Economic Modelling 64 (2017) 153–171

Contents lists available at ScienceDirect

Economic Modelling

journal homepage: www.elsevier.com/locate/econmod

Measuring the output gap in Switzerland with linear opinion pools☆

Daniel Buncic[a,*], Oliver Müller[b]

[a] Research and Modelling Division, Financial Stability Department, Sveriges Riksbank (Central Bank of Sweden), SE-103 37 Stockholm, Sweden
[b] KOF Swiss Economic Institute, ETH Zürich, Switzerland

ARTICLE INFO

JEL classification:
E17
E37
E52
E58

Keywords:
Linear opinion pool
Forecasting and nowcasting
Inflation
The output gap
Realtime
Data revision

ABSTRACT

We use the recently proposed linear opinion pool methodology of Garratt et al. (2014) to construct real-time output gap estimates for Switzerland over the out-of-sample period from 2003:Q1 to 2015:Q4. The model space consists of a large number of bivariate VAR specifications for the output gap and inflation, with each VAR specification using a different estimate of the output gap, lag order, and structural break information. We find that the linear opinion pool performs rather poorly. Real-time estimates of the output gap are no more accurate than those from some simple benchmark models, no more robust to ex post revisions than the real-time estimates of the individual univariate output gaps, and do not produce more accurate forecasts of inflation. The key driver of '*good*' forecast performance is structural break information. Once the same structural break information is conditioned upon in all prediction models, the gain from averaging over many different pools of models that utilize various output gap estimates or lag structures in the VAR specification is of negligible magnitude.

1. Introduction

Reliable real-time estimates of the output gap — the difference between actual output and unobserved potential output — are essential for timely policy making.[1] Despite their importance in policy environments, real-time estimates of the output gap are characterized by large and profound uncertainty. This fact is extensively documented in the existing empirical macroeconomic literature.[2] A number of alternative estimation methods have been developed with the aim of increasing the reliability of real-time output gap estimates (see, for instance, the general prediction pool framework of Geweke (2010), Geweke and Amisano (2011)).

In this study, we use the recently developed linear opinion pool methodology of Garratt et al. (2014) to construct real-time output gap estimates for Switzerland. The main idea behind the linear opinion pool is to utilize various univariate estimates of the output gap to compute ensemble forecasts for the output gap and inflation, using a bivariate vector autoregressive (VAR) specification for the output gap and inflation as the prediction model. The combination weights for the forecasts in the linear opinion pool are determined from the predictive densities of the individual VAR specifications. Following the approach of Camba and Rodriguez (2003), we then proceed to assess the '*goodness*' of the linear opinion pool's real-time output gap estimates by defining two main criteria. The first is that the output gap estimates should provide '*good*' forecasts for inflation. The second is that ex post revisions to the estimates of the output gap should be '*small*'.

Using an out-of-sample period from 2003:Q1 to 2015:Q4, we find that the linear opinion pool's ensemble forecasting approach performs rather poorly in our application to Swiss data. That is, the model fails to produce real-time output gap estimates that are better calibrated (or more accurate) than those from some simple univariate benchmark time series models. We find further that the key determinant of '*good*' forecast performance in any of the models that we employ in our analysis is the use of structural break information. Once the same structural break information is conditioned upon in all prediction models, the gain from averaging over many different pools of models

☆ We are grateful for comments from Francesco Ravazzolo, the editor Prof. Sushanta Mallick, as well as two anonymous referees that have significantly helped to improve the paper.

* Corresponding author.
E-mail addresses: daniel.buncic@riksbank.se (D. Buncic), omueller@kof.ethz.ch (O. Müller).

[1] Output gap estimates are routinely used in macroeconomic policy models at central banks (see, for instance, Smets and Wouters, 2003, 2007; Buncic and Melecky, 2008; Zhang and Murasawa, 2011; Komlan, 2013; Buncic and Lentner, 2016). Real-time estimates of the output gap are also becoming increasingly important for the monitoring of financial stability (see Buncic and Melecky (2013), or Buncic and Melecky (2014), Basel (2011) in the context of determining excessive credit growth in an economy).

