Erasmus University Rotterdam
Erasmus School of Economics
Bachelor Thesis Double Bachelor BSc[2] in Econometrics and Economics

The Density of Macroeconomic Data: The Effect of Incorporating Macroeconomic Expertise in Prior Selection

Bas Beekman (474613)

Supervisor:	dr. Naghi, A.A.
Second assessor:	van Meer, S.
Date final version:	2nd July 2023

Abstract

We extend the model used by Giannone et al. (2021a) to evaluate the sparsity or density of economic data. We evaluate two macroeconomic data sets and test the robustness of the conclusions of Giannone et al. (2021a) regarding the sparsity or density of these data sets when macroeconomic domain expertise is incorporated in prior elicitation. Particularly, using a t-distribution instead of normal distribution leads to much sparser models, casting doubt on the conclusions of Giannone et al. (2021a) about the suggested density of macroeconomic data sets.

1 Introduction

In the era of 'Big Data', increasingly more and larger data sets are available, more computing power is widely obtainable and accurate forecasting becomes of even greater importance. Therefore, the need for methods which can deal with these types of data sets, meaning data sets with large numbers of variables relative to the number of observations, has greatly increased. Two examples of such data sets are (i) data sets on cross-country GDP growth as used in Barro & Lee (1994) or Fernandez et al. (2001) among many others or (ii) data sets which try to explain classic macroeconomic variables like GDP growth or Industrial production in a single country using large numbers of explanatory variables as in Stock & Watson (2002) using the data obtained from Fred-MD[1]. Large numbers of macroeconomic variables across all fields of macroeconomics and over many different time periods are measured throughout the world. However, these variables are usually only reported on a monthly, quarterly or even yearly basis. This, in combination with the large number of possible explanatory variables in macroeconomic data sets, leads to an issue which is commonly referred to as high-dimensionality, which means that the number of regressors is relatively large compared to the number of observations. For high-dimensional data, classical regression methods like OLS perform extremely poorly, necessitating the development of other methods which are able to deal with this type of data adequately. A class of methods which is particularly suited to deal with high-dimensional data is the class of Bayesian econometric methods, which has led to Bayesian econometric methods becoming growingly popular in the field of empirical macroeconomics (Koop (2017)).

An ongoing debate within the Bayesian econometric academic community, started by Giannone et al. (2021a), discusses the question whether data sets in the economic fields of finance, microeconomics and macroeconomics are better described by so-called dense models, which include many predictors in the model and heavily shrink these or by so-called sparse models, which select fewer variables and apply less shrinkage to these variables (Chernozhukov et al. (2017)). Giannone et al. (2021a) state that there is an 'Illusion of Sparsity', meaning that while many methods used regularly in Bayesian econometrics assume some degree of sparsity of the data a priori, most economic data sets actually tend more towards density. In this paper the work of Giannone et al. (2021a) will be replicated specifically for two macroeconomic data sets [2] to check whether the same conclusions regarding the sparsity or density of these data sets are reached. Then, we will analyze whether incorporating insights from the fields of macroeconomics and macroeconometrics into the general framework of Giannone et al. (2021a) impacts the conclusions drawn regarding the sparsity or density of these macroeconomic data sets.

To elaborate on this, we will first go into more detail on the field of Bayesian macroeconometrics. Following Del Negro (2011), Bayesian macroeconometric methods can be roughly divided into three categories, which correspond to Bayesian approaches widely used in the field of empirical macroeconometrics. These are Bayesian Vector Autoregressions (BVAR), Dynamic Stochastic Equilibruim Models (DGSE) and Dynamic Factor Models (DFM). Below, we will go into further detail on the general intuition of BVAR models and DGSE models and specifically

[1] see McCracken & Ng (2016)

[2] These are the data sets used in Stock & Watson (2002) and Barro & Lee (1994), which are elaborated on in Section 2.

on the ways in which prior information is incorporated into these types of models.

Starting with the DGES models, Del Negro (2011) states that DGSE models are generally defined as a broad class of dynamic macroeconomic models, spanning the basic neo-classical growth model of King et al. (1988), but also the monetary model developed by Christiano et al. (2005). A common feature in these models is that decision rules are derived from assumptions on several economic factors by solving intertemporal optimization problems, where uncertainty in the productivity of factors is created through exogenous stochastic processes. Conditioning on distributional assumptions on these exogenous shocks, a joint probability distribution function can be derived for the endogenous variables. This likelihood can then be used in a Bayesian model framework to obtain a posterior distribution for the structural model parameters. Within the literature on Bayesian DGSE models informative priors are a standard practice. For example, a good way to include information which is not reflected in the joint likelihood function is to incorporate this information into a prior distribution. Del Negro (2011) identifies three possible sources from which such a prior distribution could be elicited, namely (i) information from macroeconomic series other than the evaluated variables over the same time-period, (ii) micro-level observations which contain information on the macro-level variables and (iii) macroeconomic data from years before the sample period. It is of course crucial that this information is independent from the dependent variable.

