stage to check that the system remains globally stable and that the model fit has not been significantly reduced. To deal with outliers, in each rolling window we conduct a Chow test of predictive failure on the final observation. If it is deemed an outlier, we include an additive dummy and re-estimate the model, effectively dummying out the extreme observation in the estimation and using the pre-outlier model to generate the forecast.

3.3 The Event Probability Forecasts and Recommended Fiscal Adjustments

3.3.1 Density Forecast Performance

Table 3 describes the forecast performance of the various models in predicting debt ratios, real interest rates and output at the one-year-, two-year- and three-year-ahead forecast horizons in purely statistical terms. The table focuses on the density forecasts generated by the models over the near future (up to three years) given that our interest is ultimately on event probability forecasts which depend on both the location and uncertainty surrounding the forecasts over this sort of time horizon.[18] We use the VAR models as the benchmark, reporting the log score of forecasts for each country made in real-time averaged over the period 1991-2016.[19] The table also reports the improvement or deterioration in the average obtained using the GVAR and GVECM models. The statistical significance of the improvement/deterioration is tested using the Giocomini-White test of equal forecast performance. The GVAR or GVECM model with the highest average log score is highlighted in bold if it provides a significantly better forecast than the VAR model.

The results of the table show that, in purely statistical terms, the extra complexity of the global models, compared to the VAR models, has relatively little impact on the fore-

---

[18] There is little to choose between the models in terms of point forecasts, as measured by Root Mean Squared Forecast Errors (RMSE): the GVAR model typically has RMSEs larger than the VAR model but not significantly so, while the GVECM typically has lower RMSEs but again not significantly so.

[19] The log score is based on the forecast probability of the observed outturn and is high when the observed outturn lies around the centre of the main body of the density forecast. Hence the log score reflects both the location and uncertainty captured by the density forecast (penalising both large forecast variances that reflect a lack of precision and too narrow density forecasts that are spuriously precise).

casting performance for debt-to-GDP or interest rates but are more influential for output forecasts. The differences between the log scores obtained from the global models and those from the national VAR models are statistically significant in only a small number of cases for debt-to-GDP and real interest rate forecasts, although it is worth noting that the GVAR model typically performs worse than the VAR while the GVECM typically performs better than the VAR, showing that the inclusion of the cointegrating relation between domestic and foreign real interest rates is useful in this context. On the other hand, for output, the log scores for the GVAR and GVECM models show a statistically significant improvement over the VAR model in 9/10 and 10/10 cases respectively.[20] So while there was little evidence of long run relations between the variables at home and abroad, there are important interactions through contemporaneous shocks and interacting short-run dynamics that make it worthwhile to accommodate these cross-country effects.

3.3.2 Evaluating the Event Probability Forecasts by Statistical Criteria

A model's ability to predict the occurrence of an event will be influenced by the accuracy of its density forecasts but it will also depend on the nature of the event as this will focus attention on particular features of the forecast density; for example, the event might be an extreme one that emphasises the forecast performance in the tails. Table 4 considers the ability of the VAR, GVAR and GVECM models to forecast individual events relating to the fiscal rules outlined in Table 1, again judging their performance at this stage by formal statistical criteria. As noted earlier, the individual events are themselves relatively complicated and involve functions of variables dated at a variety of different forecast horizons. We define the individual events as follows:

> "*Normal or good times*", means that the one-year-ahead (log) output gap is greater than -1.5, so that actual output is no more than 1.5% lower than a potential output level, itself defined by the smoothed moving average of output over the past five years adjusted to accommodate trend growth over that time;

---

[20] This echoes the findings of Garratt et al. (2018) who found global effects to be as important as national ones in explaining output fluctuations in the G7.

"*Above-trend output growth*", means that one-year-ahead output growth is expected to exceed one-year-ahead growth in potential output;

"*Acceptable debt-to-GDP*", means that the three-year-ahead debt-to-GDP ratio is expected to be less than 60%; and

"*No sustainability risk*", means that the indicator $S1$ is less than zero where, given the current debt-to-GDP level and projected output and interest rates, $S1$ is the cumulative required change in the primary balance-to-GDP over the next five years to ensure debt-to-GDP ratio is less than 60% within 15 years.

To judge which of the VAR, GVAR and GVECM models are best able to predict these events, the statistics of Table 4 report the accuracy rates, associated Kuipers Scores [KS] and Pesaran and Timmermann (2009) [PT] statistics assuming that we predict these events will occur if the forecast probabilities from the different models exceed 0.5.[21] Table 4a, for example, reports accuracy rates and KSs of the three models' predictions that, one year ahead, the economy will be in 'normal or good' times, calculated recursively over the forecast evaluation period 1991-2011.[22] The unconditional probability of the event is around 0.90 for most countries and the fact that most countries are nearly always in normal times (by definition) means that the performance of the models - against each other and against an uninformed guess based on the unconditional probability - depends on their ability to forecast relatively unusual events. As it turns out, the superior performance of the global models over the nation-specific models in density forecasting for output does carry over to the event probability forecasts, with the GVAR model being the best-performing model in nine out of ten countries (matched by the GVECM in seven) and - in contrast to the national models - with most global models showing significant

---

[21] If 'YY' indicates occassions where a safe debt level is forecasted to be achieved and it is, 'YN' indicates a safe debt level is forecasted to be achieved but it is not, and so on, then the accuracy rate is calculated as $AR = (YY + NN)/(YN + NN + NY + YY)$. The Kuipers Score is defined by $H - F$, where $H = YY/(YY - YN)$ and $F = NY/(NY + NN)$ and takes values between -1 and 1 and is equal to zero when predictions are made at random according to the unconditional probabilities. The PT statistic tests the associated null that the model is no more successful in predicting outcomes than forecasts based only on the unconditional probability.

[22] The evaluation period ends in 2011 because the 'no sustainability risk' event is based on five-year-ahead forecasts.

improvements over forecasts based on the unconditional probability alone.

The event considered in Table 4b is based on the difference between forecasts of two relatively volatile growth series with a difficult-to-predict outcome reflected by the unconditional probabilities that are around 50/50 for all countries. The results show that none of the models performs systematically better than the others and forecast probabilities are typically no better than those based on the unconditional probability. Table 4c relates to an event at the other end of the spectrum in the sense that an 'acceptable debt-to-GDP' is not achieved in any year of the forecast evaluation period in three of our ten countries (so that outcomes are completely predictable). The global models outperform the nation-specific VARs in five of the remaining seven countries though and significantly outperform the forecasts based on the unconditional probability in nearly all of these cases. Finally Table 4c relates to a complex event but one which is defined averaging over a five-year period and so is relatively easy to predict. This means that all of the models perform well (with Kuipers scores often greater than 0.9) with none clearly outperforming the others.

