Now-casting the Italian budget deficit: a mixed frequency BVAR approach[1]

Jacopo Cimadomo (European Central Bank)

Domenico Giannone (Federal Reserve Bank of New York and CEPR)

Michele Lenza (European Central Bank and ECARES-ULB)

December 2015

Abstract. Budget balance data are only available at quarterly frequency and their release is particularly delayed coming, in general, later than national quarterly accounts. However, monthly data are available on cash flows from government, with a very limited delay (two days after the end of the reference month, in the case of Italy). Though very timely, due to a different accounting methodology compared to the one for assessing the budget balance, monthly cash flows are a noisy indicator of the budget balance. In order to extract information on the budget balance while, at the same time, discounting the noisy content of monthly cash-flows, this paper proposes a Bayesian Mixed VAR approach to now-cast the Italian budget balance.

1) Introduction

The deficit to GDP ratio is a synthetic indicator of state of public finances in one country, and it has a core role in the surveillance process in the context of the EU fiscal framework. Timely monitoring the tendency of such ratio is of fundamental importance, especially for countries that exceeded the 3% of GDP threshold and are subject to an "Excessive Deficit Procedure" (EDP). This paper describes and evaluates a new methodology to implement this task. In this paper, we focus on Italian data as an illustration, but the methodology we propose is more general and can be applied to other cases, as well.

Budget revenues and expenditures, the two constituencies of the budget balance, are generally quarterly variables and are released with a considerable delay. In Italy, these variables are released only on the first business day of the fourth month after the end of the reference quarter. For example, the budget balance for the fourth quarter of 2012 will be only released at the beginning of April 2013. However, two business days after the end of, say, month *tm*, the Italian Treasury publishes its cash flow in month *tm*. The sum of the cash flows in the quarter do not generally exactly sum to the budget balance of that quarter, due to different accounting methods. In fact, the ESA95 data on the budget balance, which are those relevant for fiscal surveillance, are not reported in terms of cash flows but are rather characterized by the accrual recording method. However, the cash flows should still reflect a large part of the items included in the evaluation of the Italian budget balance data relevant for fiscal surveillance.

This paper proposes a methodology, based on a mixed frequency vector autoregressive model, which aims to reap the benefits of the timeliness in the releases of monthly cash data while, at the same time, trying to filter out the noise in the relationship with quarterly budget balance data induced by the different accounting procedures.

---

[1] Preliminary version, do not quote without authors' permission. Domenico Giannone gratefully acknowledges the Fiscal Policies Division of the ECB for its hospitality. The opinions expressed herein are those of the authors and do not necessarily reflect those of the ECB or the Eurosystem.

Section two provides a brief literature review, section three describes the data, the now-casting problem that we set to solve and the estimation methodology. Section four reports the results, while section five summarizes the main point of the paper and discusses the next steps in the project.

2) **Literature review** (*to be completed*)

The fiscal forecasting literature is quite limited (for a survey, see Leal et al., 2008). Few papers in this literature have highlighted that – while accrual data on government deficits are only available with a relatively long time lag – monthly or quarterly intra-annual data are available with much shorter time lags, and can be used to derive accurate forecasts for end-of-year fiscal outcomes. (see e.g. Pérez, 2007; Pedregal and Pérez, 2008; Onorante et al., 2008).

In particular, exploiting a Mixed Data Sampling approach (MiDaS), Asimakopoulos et al. (2012) assess the news content of quarterly fiscal data releases and their implications for the annual outturn of those series. Focusing on a sample of EU countries, they show that quarterly information is indeed very important to estimate annual outcomes.

Hughes Hallet et al (2010) focus on monthly cash data. They assess the relevance of such data as instruments for constructing early warnings indicators for future deficit deviating from targets. They also examine and compare two different strategies for correcting excessive.

In this paper, we adopt a different approach: we use a mixed-frequency Bayesian VAR model to extract information from monthly data that can be useful for the estimation of annual deficit outturns.

3) **Data, now-casting problem and methodology**

3.1 *Data*

As a convention, in this paper, t indicates time in quarters while tm time in months and ta time in years.