The use of the so-called Vector Autoregression (VAR) in macroeconomics began when Sims (1980) proposed that VARs should be used for empirical macroeconomic analysis rather than the large-scale macroeconometric models inherited from the decades before. According to Sims (1980), these models imposed *incredible* restrictions, which were not consistent with the actual macroeconomic reality. He also suggested that Bayesian methods could improve the estimation of model coefficients compared to frequentist methods. BVARs were first applied to forecasting by Litterman et al. (1979) and Doan et al. (1984), preparing the ground for a large body of research in the following decades. To give a more general overview of all different methods and specifications which have been researched over the past decades, we refer to Karlsson (2013) for an extensive survey of BVARs with a focus on forecasting. Canova (2007) provides an overview of VARs and BVARs in the context of applied macroeconomic research methods. Additionally, Del Negro (2011) provides a clear framework for the different ways in which BVARs can be used and explains these in the context of other Bayesian Macroeconometric methods. An aspect of BVAR methods which will be further elaborated on here is prior selection and inference. A commonly used prior in the BVAR framework is the 'Minnesota' prior, first proposed by Litterman (1980) and Doan et al. (1984) and evaluated and corroborated in Litterman (1986). The basic intuition of this prior is that most macroeconomic variables can be reasonably approximated by a random walk with drift. Furtermore, it incorporates very general economic beliefs on the behaviour of economic time-series. For example, it assumes that the own lags of a variable are more informative than the lags of other variables and that recent lags are more informative than distant lags. The Minnesota prior can be implemented using dummy observations, an insight which dates back to Theil & Goldberger (1961). In the spirit of using economic theory in the determination of priors, Giannone et al. (2019) suggest theory-based priors on the long-run of persistent variables as to provide guidance on the long run joint dynamics of macroeconomic variables. Another interesting method of prior selection bridges the BVAR approach and the aforementioned DGSE models. First proposed by Ingram & Whiteman (1994) and later generalized by Del Negro & Schorfheide (2004), the broad idea is to construct prior distributions based on the restrictions which a DGSE model implies for the VAR coefficients. Other possible priors for BVAR are priors which, rather than using economic theory and intuition, focus on model selection or try to impose sparsity. This class of so-called sparse priors is much larger than the literature on BVARs alone. To elaborate further on these sparse priors, a short overview of the literature on this subject will be given below.

The basic idea of so-called sparse models is to select a relatively small number of non-zero parameters and set the rest of the parameters to zero. An example of this is the well-known lasso method. The lasso method was first proposed by R. Tibshirani (1996) for regular regressions and imposes an ℓ_1 penalty for both fitting and penalization of the coefficients. Park & Casella (2008) adapted the lasso method to a Bayesian framework, stating that the lasso estimate in a linear regression framework can be interpreted as the estimate of the posterior mode when the regressors have independent Laplace priors. Over the years many variations and adaptations of the original lasso method have been proposed which have subsequently been adapted into Bayesian frameworks. An example of one of these adaptations is the very widely used Elastic Net Zou & Hastie (2005), which uses a combination of an ℓ_1 and ℓ_2 term for penalization. The elastic net was then adapted into a Bayesian framework by Li & Lin (2010). Yuan & Lin (2006) proposed the Grouped Lasso method, which allows certain groups of (categorical) predictors to be included and excluded only together. The grouped lasso was adapted into a Bayesian framework by Raman et al. (2009) and Xu & Ghosh (2015). R. J. Tibshirani (2011) proposed the Generalized Lasso, which is - as the name suggests - a generalized form of the original lasso of R. Tibshirani (1996) in which structural constraints can be imposed through the use of a so-called specify penalty matrix. This matrix is equal to I in the standard lasso case. Yet another variant is the so-called Square Root Lasso, first proposed by Belloni, Chernozhukov & Wang (2011). The square root lasso is meant for problems in which the number of possible regressors is very large, but the number of significant regressors is expected to be relatively small. Furthermore, methods were developed specifically within the Bayesian framework rather than being adapted from a standard regression framework. An example of this is the horseshoe prior put forward by Carvalho et al. (2010), which uses a prior based on multivariate-normal scale mixtures and is able to deal with the shortcomings of pure ℓ_1 or ℓ_2 penalization due to the unique, horseshoe-like, shape of the distribution. Piironen & Vehtari (2017) extend upon the horseshoe prior and propose the so-called regularized horseshoe prior, which allows specification of a minimum amount of regularization to the largest values.

The introduction of sparse priors brings us back to the discussion regarding the sparsity or density of economic data mentioned before. Chernozhukov et al. (2017) pose that the spectrum Bayesian methods can broadly be divided into two categories, namely sparse models and dense models. A fairly extensive overview of the many sparse methods has been given above. Dense methods, on the other hand, rather than selecting a small number of larger coefficients, consist of a very large number of coefficients which are heavily shrunk towards zero, but are not equal to zero. An example of a dense method is ridge regression, first proposed by Hoerl & Kennard

(1970), which applies ℓ_2 penalization. Another example is Dynamic Factor Modeling, already briefly mentioned above, which is based on the belief that the behaviour of macroeconomic variables is driven by several unobserved underlying factors, of which principal components are an often-used example.