The results of Table 4 highlights that a model's usefulness depends not only on its statistical adequacy but also on its relevance to the task at hand. There are events which are close-to-perfectly predictable and there are events that are close-to-impossible to predict for any model and there is no way to distinguish between models in these cases. For the specific individual events discussed in Table 4, there is evidence that the global models outperform the nation-specific models in some dimensions. But it remains to be seen whether this matters when looking at the particular conjunction of events described in Table 1.

3.3.3 Evaluating the Event Probability Forecasts when Making Fiscal Adjustments

Table 5 describes the losses - expressed in terms of the size of mistakes made in making fiscal adjustments - if the VAR, GVAR and GVECM models were used to inform fiscal policy decisions following the rules set out in Table 1.[23] Table 5a refers to the case described in (9) where the government cares equally about making an adjustment that is too large or too small, and Table 5b refers to the case in (10) where we set $\lambda = 2$ so that the government cares twice as much about reducing fiscal spending unnecessarily than when it fails to reduce net spending by as much as is warranted by the outturn of events. The statistics are calculated as in (12), showing the average loss calculated over the forecast evaluation period when the various models are estimated and used recursively to minimise expected average loss as in (11).

Both sets of results provide support for the use of the global models in making countries' fiscal adjustments. In Table 5a, losses in the GVAR model are more than 5% lower than the losses incurred using the VAR models in six out of ten countries, broadly equal in two countries and more than 5% larger in two countries. The corresponding split is 5:3:2 in the GVECM case. In Table 5b, losses in the GVAR model are more than 5% lower than those from the VAR model in seven countries, broadly equal in one and larger in two (and the split is the same for the GVECM). The results of Table 6 give some insights on where these benefits arise, reporting the (average) size of the underlying errors made over the whole forecast evaluation period 1990-2011 and over the period 2008-2011 following the GFC. The statistics of Table 6a show that, across the whole evaluation period and making decisions to minimise average losses according to (9), the recommendations from the VAR model produced negative average errors across almost all countries while those from the global models produced (smaller but) positive average errors in all countries. This means the recommendations from the VAR model were to pursue austerity in fiscal policy more actively than necessary. Indeed, averaging across the ten countries, the adjustments were 0.13 percentage point of GDP more austere than those from the GVAR model for example.[24] This is especially true in the years following the GFC where the errors were on average 0.37 percentage point of GDP more austere than those from the

---

[23] In practice, we interpret the recommendation to 'adjust by >0.5' to mean 'adjust by 0.625', and the recommendation to 'adjust by >0.75' is interpreted to mean 'adjust by 0.875'.

[24] With the error in each period defined by the warranted adjustment minus the recommended adjustment, a negative error means that the recommendation from the model was to cut spending by more than necessarily.

GVAR model . When the decision takes into account the asymmetry in losses captured by (10), Table 6b shows that the recommendations based on the VAR produced errors that are positive on average across the whole evaluation period, but they are still smaller than those from the global models. The VAR models continue to recommend more austerity than the global models therefore (to the value of 0.11 percentage points of GDP on average). The same conclusion holds again, and more strikingly, in the period following GFC.

We noted earlier that, in terms of forecasting debt-to-gdp at least, there is relatively little to choose between the performance of the models when judged by statistical criteria. But, given the slow-moving nature of debt and the simple dichotomous way in which debt enters decisions, this has little impact on the economic evaluation of the models. Rather it is the superior statistical performance of the forecasts of the output gap which, given the rules and loss functions we have employed, translates into improved decision making when cross-country interactions are taken into account. In short, in this analysis, the inclusion of global interactions in our models generates benefits in estimating some density forecasts and individual event probability forecasts, and these benefits carry over to forecasting the conjunction of events relevant to fiscal adjustments and to the decisions on fiscal adjustments themselves. Moreover, we observe that the greater losses incurred by using the nation-specific VAR models instead of the global models are caused by the pursuit of fiscal policies that are more austere than is necessary.

4 Conclusion

The experiences of the GFC and European Sovereign Debt Crisis exposed very clearly the need for countries to pursue a transparent and systematic approach to making fiscal policy adjustments. A key component of such an approach is the formulation of nuanced fiscal rules - describing explicitly the appropriate adjustments in a range of circumstances - and an understanding of the losses incurred when mistakes are made. The approach also requires a forecasting model to acknowledge that the future values of the relevant variables are time-varying and determined jointly and endogenously and to take into account the uncertainties surrounding the future values of these variables. And, of course, it requires the model to relate to the circumstances identified as important by the fiscal rules. This paper shows that, in the context of the decision rules employed in the EU's Growth and Stability Pact, this can be done in a straightforward way using event probability forecasts obtained from relatively simple times series models involving just a small number of variables. The analysis of the data from ten OECD countries demonstrates the strength and flexibility of the approach and, in establishing the importance of accommodating cross-country interactions in the models underlying these decisions, it also provides a clear illustration of the importance of judging models by economic, as opposed to statistical, criteria.

5 Data Appendix

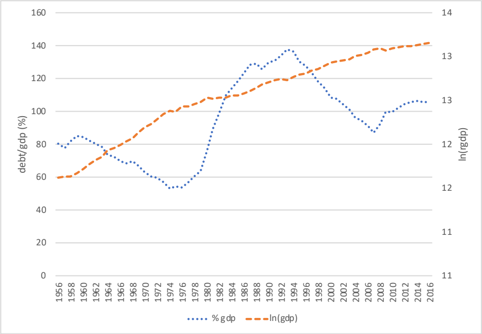
Public debt. The debt-to-GDP series is the ratio of gross general government debt (including central and local government debt) expressed as a ratio to GDP. The debt is valued at par. For three countries (Italy, the Netherlands and the United Kingdom), annual observations for 1956-2016 are obtained from the *Annual Macro-Economic (AMECO) Database* of the European Commission's Directorate General for Economic and Financial Affairs. For other countries, the same source is used for 1980-2016 and this is spliced with data for the earlier period using central government debt. The latter series are from http://www.carmenreinhart.com/data/browse-by-topic/topics/9/ updating the series discussed in Reinhart and Rogoff (2011).

Real Interest Rate. The real interest rate is the ten-year government bond yield deflated contemporaneously by the annual inflation rate. Bond yields are from the IMF's *International Financial Statistics* (*IFS*) and are the average yield over the year. Inflation is measured by year-on-year price changes measured by the CPI from the OECD's *Main Economic Indicators* for Germany and the UK and from the *IFS* for the other countries.