[2] The seminal paper by Orphanides and van Norden (2002) is one of the first to show that many output gap estimation methods for the US are unreliable in real-time. Marcellino et al. (2006) come to a similar conclusion using data for the Euro area. In a fiscal policy setting, Ley and Misch (2014) find that output gap revisions significantly undermine the reliability of real-time estimates of fiscal balances in a large cross-section of 175 countries. Several other studies have found that inaccurate real-time output gap estimates can introduce a pro-cyclicality bias into policy decisions (see, for instance, Orphanides, 2001; Smets, 2002; Grigoli et al., 2015).

http://dx.doi.org/10.1016/j.econmod.2017.03.007
Received 30 August 2016; Received in revised form 28 December 2016; Accepted 8 March 2017

that utilize various output gap estimates and/or different lag orders in the bivariate VAR specification is negligible. We confirm that these results also hold for the US data analysed in Garratt et al. (2014) in our independent replication of their study. Another key driver behind the poor performance of the linear opinion pool for Swiss data, and which is absent from US data, is the magnitude of the revisions to the GDP price deflator series that is needed to construct the inflation measure. These revisions can be large, even as many as eight quarters after the initial release. This leads to further increases in the uncertainty about the level of inflation.

In our analysis we find also that the real-time estimate of the output gap provided by the linear opinion pool over the evaluation period is no more robust to ex post revisions than the real-time estimates of the individual univariate estimates of the output gap. Our overall findings are thus in line with a large body of existing literature which fails to find output gap estimates to have predictive power for inflation, or pooling of various real-time output gap estimates to improve the robustness to ex post revisions (see, for example, Orphanides and van Norden, 2005; Marcellino and Musso, 2011; Ince and Papell, 2013).

The rest of the paper is structured as follows. Section 2 describes the different VAR specifications that constitute the model space of the ensemble forecast, and how the linear opinion pool is constructed from the individual forecast densities. Section 3 describes the data and the VAR model space used in our application to Swiss data. The empirical results are presented in Section 4. In Section 5, we discuss our findings and provide a robustness analysis using US data. Finally, Section 6 concludes the study. Additional results are reported in an Appendix A.

2. Methodology

This section describes in more detail the linear opinion pool modelling methodology of Garratt et al. (2014) that we implement, defining initially the bivariate vector autoregressive (VAR) specification for the inflation and output pair that is used. A detailed outline of how the individual forecasts are constructed and aggregated follows this description.

2.1. Vector autoregressive specifications

A simple Phillips Curve type relationship between inflation and the output gap is frequently modelled as a bivariate VAR(p) process of the form (see for instance Garratt et al., 2014; Tiwari et al., 2014):

$$\mathbf{A}^j(\mathrm{L})\mathbf{X}_t^j = \mathbf{C}^j + \boldsymbol{\epsilon}_t^j, \tag{1}$$

where $\mathbf{X}_t^j = [\pi_t; y_t^j]$ is a (2×1) dimensional vector containing inflation π_t and the j-th output gap estimate y_t^j, $\forall\, j = 1, \ldots, J$. The corresponding (2×1) dimensional intercept vector is denoted by $\mathbf{C}^j$, $\mathbf{A}^j(\mathrm{L})$ is a matrix polynomial in the lag operator L of order p, and $\boldsymbol{\epsilon}_t^j$ is a (2×1) vector of error terms with mean vector of zero and variance/covariance matrix $\boldsymbol{\Omega}^j$, which is assumed to be normal distributed. The j superscript in the notation signifies that the VAR(p) model was computed using the j-th output gap estimate. The j-th output gap estimate y_t^j is defined as the difference between the logarithms of observed output and the j-th estimate of the unobserved trend (permanent) component of output (Section 3 describes in more detail the construction of the output gaps).