In this application, although our methodology can be extended to consider a very large set of variables, we only consider those variables that are instrumental to compute the variable we wish to now-cast, i.e. the Italian annual budget balance to GDP ratio. Our data on Italian expenditures (E_t) and revenues (R_t) are quarterly and range from 1999Q1 to 2011Q2 (50 observations); rather than including both variables, in our model, we include their difference (D_t), to keep consistency with the cash data, which are timely available only in the form of balance (while the sub-components of the cash balance become available only a month and a half after the end of the reference month).

In order to be able to compute future budget balance ratios, we also include real GDP (Y_t) and the GDP deflator (P_t) in our model, which are also quarterly variables, in the sample 1999Q1-2011Q2.

Finally, we include the monthly cash balance (D_{tm}) of the Italian state, in the sample January 1999 to October 2011.

3.2 *The now-casting problem*

Our target variable is the annual budget balance to GDP ratio (b_{ta}) of the Italian state in each specific year *ta*, i.e.

$$b_{ta} = \frac{\sum_{t=ta.Q1}^{ta.Q4} D_t}{\sum_{t=ta.Q1}^{ta.Q4} Y_t * P_t}$$

In this application, we focus on *monthly* now-casts of b_{ta}[2], i. e. the assessment of the budget balance to GDP ratio for the whole year *ta* conducted in each of the twelve months of the same year. Notice that this is not a trivial problem, given that, especially in the first months of the year, it implies to forecast the path of the budget balance to GDP ratio (hence, the path of the difference between revenues and expenditures, GDP and the GDP deflator) several months ahead.

In order to provide a realistic assessment of the challenges faced in now-casting the state of Italian public finances, we should also take into account the real time data availability faced by pratictioners.

At this stage, we cannot issues related to data revisions, given that we have only ex-post revised data, for the time being. However, we fully address the issue of the end-sample data imbalance caused by the staggered nature of data releases. In order to mimic the data availability at the time of the now-cast production, which we assume to be the 15th of each month, we have reconstructed the data availability at the end of the sample that a praticitioner would face in each of the twelve months of each year. Table 1 reports the data availability on the 15th of each month.

Table 1. Data availability for Italy in the dates of the now-cast production

Date of now-cast	GDP	GDP Deflator	Budget balance	Cash data
15-Jan	ta-1.Q3	ta-1.Q3	ta-1.Q3	ta-1.December
15-Feb	ta-1.Q3	ta-1.Q3	ta-1.Q3	ta.January
15-Mar	ta-1.Q4	ta-1.Q4	ta-1.Q3	ta.February
15-Apr	ta-1.Q4	ta-1.Q4	ta-1.Q4	ta.March
15-May	ta-1.Q4	ta-1.Q4	ta-1.Q4	ta.April
15-Jun	ta.Q1	ta.Q1	ta-1.Q4	ta.May
15-Jul	ta.Q1	ta.Q1	ta.Q1	ta.June
15-Aug	ta.Q1	ta.Q1	ta.Q1	ta.July
15-Sep	ta.Q2	ta.Q2	ta.Q1	ta.August
15-Oct	ta.Q2	ta.Q2	ta.Q2	ta.September
15-Nov	ta.Q2	ta.Q2	ta.Q2	ta.October
15-Dec	ta.Q3	ta.Q3	ta.Q2	ta.November

Table 1 reflects both the different timeliness of the variables (different dates of data releases) and their different sample frequency.

[2] The modelling framework we employ allows also to produce forecasts and backcasts of the budget position.

In column 2 and 3, we report the available releases of GDP and GDP deflator at each mid-month now-casting round. National accounts are released with a quarterly frequency and around mid-month, in the third month after the end of the reference quarter. Hence, say, the now-casts of the budget balance to GDP ratio produced in January and February are based on GDP and GDP deflators data until the third quarter of the previous year (ta-1). At mid-march, instead, the release of the fourth quarter for the previous year becomes available. Successive national account releases follow the same path just described discussed for the first quarter.

Government accounts are released with a few weeks delay compared to Quarterly National accounts, generally at the beginning of the fourth month after the end of the reference quarter. Hence, differently from the case of GDP and the GDP deflator, even in March the now-casts are still based on budget balance data only until Q3 of the previous year and the fourth quarter release will only be factored in the now-casts from April onward. Again, the same pattern of releases then follows in the successive months.

Cash data, instead, are released with monthly frequency and for, say, month *tm*, right at the beginning of the successive month (second business day after the end of the month). Hence, at the date of each now-cast, we have cash data releases ranging until the previous month.