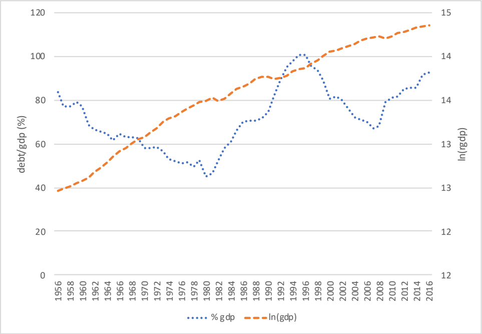
Output. Output is the real gross domestic product (GDP) based on purchasing power parity and measured in 2016 U.S. prices obtained from the *Total Economy Database* maintained by the Conference Board (previously known as the Groningen Growth and Development Centre).

Figure 1: **Output Levels and Debt-to-GDP Ratios**

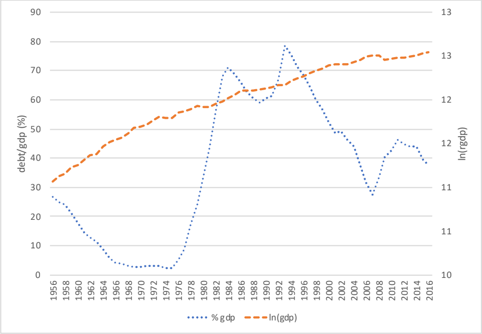
(a) **Belgium**



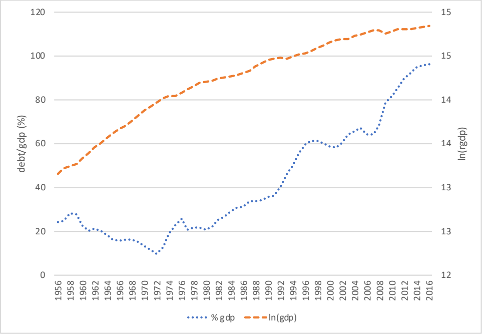
(b) **Canada**



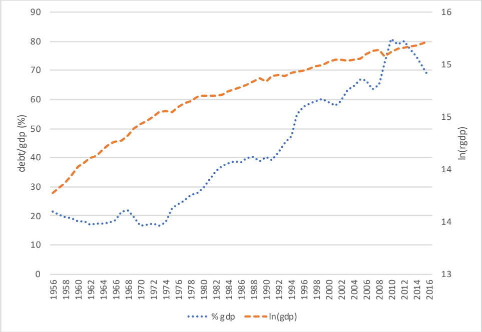