Let N denote the number of different VAR models in our model space. Using J different output gap estimates, this means that we will have N=J different VAR specifications to aggregate from. We further account for the possibility of a changing relationship between inflation and the output gap, by allowing for structural breaks, that is, changes in the conditional mean and variance of the series. In our set-up, every structural break defines a new model. With K unique structural break dates, the number of possible models increases from N=J to $N = J \times K$. Finally, we also allow the number of lags in the VAR(p) specification to vary between 1 and P, with P being the maximum lag length, so that the total number of models is then $N = J \times K \times P$, producing N different forecasts for inflation and the output gap from the VAR specification in (1). Note here that we purposefully do not want to make any assumptions about the true structure of the data generating process for inflation and the output gap. Moreover, following the reasoning of Garratt et al. (2014), we work with the proposed model space because of its flexibility and simplicity. Utilizing a variety of different VAR specifications allows us to bypass the problem of defining a single output gap measure by considering all commonly used measures, and combining those that produce the best forecasts for inflation.

2.2. Constructing the ensemble forecasts

We use a linear opinion pool (LOP) to construct ensemble forecast densities for inflation and the output gap. The opinion pool approach has a long history in management science, where the focus is on combining the evidence supplied by many experts to a decision maker (wisdom of crowds analogy). Since opinion pools only require the out-of-sample forecasts of every expert (in our case the N VAR specifications) as inputs, they are particularly useful when combining survey information (see also Wallis (2005)).[3] Following Garratt et al. (2014), we construct the forecast density for the output gap by combining the forecast densities of the N individual VAR specifications with the weights determined by the out-of-sample forecast performance of inflation. The h-step ahead ensemble forecast density for inflation, given the N different VAR specifications, is then obtained from the following linear opinion pool:

$$p_\pi(\pi_{\tau+h}) = \sum_{j=1}^{N} p_{\pi,j}(\pi_{\tau+h}|\mathcal{I}_{j,\tau})w_{j,h,\tau}, \quad \tau = \underline{\tau}, \ldots, \overline{\tau} \tag{2}$$

where $p_{\pi,j}(\pi_{\tau+h}|\mathcal{I}_{j,\tau})$ is the h-step ahead forecast density from model j for inflation $\pi_{\tau+h}$ given the information set $\mathcal{I}_{j,\tau}$. The information set $\mathcal{I}_{j,\tau}$ consists of the data vintages up to and including period τ. Note that in the notation in (2) we use τ to label the vintage of the data, that is, data up to and including τ, regardless of the publication lag. Thus, when the real-time data are released with a one period lag, as it is usually the case, the one-step ahead forecast (h=1) is the nowcast. The weights $w_{j,h,\tau}$ are by definition non-negative and sum up to unity for every forecast horizon h and period τ.

We use a logarithmic scoring rule to construct the aggregation weights $w_{j,h,\tau}$, which gives a high score (or weight) to a density forecast whenever a high probability of the realized inflation value is obtained. The logarithmic score of the density forecast for model j is defined as $\ln p_{\pi,j}(\pi_{t+h}^*|\mathcal{I}_{j,t})$ and is evaluated at the realized inflation value $\pi_{\tau+h}^*$. The weight $w_{j,h,\tau}$ for the h-step ahead forecast density of model j at τ is calculated as:

$$w_{j,h,\tau} = \frac{\exp\left\{\sum_{t=\underline{\tau}-h-r-\kappa+1}^{\tau-h-r} \ln p_{\pi,j}\left(\pi_{t+h}^*|\mathcal{I}_{j,t}\right)\right\}}{\sum_{j=1}^{N}\exp\left\{\sum_{t=\underline{\tau}-h-r-\kappa+1}^{\tau-h-r} \ln p_{\pi,j}\left(\pi_{t+h}^*|\mathcal{I}_{j,t}\right)\right\}}, \quad \tau = \underline{\tau}, \ldots, \overline{\tau} \tag{3}$$

where κ is the length of the minimal training period used to initialize the weights and r is the revision horizon. Inflation measurements are usually subject to revisions in later periods. Revised inflation measurements are considered to be more accurate and thus constitute a better target for forecasting. Given a revision horizon r, the realized inflation

[3] The primary motivation for using an opinion pool for the real-time measurement of the output gap is the view that policymakers typically face at central banks. Policymakers generally discuss the forecast densities supplied by the various experts on their staff, who typically utilize linear (or linearised) and Gaussian models, which the policymakers inherently believe to be false. In our context, this translates to the policymaker believing the various output gap measures to differ from the '*true*' output gap by more than the conventional white noise measurement error and that the true output gap is never observed, even ex post.