3.3 *Estimation and now-casting methodology*

The now-casting problem defined above requires the solution of two issues of missing data.

First, the variables are sampled with different frequency, quarterly and monthly. We assume that quarterly variables are monthly variables with missing observations in the first two months of the quarter.

Second, due to the staggered nature of data releases highlighted in table 1, several observations of the quarterly variables are missing at the end of the sample. This section briefly sketches the methodology we use in order to address these issues.

We assume that the *levels* of our N (=4) variables (collected in the N-dimensional vector X_{tm}) are described by the following monthly vector autoregressive process with p (=13) lags:

$$X_{tm} = A_0 + A_1 X_{tm} + \ldots + A_p X_{tm-p} + e_{tm}$$

where A_p is the $N*N$ matrix collecting the coefficients of the p-th lag and e_{tm} is a normally distributed multivariate white noise with covariance matrix Σ.

The choice of accounting for rich dynamics (*p = 13 lags*) is motivated by two main considerations. First, we want a general and flexible model which does not a-priori constraints the dynamic interrelationships among our variables. Second, the data are not seasonally adjusted and this dynamic specification is able to account for the seasonal fluctuations in the variables.

The rich dynamics we want to allow for in our VAR model imply that we face an issue of over-fitting, owing to the large number of parameters (the so-called “curse of dimensionality”). We address this issue by shrinking the model’s coefficients toward those of the naïve and parsimonious random walk with drift model, $X_{i,tm} = \delta_i + X_{i,tm-1} + u_{i,tm}$. De Mol et al. (2008) and Banbura et al. (2010) have shown that this approach reduces estimation uncertainty without introducing substantial bias. This is achieved thanks to the tendency for macroeconomic time series to co-move over the business cycle, which creates scope for the data to point "massively" in the same direction against a naïve prior model that does not allow for any dynamic interaction. The resulting model offers a parsimonious but reliable estimate of the complex dynamic interactions among the macro, monetary and financial variables included in the data set.

More specifically, we use a Normal-Inverted Wishart prior centred on a random walk model. For Σ, the covariance matrix of the residuals, we use an inverted Wishart with scale parameter given by a diagonal matrix Ψ and d=N+2 degrees of freedom. This is the minimum number of degrees of freedom that guarantees the existence of the prior mean of Σ, which is equal to $\Psi/(d-N-1)=\Psi$. For the constant A0 term, we use a flat prior. For the autoregressive coefficients ($A_1 \dots A_p$), we use the Minnesota prior, as originally proposed by Litterman (1980).

As regards the Minnesota prior, conditional on the covariance matrix of the residuals, the prior distribution of the autoregressive coefficients is normal with the following means and variances:

$$E(A_1) = I_N \text{ while } E(A_2) = \dots = E(A_p) = 0_{N,N}$$

$$Cov[(A_s)_{ij},(A_r)_{hm}|\Sigma] = \lambda^2\Sigma_{ih}/(s^2\Psi_{ii}) \text{ if } m=j \text{ and } r=s, \text{ zero otherwise.}$$

Notice that the variance of this prior distributions decays with the lag, and that coefficients associated with the same variables and lags in different equations are allowed to be correlated. The key hyperparameter is λ, which controls the scale of all the prior variances and covariances, and effectively determines the overall tightness of this prior. For λ = 0 the posterior equals the prior and the data do not influence the estimates. If $\lambda \to \infty$, on the other hand, posterior expectations coincide with the Ordinary Least Squares (OLS) estimates. The factor $1/s^2$ is the rate at which the prior variance decreases with increasing lag length and Σ_{ii}/Ψ_{jj} accounts for the different scale and variability of the data.

Summing up, the setting of these priors depends on the hyperparameter λ, which reflects the informativeness of the prior distribution for the model's coefficients. This parameter is usually set on the basis of subjective considerations or rules of thumb. For the sake of simplicity, at this stage, we set the value of this hyperparameter to 0.2, as it suggested in Sims and Zha (1998).

If we did not face the issue of missing data, the Bayes rule would allow us to easily draw parameters from the posterior distributions implied by the likelihood and the prior set-up just described. Then, the algorithm to produce conditional forecasts developed in Banbura et al. (2012) based on the simulation smoother of Carter and Kohn (1994), could be employed in order to produce the out-of-sample forecasts of the budget balance and nominal GDP. Notice that the need of an algorithm to produce conditional forecasts is due to the end-of-sample imbalance in our panel caused by the staggered data releases. The idea here is that we treat more timely data releases as future "conditions" on which we condition the other forecasts.