(c) **Denmark**



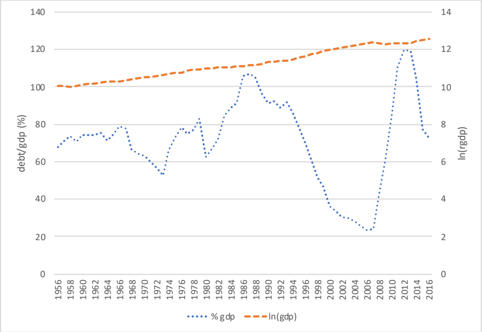
(d) **France**



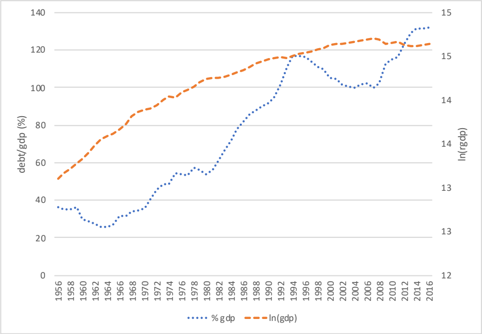
(e) **Germany**



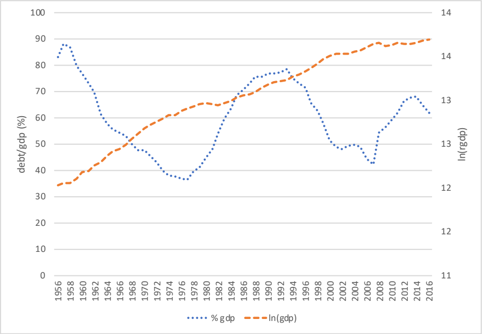
(f) **Ireland**



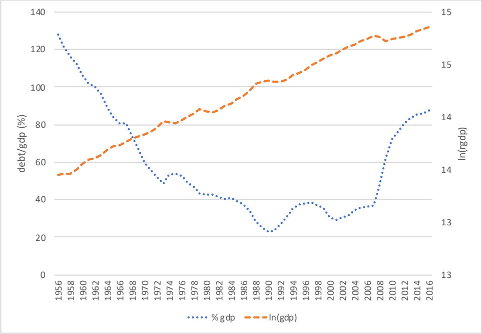
(g) **Italy**



(h) **Netherlands**



(i) **UK**



(j) **US**

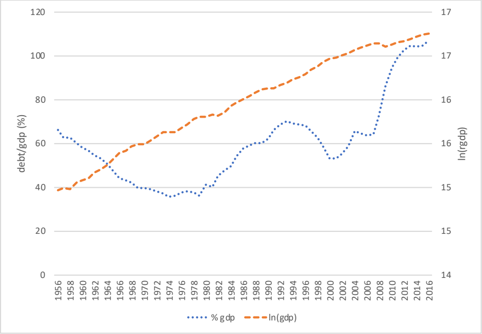
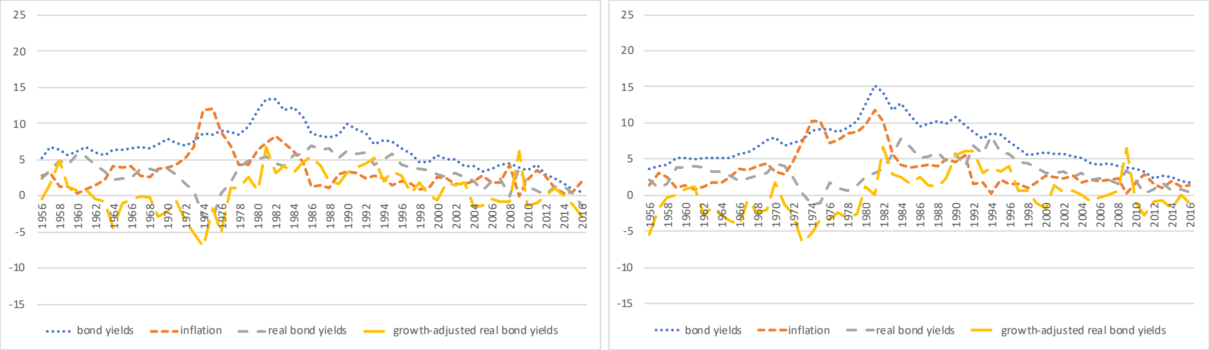
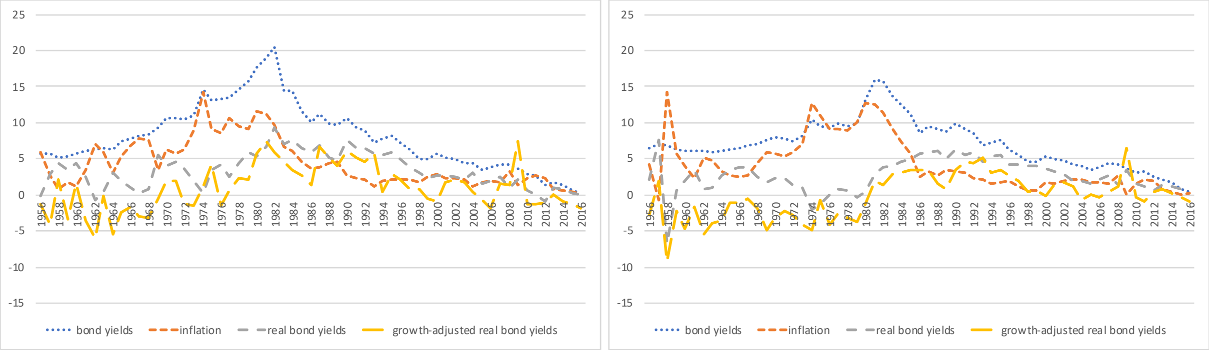
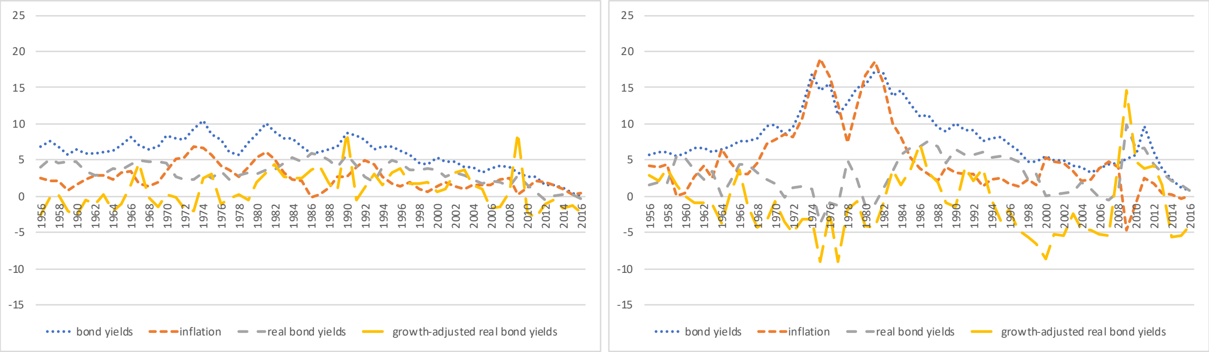
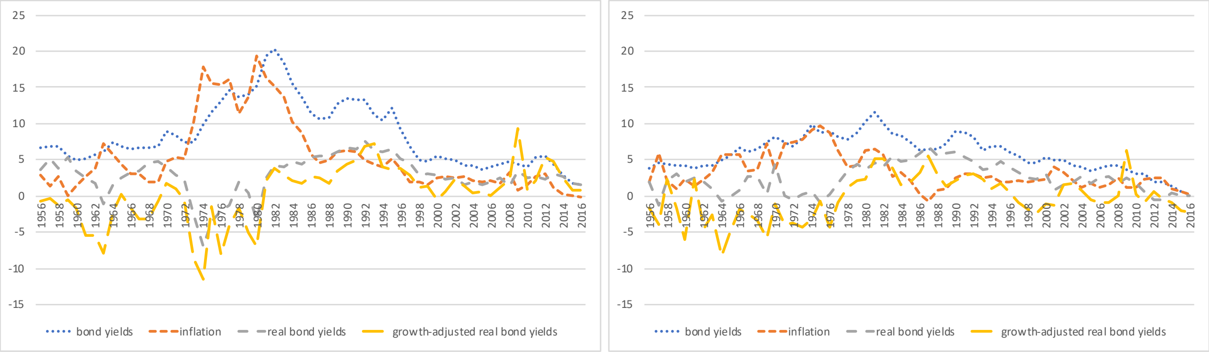