However, as described above, we have to tackle also a further issue of missing data in this set-up, due to the mixed frequency of the variables. We tackle the issue of missing data by setting up a recursive procedure that, first, balances the database by providing a draw of the missing data *conditional* on a draw from the posterior of the model parameters and, then, provides another draw of the parameters conditional on the previous draw of the variables.

A schematic way of representing our recursive algorithm for the panel available in month t_m and for a forecast horizon *h*, is the following.

1) Initialization: $\mathbf{X(0)_{tm}}$ is obtained by interpolating the unbalanced panel by means of standard univariate non-parametric interpolation techniques.

2) First draw of the parameters from their posterior distribution, conditional on initialization of the variables: $\mathbf{A(1)_{0} \ldots A(1)_{p}}$.

3) First draw of the past, present and future of the variables from the distribution of their conditional expectation, conditional on $A(1)_0 \ldots A(1)_p$: $\mathbf{X(1)_0 \ldots X(1)_{tm} \ldots X(1)_{tm+h}}$ by means of the simulation smoother of Carter and Kohn (1994).

4) Second draw of parameters from their posterior distribution, conditional on previous draw of the variables conditional on $X(1)_0 \ldots X(1)_{tm}$: $\mathbf{A(2)_0 \ldots A(2)_p}$ **.**

5) Second draw of the past, present and future of the variables from the distribution of their conditional expectation, conditional on $A(2)_0 \ldots A(2)_p$: $\mathbf{X(2)_0 \ldots X(2)_{tm} \ldots X(2)_{tm+h}}$ by means of the simulation smoother of Carter and Kohn (1994).

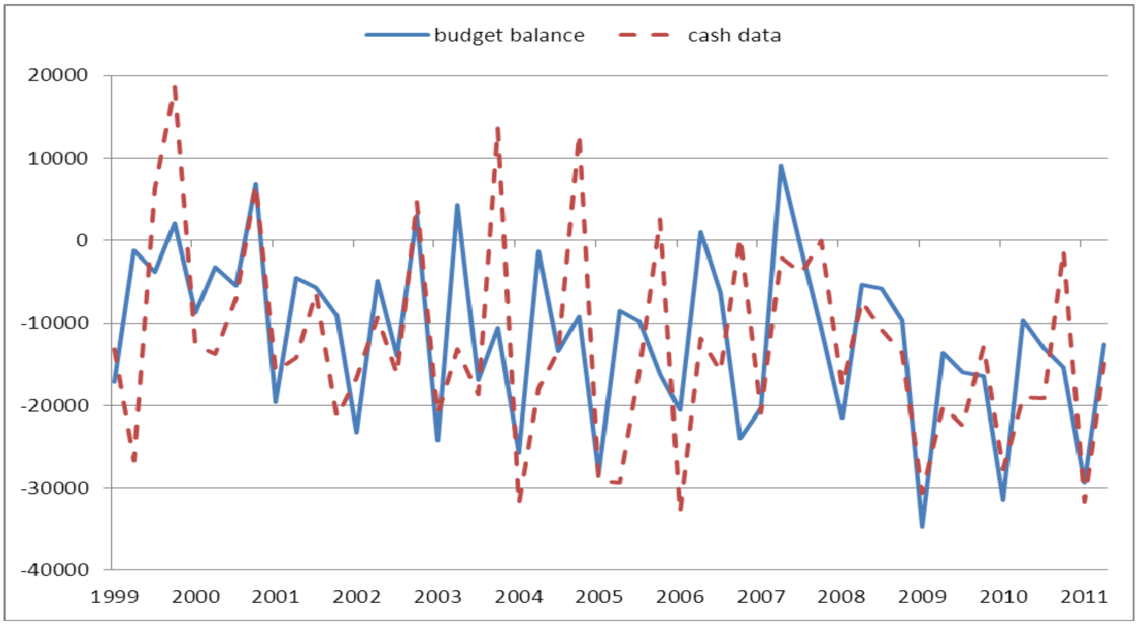
6) Iterate 4 and 5 *M* times.