Figure 2: **Bond Yields, Inflation and Growth-Corrected Real Interest Rates**


(a) **Belgium** (b) **Canada**


(c) **Denmark** (d) **France**


(e) **Germany** (f) **Ireland**


(g) **Italy** (h) **Netherlands**

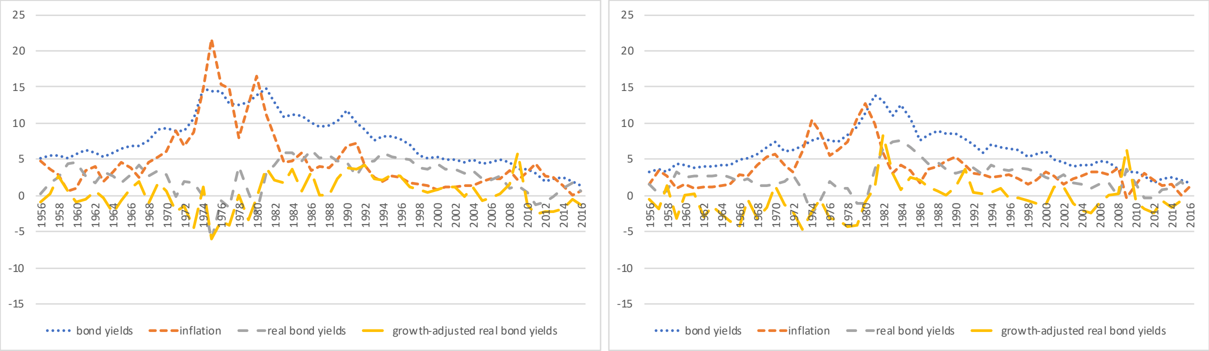
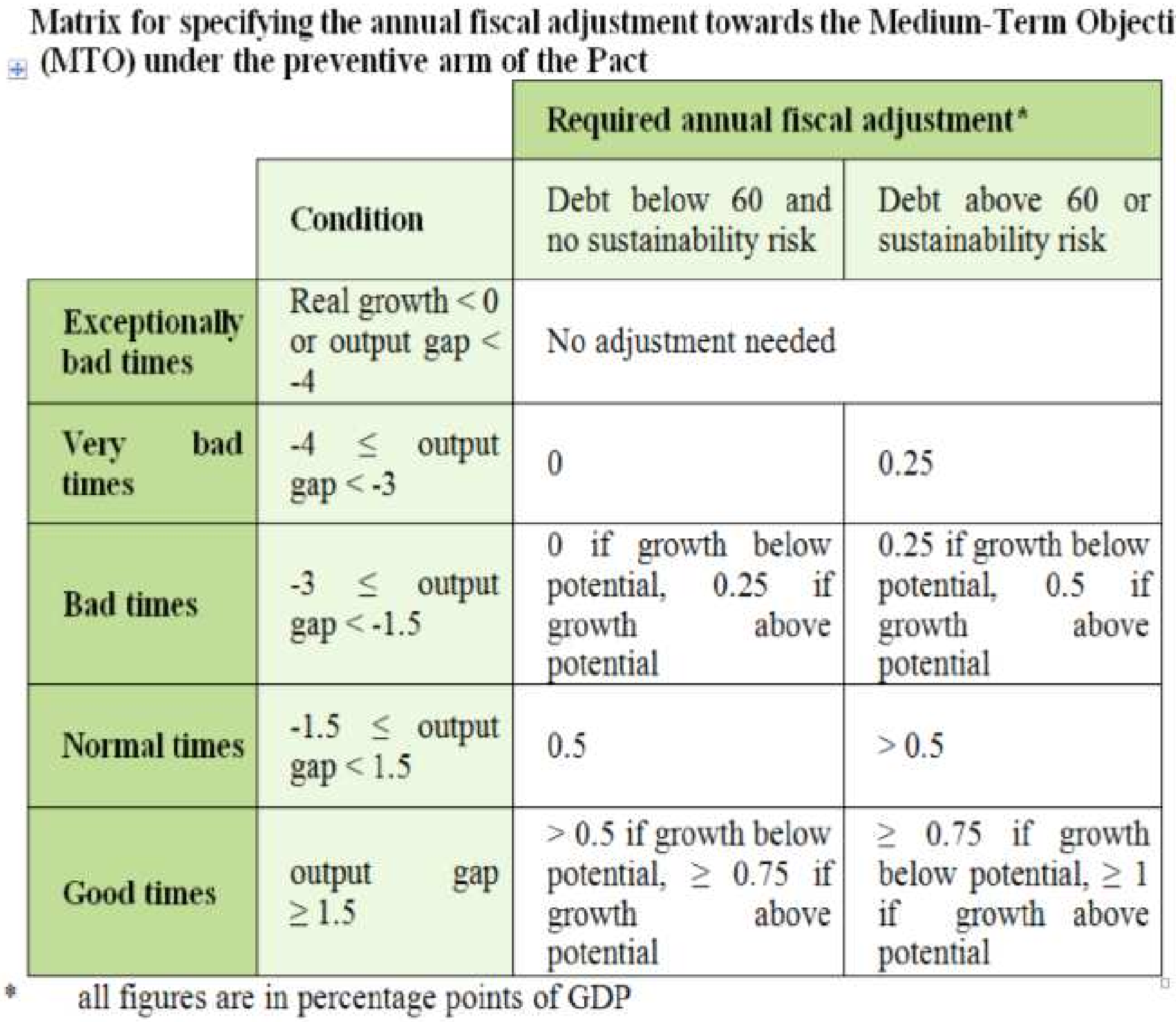

(i) **UK** (j) **US**

Table 1



Matrix for specifying the annual fiscal adjustment towards the Medium-Term Objectiv
(MTO) under the preventive arm of the Pact

		Required annual fiscal adjustment*	
	Condition	Debt below 60 and no sustainability risk	Debt above 60 or sustainability risk
Exceptionally bad times	Real growth < 0 or output gap < -4	No adjustment needed	
Very bad times	-4 ≤ output gap < -3	0	0.25
Bad times	-3 ≤ output gap < -1.5	0 if growth below potential, 0.25 if growth above potential	0.25 if growth below potential, 0.5 if growth above potential
Normal times	-1.5 ≤ output gap < 1.5	0.5	> 0.5
Good times	output gap ≥ 1.5	> 0.5 if growth below potential, ≥ 0.75 if growth above potential	≥ 0.75 if growth below potential, ≥ 1 if growth above potential

* all figures are in percentage points of GDP

Table 2a: Single Country Unit Root Tests, 1956 - 2016

	CADF(2)			CADF(1)		
	b	r	y	Δb	Δr	Δy
Bel	-1.07	-4.75***	-2.74	-2.88	-5.39***	-3.93***
Can	-2.80	-5.18***	-2.62	-3.86**	-6.90***	-5.00***
Den	-0.94	-1.43	-2.75	-2.83	-7.67***	-6.16***
Fra	-1.78	-5.78**	-3.39	-3.29***	-6.22***	-3.96***
Ger	-2.06	-3.36	-2.15	-3.53**	-7.26***	-4.92***
Ire	-4.78***	-3.05	-2.11	-3.36**	-5.71***	-3.32**
Ita	-1.79	-2.65	-1.50	-3.37***	-7.23***	-4.60***
Net	-1.35	-1.93	-2.45	-3.98***	-8.42***	-4.53***
UK	0.36	-3.23	-2.96	-1.97	-7.15***	-5.32***
US	0.34	-3.86*	-3.15	-3.57**	-6.28***	-5.21***
CIPS	-1.59	-3.52***	-2.58	-3.26***	-6.82***	-4.69***

Note: CADF(p) statistics for levels use ADF regressions with an intercept, linear trend and p lagged differences of the dependent variable, plus the lagged level and contemporaneous and p lagged differences of the cross-section average. The CADF(p) statistics for first differences omit the time trend. Relevant critical values are tabulated in Pesaran (2007). The CIPS test statistic is the cross-section mean of the CADF tests and is compared with standard normal distribution. The '*' '**' and '***' denote significance at the 10%, 5% and 1% level respectively.

Table 2b: Cross-Country Cointegration Tests, 1956 - 2016

	Trace Test			CADF		
	b,b*	r,r*	y,y*	b-b*	r-r*	y-y*
Bel	0.71	0.04**	0.08*	0.80	0.17	0.81
Can	0.13	0.00***	0.01***	0.47	0.17	0.85
Den	0.78	0.02**	0.00***	0.93	0.10*	0.97
Fra	0.31	0.00***	0.01***	0.98	0.00***	0.88
Ger	0.23	0.22	0.00***	0.44	0.19	0.98
Ire	0.07*	0.01***	0.00***	0.74	0.02**	1.00
Ita	0.36	0.14	0.00***	0.65	0.06*	1.00
Net	0.23	0.01***	0.07*	0.83	0.23	0.31
UK	0.14	0.01***	0.01***	0.51	0.01***	0.00***
US	0.52	0.00***	0.02**	0.99	0.06*	0.94

Note: The table reports p-values for the null of no cointegration between domestic and foreign variables. The Trace test is Johansen's (1988) rank test for cointegration. CADF is the unit root test applied to the difference between the series, imposing the cointegrating vector to be $(1, -1)'$.

Table 3: Evaluation of the Density Forecasts, 1990 - 2011

Average Log Scores for debt-to-GDP Forecasts

	VAR	GVAR	GVECM
Bel	*-1.52*	**-0.51***	**-0.44***
Can	*-1.89*	-0.31	-0.07
Den	*-2.55*	-0.05	0.10
Fra	*-2.23*	-0.18	**0.22***
Ger	*-2.57*	0.07	**0.17***
Ire	*-2.25*	-0.06	0.05
Ita	*-2.71*	-0.15	**0.13***
Net	*-2.32*	-0.13	0.08
UK	*-3.77*	-0.19	0.27
US	*-2.72*	-0.11	-0.10

Average Log Scores for Real Interest Rates Forecasts

	VAR	GVAR	GVECM
Bel	*-1.40*	-0.02	0.08
Can	*-1.54*	-0.04	0.09
Den	*-1.23*	-0.11	-0.03
Fra	*-1.43*	-0.06	**0.19***
Ger	*-1.75*	0.03	0.09
Ire	*-1.83*	**0.29***	0.20
Ita	*-1.20*	-0.01	**0.12***
Net	*-1.28*	-0.08	-0.03
UK	*-1.33*	-0.12	-0.04
US	*-1.50*	0.04	0.16

Average Log Scores for Output Growth Forecasts

	VAR	GVAR	GVECM
Bel	*-1.68*	**0.34***	**0.37***
Can	*-1.50*	**0.14***	**0.14***
Den	*-1.88*	**0.25***	**0.28***
Fra	*-2.01*	**0.36***	**0.41***
Ger	*-1.55*	**0.15***	**0.19***
Ire	*-1.91*	**0.09**	**0.19***
Ita	*-2.91*	**0.57***	**0.60***
Net	*-1.77*	**0.32***	**0.33***
UK	*-1.70*	**0.20***	**0.27***
US	*-1.94*	**0.23***	**0.24***

Notes: The VAR , GVAR and GVECM models are described in the text. The ALS's of the VAR forecasts are in italics; the ALS's of GVAR and GVECM are expressed relative to the VAR as: $\tilde{ALS}_i = (ALS_{VAR} - ALS_i)/ALS_{VAR}$ i = GVAR and GVECM. A * denotes that the ALS is significantly different from the VAR model, at the 10% level of significance, applying Giacomini and White's (2006) test of equal forecast performance. The highest ALS which is statistically better than the benchmark is in bold.

Table 4a: Evaluation of 'Normal or Good Output Gap' Probability Forecasts, 1990 - 2011

	Accuracy Rate				Kuipers Score		
	p	VAR	GVAR	GVECM	VAR	GVAR	GVECM
Bel	0.95	0.95	**0.95**	0.91	0.00 (-,na)	**0.95** (***,***)	0.90 (***,***)
Can	0.91	0.91	**0.91**	**0.91**	0.00 (-,na))	**0.45** (na, na)	**0.45** (na, na,)
Den	0.95	0.95	**1.00**	**1.00**	0.00 (-,na)	**1.00** (na, na)	**1.00** (na, na)
Fra	0.95	0.95	**0.91**	**0.91**	0.00 (-,na))	**0.90** (**,***)	**0.90** (**,***)
Ger	0.95	0.91	**0.91**	**0.91**	−0.05 (na,na)	**0.90** (***,***)	**0.90** (***,***)
Ire	0.77	0.82	**0.82**	**0.82**	0.34 (*,*)	**0.34** (*,*)	**0.34** (*,*)
Ita	0.95	0.95	**0.95**	**0.95**	0.00 (-,na))	**0.95** (***,***)	**0.95** (***,***)
Net	0.91	0.86	0.91	0.86	−0.05 (-, -)	0.40 (na, na)	0.45 (−,*)
UK	0.91	0.91	**1.00**	**1.00**	0.00 (-,na)	**1.00** (na, na)	**1.00** (na, na)
US	0.95	0.95	**1.00**	**1.00**	0.00 (-,na)	**1.00** (na, na)	**1.00** (na, na)

Notes: p is the unconditional probability. The 'Normal or Good Output Gap' event is defined in the text. The model with the highest accuracy rate and the largest Kuipers Score is in bold. The figures (a,b) below the KS are the static and dynamic Pesaran and Timmerman (2009) tests statistics for no additional predictive power beyond the unconditional probability; '***', '**' and '*' indicate significance at 1%, 5% and 10% levels respectively; '-' indicates no significance at 10% level; and 'na' indicates that the test cannot be conducted (e.g. if the event is forecast to occur in every period).

Table 4b: Evaluation of 'Above-Trend Output Growth' Probability Forecasts, 1990 - 2011

	Accuracy Rate				Kuipers Score		
	p	VAR	GVAR	GVECM	VAR	GVAR	GVECM
Bel	0.55	0.64	0.64	**0.68**	0.20 (-, -)	0.30 (-, -)	**0.38** (*, -)
Can	0.55	**0.77**	0.68	0.73	**0.53** (**, *)	0.38 (*, -)	0.47 (**, -,)
Den	0.64	**0.68**	0.50	0.55	**0.29** (-, -)	0.11 (-, -)	0.23 (-, -)
Fra	0.55	0.64	**0.68**	0.64	0.25 (-, -)	**0.38** (*, -)	0.28 (-, -)
Ger	0.50	**0.55**	**0.55**	**0.55**	**0.09** (-, -)	**0.09** (-, -)	**0.09** (-, -)
Ire	0.59	**0.59**	0.41	0.41	**0.27** (-, -)	−0.03 (-, -)	−0.03 (-, -)
Ita	0.50	0.59	0.59	**0.64**	0.18 (-, -)	0.18 (-, -)	**0.27** (-, -)
Net	0.55	0.73	0.68	**0.73**	0.43 (*, -)	0.42 (**, -)	**0.50** (**,*)
UK	0.55	0.73	0.68	**0.77**	0.48 (**, -)	0.42 (**, **)	**0.58** (***, ***)
US	0.59	0.64	**0.64**	0.55	0.21 (-, -)	**0.35** (−, *)	0.13 (-, -)

Notes: The 'Above-Trend Output Growth' event is defined in the text. See notes to Table 4a.