4) Empirical results

4.1 *Descriptive analysis of the relationship between cash flows and budget balance data*

Chart 1 offers a visual impression of the relationship between budget balance data (revenues minus expenditures) and cash flow data for the Italian economy. In order to plot the data on the same time scale, we derive quarterly cash data by summing the three consecutive monthly values in each quarter.

Chart 1: Quarterly budget balance and cash data



Note: the sample is from 199Q1 to 2011Q2. Units in euro

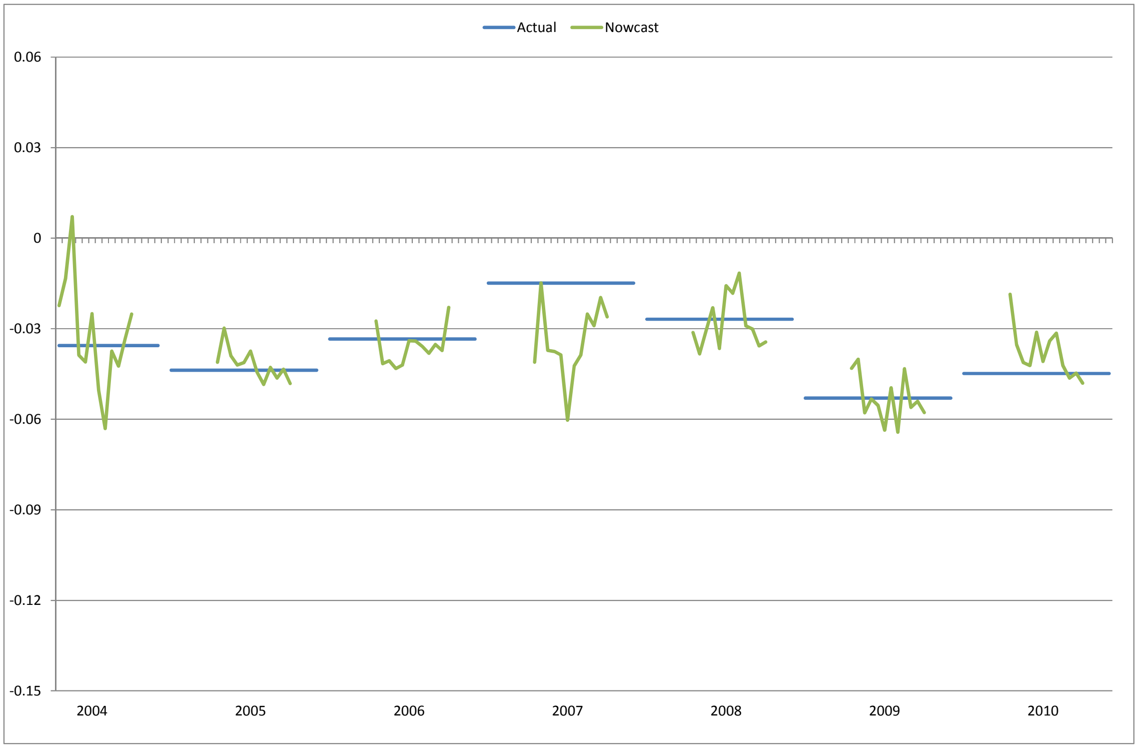
Chart 1 shows that the medium-low frequency developments in the data on cash flows are definitely in line with the medium-low frequency in budget balance data. Hence, the very timely releases of cash data can be a very important asset in order to predict the budget balance. However, cash data are also quite noisier than budget balance data and modelling devices should be used in order to appropriately filter out such noise without eliminating too much of their informative content.

4.2 *Nowcasts of the annual budget balance to GDP ratio for the period 2004-2010*

In this sub-section, we report the results of a rough analysis of accuracy of now-casts (i.e. evaluation of budget balance ratios conducted in the twelve months of the reference year) of the budget balance to GDP ratio in Italy, in the years from 2004 to 2010. We limit our analysis to this sample because our observations for the budget balance start in 1999 and we use roughly a third of the sample in order to estimate the model for the first now-casts of 2004.

Chart 2 reports the results. The green line indicates the twelve now-casts while the blue straight lines indicate the outcomes for the budget balance ratio in a specific year. We plot point forecasts, which are given by the median of the predictive distribution produced by our model.