Table 4c: Evaluation of 'An Acceptable Debt-to-GDP Ratio' Probability Forecasts, 1990 - 2011

73

	Accuracy Rate				Kuipers Score		
	p	VAR	GVAR	GVECM	VAR	GVAR	GVECM
Bel	0.00	1.00	1.00	1.00	1.00 (na, na)	1.00 (na, na)	1.00 (na, na)
Can	0.00	1.00	1.00	1.00	1.00 (na, na)	1.00 (na, na)	1.00 (na, na)
Den	0.68	**0.95**	0.91	0.91	**0.93** (***,-)	0.87 (***, -)	0.87 (***, -)
Fra	0.36	0.77	0.77	**0.82**	0.48 (**, -)	0.48 (**, -)	**0.55** (**, -)
Ger	0.45	0.68	0.73	**0.77**	0.33 (-, ***)	0.43 (*, ***)	**0.53** (**,–)
Ire	0.50	0.86	**0.91**	**0.91**	0.73 (***,-)	**0.82** (na, na)	**0.82** (na, na)
Ita	0.00	1.00	1.00	1.00	1.00 (na, na)	1.00 (na, na)	1.00 (na, na)
Net	0.55	**1.00**	0.95	0.95	**1.00** (na, na)	0.92 (***,***)	0.92 (***,***)
UK	0.77	0.91	0.91	**0.95**	0.60 (***, -)	0.60 (***, -)	**0.80** (***, -)
US	0.23	0.86	0.86	**0.91**	0.54 (***,-)	0.68 (***,-)	**0.74** (na, na)

Notes: The 'Acceptable Debt-GDP Ratio' event is defined in the text. See notes to Table 4a.

Table 4d: Evaluation of 'No Sustainability Risk' Probability Forecasts, 1990 - 2011

	Accuracy Rate				Kuipers Score		
	p	VAR	GVAR	GVECM	VAR	GVAR	GVECM
Bel	0.36	**0.91**	0.91	0.91	**0.86** (***,***)	0.80 (na, na)	0.80 (na, na)
Can	0.41	0.91	0.91	**1.00**	0.78 (***,***)	0.78 (***,***)	**1.00** (na, na)
Den	0.73	**1.00**	**1.00**	**1.00**	**1.00** (na, na)	**1.00** (na, na)	**1.00** (na, na)
Fra	0.32	**1.00**	**1.00**	0.95	**1.00** (na, na)	**1.00** (na, na)	0.93 (***,***)
Ger	0.23	**1.00**	**1.00**	**1.00**	**1.00** (na, na)	**1.00** (na, na)	**1.00** (na, na)
Ire	0.77	**1.00**	0.95	0.95	**1.00** (na, na)	0.94 (***,***)	0.94 (***,***)
Ita	0.23	0.86	0.86	**0.91**	0.54 (***,***)	0.54 (***,***)	**0.74** (na, na)
Net	0.77	**0.95**	**0.95**	0.91	**0.94** (***,***)	**0.94** (***,***)	0.88 (***,***)
UK	0.55	0.95	**1.00**	0.95	0.90 (***,***)	**1.00** (na, na)	0.90 (***,***)
US	0.41	**1.00**	**1.00**	**1.00**	**1.00** (na, na)	**1.00** (na, na)	**1.00** (na, na)

Notes: The 'No Sustainability Risk' event is defined in the text. See notes to Table 4a.

Table 5a: Loss-Based Evaluation of Forecasts (Symmetric in Fiscal Adjustment Errors), 1990 - 2011

		Loss	
	VAR	GVAR	GVECM
Bel	*0.11*	**0.88**	1.00
Can	*0.11*	**0.85**	**0.85**
Den	**0.09**	1.22	1.35
Fra	*0.13*	**0.78**	0.81
Ger	**0.18**	1.14	1.07
Ire	**0.21**	1.02	1.02
Ita	*0.13*	0.80	**0.72**
Net	*0.15*	**0.83**	0.85
UK	**0.13**	1.04	1.06
US	*0.10*	**0.62**	0.68

Table 5a: Loss-Based Evaluation of Forecasts (Asymmetric in Fiscal Adjustment Errors), 1990 - 2011

		Loss	
	VAR	GVAR	GVECM
Bel	*0.14*	0.88	**0.86**
Can	*0.16*	**0.76**	0.79
Den	*0.14*	0.90	**0.88**
Fra	**0.12**	1.10	1.15
Ger	**0.18**	1.21	1.21
Ire	**0.23**	1.02	1.02
Ita	*0.18*	**0.64**	0.67
Net	*0.20*	**0.67**	**0.67**
UK	*0.15*	**0.92**	0.94
US	*0.10*	**0.92**	**0.92**

Notes: For the VAR models, the loss function is defined by expressions (9) and (10) for Tables 5a and 5b respectively. Losses are measured in percentage points of GDP. For the GVAR and GVECM models, the models' losses are expressed as a proportion of the VAR models' losses. The model with the lowest loss is highlighted in bold.

Table 6a: Average Fiscal Adjustment Error in the Decisions Underlying Table 5a

	1990-2011			2008-2011		
	VAR	GVAR	GVECM	VAR	GVAR	GVECM
Bel	-0.18	0.09	0.08	-0.25	0.19	0.13
Can	-0.08	0.01	-0.04	-0.25	0.13	0.00
Den	-0.04	0.07	0.08	-0.25	0.38	0.25
Fra	-0.13	0.06	0.06	-0.38	0.19	0.19
Ger	-0.04	0.08	0.10	0.08	0.31	0.38
Ire	0.11	0.10	0.10	0.33	0.33	0.33
Ita	-0.10	0.04	0.04	-0.29	0.13	0.08
Net	-0.09	0.08	0.07	-0.08	0.17	0.17
UK	0.05	0.09	0.05	-0.13	0.25	0.25
US	-0.09	0.09	0.08	-0.25	0.13	0.13

Table 6b: Average Fiscal Adjustment Error in the Decisions Underlying Table 5b

	1990-2011			2008-2011		
	VAR	GVAR	GVECM	VAR	GVAR	GVECM
Bel	0.00	0.13	0.13	-0.13	0.25	0.25
Can	0.02	0.14	0.12	-0.25	0.13	0.13
Den	0.03	0.14	0.16	-0.13	0.21	0.21
Fra	0.00	0.08	0.10	-0.19	0.25	0.19
Ger	0.05	0.15	0.15	0.09	0.42	0.42
Ire	0.15	0.15	0.15	0.38	0.42	0.42
Ita	-0.08	0.13	0.13	-0.29	0.25	0.25
Net	0.01	0.14	0.14	-0.08	0.38	0.38
UK	0.08	0.15	0.15	-0.13	0.21	0.21
US	0.02	0.14	0.11	-0.25	0.13	0.13

Notes: Forecast error is defined as the fiscal adjustment warranted by the actual outcomes minus the recommended fiscal adjustment based on decisions where the loss function is defined by expressions (9) and (10) for Tables 6a and 6b respectively. Hence the forecast error is negative if the recommended adjustment is more austere than is warranted by the actual outcomes. Errors are measured in percentage points of GDP.

6 References

(Chapter head:)*
Literaturverzeichnis

Baldwin, R and F Giavazzi (2015), "The eurozone crisis: A consensus view of the causes and a few possible solutions.", *VoxEU.org*, September.

Becker, B. and V. Ivashina (2018) "Financial Repression in the European Sovereign Debt Crisis", *Review of Finance*, 22, 1, 83–115.

Blanchard, O., J-C Chouraqui, R. Hagemann and N. Sartor (1990), "The sustainability of fiscal policy: New answers to an old question.", *OECD Economic Studies*, no. 15, 7-36.

Bohn, H. (2005), "The Sustainability of Fiscal Policy in the United States", *CES-IFO Working Paper* no. 1446.

Buiter, W.H. (1985), "A guide to public sector debt and deficits", *Economic Policy,* 1, 1, 13-79.

Celasun, O., X. Debrun, and J.D. Ostry (2006), "Primary Surplus Behavior and Risks to Fiscal Sustainability in Emerging Market Countries: A Fan-Chart Approach.", *IMF Staff Papers*, 53, 3.

Chung, H. and E. Leeper (2007), "What has financed Government debt?", *NBER Working Paper*, no. 13425.

Cottarelli, C. and J. Vinals (2007), "A strategy for re-normalising monetary and fiscal policy in advanced economies.", *IMF Staff Position Note,* 09/22.

Debrun, X. and T. Kinda (2014), "Strengthening Post-Crisis Fiscal Credibility: Fiscal Councils on the Rise — A New Dataset.", *IMF Working Papers 14/58*, International Monetary Fund.

Dungey, M., R. Fry, B. González-Hermosillo and V.L. Martin, (2005), Empirical modelling of contagion: a review of methodologies, *Quantitative finance*, 5 (1), 9-24, 587

Dungey, M. and R. Fry (2009) The identification of fiscal and monetary policy in a structural VAR, *Economic Modelling* 26 (6), 1147-1160

Dungey, M., J. Harveya and V. Volkov, (2018a), The changing international network of sovereign debt and financial institutions, *Journal of International Financial Markets, Institutions and Money*, 6

Dungey,M., M. Luciani and D. Veredas, (2018b), Systemic risk in the US: Interconnectedness as a circuit breaker, *Economic Modelling*, 71, 305-315

Dungey,M. and V.L. Martin, (2007) Unravelling financial market linkages during crises, *Journal of Applied Econometrics* 22 (1), 89-119 210

Dungey, M. and D.R. Osborn (2014), Modelling large open economies with international linkages: the USA and Euro area, *Journal of Applied Econometrics*, 29 (3), 377-393.

Dungey, M. and A. Pagan (2000), A structural VAR model of the Australian economy, *Economic record*, 76 (235), 321-342 268

Dungey, M. and A. Pagan (2009) Extending a SVAR model of the Australian economy, *Economic Record* 85 (268), 1-20

European Commission (2014), “Report on Public Finances in EMU.”, *Technical Report*, European Economy, Directorate-General for Economic and Financial Affairs.

European Union (2016), “Specifications on the Implementation of the Stability and Growth Pact”, *EU Economic and Financial Committee Opinion Piece*, https://ec.europa.eu/info/sites/info/files/code_of_conduct_en1.pdf, July 2016.

Garratt, A., K. Lee, M.H. Pesaran, and Y. Shin (2003), “Forecast Uncertainty in Macroeconometric Modelling: An Application to the UK Economy”, *Journal of the American Statistical Association*, 98, 464, 829-838.

Garratt, A., K. Lee, M.H. Pesaran, and Y. Shin (2012), *Global and National Macroeconometric Modelling: A Long-run Structural Approach*. Oxford University Press.

Garratt, A., K. Lee and K. Shields (2018),.“The Role of Uncertainty, Sentiment and Cross-Country Interactions in G7 Output Dynamics”, *Canadian Journal of Economics*, 51, 391-418.

Giacomini, R. and H. White (2006), “Tests of conditional predictive ability”, *Econometrica*, 74, 6, 1545-1578.

Granger, C.W.J and M. Machina (2006), “Forecasting and Decision Theory”, in G. Elliott, C.W.J. Granger and A.Timmermann (eds.), *Handbook of Economic Forecasting*, 81-98.

Granger, C.W.J. and M.H. Pesaran (2000), "Economic and Statistical Measures of Forecast Accuracy", *Journal of Forecasting*, 19, 537-560.

Horne, J. (1991), “Criteria of external sustainability”, *European Economic Review*, 35, 1559-1574.

Johansen, S. (1988), “Statistical analysis of cointegration vectors”, *Journal of Economic Dynamics and Control* 12, 2-3, 231-254.

Kose, M.A., S. Kurlat, F. Ohnsorge and N. Sugawara (2017), A Cross-Country Database of Fiscal Space, *Policy Research Working Paper 8157*, World Bank, Washington, DC.

OBR (2015), “Fiscal sustainability report.”, *Technical Report*, Office for Budget Responsibility.

Pesaran, M.H. (2007), “A simple panel unit root test in the presence of cross-section dependence.”, *Journal of Applied Econometrics*, 22, 265–312.

Pesaran, M.H., T. Schuermann, and S.M. Weiner (2004), “Modeling Regional Interdependencies Using a Global Error-Correcting Macroeconometric Model.”, *Journal of Business & Economic Statistics*, 22, 129–162.

Pesaran, M.H. and S. Skouras (2002), “Decision-based Methods for Forecast Evaluation”, in M.P. Clements and D.F. Hendry (eds.), *A Companion to Economic Forecasting*, Oxford: Blackwell.

Pesaran, M.H., L.V. Smith, and T. Yamagata (2013), "Panel unit root tests in the presence of a multifactor error structure", *Journal of Econometrics*, 175, 94–115.

Pesaran, M.H. and A. Timmermann (1992), "A Simple Nonparametric Test of Predictive Performance.", *Journal of Business and Economic Statistics*, 10, 461–465.

Pesaran, M. H. and A. Timmermann (2009), "Testing Dependence Among Serially Correlated Multicategory Variables", *Journal of American Statistical Association*, 104, 325–337.

Polito, V. and M. Wickens (2012), "A model-based indicator of the fiscal stance.", *European Economic Review*, 56, 526–551.

Popov, A. and N. Van Horen (2015) "Exporting Sovereign Stress: Evidence from Syndicated Bank Lending during the Euro Area Sovereign Debt Crisis", *Review of Finance*, 19, 5, 1825–1866.

Reinhart, C. M. and K. S. Rogoff (2011), "From Financial Crash to Debt Crisis",. *American Economic Review*, 101, 1676–1706.

Wyplosz, C. (2007), "Debt sustainability assessment: The IMF approach and alternatives.", *Technical Report, Economics Section*, The Graduate Institute of International Studies.