Chart 2: Forecasts, now-casts and back-casts of budget balance to GDP ratio



Note: The budget balance ratios are expressed in percentage terms. Point forecasts are given by the median of the predictive distribution. The green line indicates the twelve nowcasts successively produced every year (from left to right the nowcast factor in increasingly more information) while the blue straight lines indicates the outcomes for the budget balance to GDP ratio in a particular year.

Chart 2 shows that, in spite of some volatility, our model provides a quite accurate account of the annual budget balance to GDP ratio. In particular, nowcasts produced around mid-year, i.e. now-casts produced about nine months before the release of the annual budget balance ratio and only based on the knowledge of the budget balance in the first quarter of the current year, are already pretty close to the outcomes.

It is also generally the case that further releases of cash data improve the quality of the now-casts, pushing them closer to the final outcomes.

This very informal evaluation of the model performance reveals that, in spite of the noisy nature of cash flow data, our model is able to extract information from the latter in order to inform our view on the state of public finances in Italy.

5) Conclusions and ongoing work

This paper describes a methodology to extract information from monthly cash data in order to now-cast the annual budget balance ratio to GDP in Italy.

The methodology we propose is able to handle both staggered data releases and missing data in the estimation sample in a unified framework and its outcome is the predictive distribution of the budget balance ratio.

Our empirical application, in this paper, is on Italian data. We provide quite an accurate account of the Italian budget balance to GDP ratio, which allows us to conclude that our model is able to successfully extract information from the noisy cash flow data.

Ongoing work is devoted to:

- extend the estimation sample making use of available data on the annual budget balance before 1999;
- impose more sophisticated priors on the sum-of-coefficients, which generally help to improve the forecast accuracy in Bayesian vector autoregressive models;
- equip the model with state-of-the-art techniques in order to select the informativeness of the prior distributions;
- evaluation of density forecasts; extend evaluation also to forecasts and back-casts;
- extension of the cross-section of data in order to improve forecast accuracy (for example, including monthly surveys to better forecast GDP) and extend the possible applications of the model (which, with a suitable variable choice, can be also used in order to provide scenario analysis).

References

Asimakopoulos, A., Paredes, J. and T. Warmedinger (2012): “Forecasting fiscal time series using Mixed Frequency Data Sampling”, European Central bank, mimeo.

Banbura, M., Giannone, D. and M. Lenza (2012): “Conditional forecasts in a large bayesian euro area VAR”, ECB mimeo.

Banbura, M., Giannone, D. and L. Reichlin (2010): "Large Bayesian vector auto regressions," Journal of Applied Econometrics, John Wiley & Sons, Ltd., vol. 25(1), pages 71-92.

Carter, C. and R. Kohn (1994): “On Gibbs sampling for state space models”, Biometrika, 81, 541-553

De Mol, C., Giannone, D. and L. Reichlin (2008): "Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components?," Journal of Econometrics, Elsevier, vol. 146(2), pages 318-328, October.

Giannone, D., M. Lenza and G. Primiceri (2012): “Prior selection for vector autoregressions”, NBER WP 18467.

Hughes Hallett, A., Kuhn, M. and T. Warmedinger (2010) “The gains from early intervention in Europe: Fiscal surveillance and fiscal planning using cash data”, ECB Working paper Series, 1220, July 2010.

Leal, T., Pérez, J., Tujula, M., Vidal J-P, 2008. "Fiscal Forecasting: Lessons from the Literature and Challenges", Fiscal Studies, Institute for Fiscal Studies, vol. 29(3), pages 347-386, 09.

Litterman, D. (1980): “A Bayesian Procedure for Forecasting with Vector Autoregression.,” Working paper, Massachusetts Institute of Technology, Department of Economics.

Onorante, Luca, Diego Pedregal, Javier Pérez, and Sara Signorini (2008). “The usefulness of infra-annual government cash data for fiscal forecasting in the euro area”. ECB Working paper Series, 901, May 2008.

Pedregal, Diego and Javier Pérez (2008). “Should quarterly government finance statistics be used for fiscal surveillance in Europe?” ECB Working paper Series, 937, September 2008.

Pérez, Javier (2007), “Leading indicators for euro area government deficits”. International Journal of Forecasting, 23: 259 – 275.

Sims, C. A. and T. Zha (1998): “Bayesian Methods for Dynamic Multivariate Models,” International Economic Review, 39(4), 949–68.