Giannone et al. (2021a) set out to determine whether economic predictive problems are more likely to be characterized by a sparse or a dense model. They formulate a very general model including two hyperparameters q and γ which correspond to the probability that a certain variable is included in the model and the level of shrinkage which is then applied to this coefficient respectively. A more technically detailed explanation of this model is given in Section 3.1. Giannone et al. (2021a) then apply this model to six widely used data sets in the fields of microeconomics, finance and macroeconomics and evaluate the posterior densities of the hyperparameters. They conclude that for almost all data sets, except one microeconomic data set, the data sets tend more towards density than sparsity. Seeing that, as was shown before, many of the newer proposed Bayesian methods assume some form of sparsity, Giannone et al. (2021a) conclude that there is an 'Illusion of Sparsity' in the academic sphere. Furthermore, Giannone et al. (2021a) show that model uncertainty, expressed through the hyperparameter q, the probability of inclusion, is extremely persistent and they accordingly stress the relevance and importance of using model averaging methods rather than model selection methods in forecasting exercises. The debate on the veracity of these claims is still ongoing. For example, Fava & Lopes (2021) pose that the approach of Giannone et al. (2021a) is not consistently able to distinguish randomly generated predictors from the real data and show that the results regarding the level of density are sensitive to the choice of prior distribution. Gruber & Kastner (2022) state that the level of sparsity or density varies between data sets and time frames and touch upon the potential of dynamic model averaging to combat this issue.

Over the past years, the importance and effects of properly modeling the prior distributions when using Bayesian methods have become more and more clear (Gelman et al. (2017), Betancourt (2021), Team (2017)). Whereas before it was often standard practice to choose a prior to be as uninformative as possible, Gelman et al. (2017) state the following:

> We view much of the recent history of Bayesian inference as a set of converging messages from many directions—theoretical, computational, and applied—all pointing towards the benefits of including real, subject-matter-specific, prior information in order to get more stable and accurate inferences

Of course, as was discussed earlier, many of the methods used in Bayesian macroeconometrics, specifically with respect to BVARs and DGSE models already implement economic intuition and theory heavily in the elicitation of priors. Taking this general sentiment into account and building on widely established practices in the field of Bayesian macroeconometrics, we will adapt the approaches formulated by Giannone et al. (2021a) specifically for the macroeconomic data sets[3] they use by incorporating macroeconomic domain expertise into the elicitation of the priors of this framework and the general set-up of the framework. Then, the effects of incorporating macroeconomic expertise on the behaviour of the prior and posterior distribution of

[3]These are a data set on US industrial production used among others by Stock & Watson (2002) and a data set on cross-country GDP, see Barro & Lee (1994), these will be further explained in Section 2.

the hyperparameters q and γ will be analyzed. Specifically, the effects of changing the prior distributions of the 'first layer' model parameters and the effect of 'fixing' certain variables so they will always be included in the model will be analyzed. The results with respect to the prior and posterior distribution of the hyperparameters will then be compared to the original model proposed by Giannone et al. (2021a) and accordingly we will discuss whether the conclusions of Giannone et al. (2021a) regarding the sparsity or density of these data sets can be corroborated when the proposed changes are made.

We find that using a t-distribution as a prior on β instead of a normal distribution leads to sparser models for the two macroeconomic data sets, extremely so in case of data set on cross-country GDP growth. On the other hand, always including certain variables based on a benchmark of the posterior probability of inclusion leads to significantly denser representations. In contrast to this effect, always including variables based on macroeconomic theory has a very limited effect on the results.

This paper adds to the existing literature by expanding upon the original framework of Giannone et al. (2021a) by incorporating macroeconomic domain expertise into the formulation of the priors and evaluating the effects of doing so on the conclusions. This paper adds in a practical sense to the growing body of literature advocating for the use of more informative priors by evaluating the effects of incorporating more informative priors in practice. Lastly this paper adds to the ongoing 'Illusion of Sparsity' debate by evaluating whether the original conclusions of Giannone et al. (2021a) are corroborated in these more specific cases. To this end, first the results regarding the prior and posterior distribution of the two macroeconomic data sets used in Giannone et al. (2021a) are replicated, and then the proposed strategies are implemented and the results evaluated. In Section 2 the relevant data sets will be explained. Finally in Section 3.1 and 3.2 the original framework of Giannone et al. (2021a) and the proposed adaptations will be discussed respectively. In Section 4 the results will be discussed and finally in Section 5 some concluding remarks will be given.

2 Data

In this research two data sets which are also analyzed in the research of Giannone et al. (2021a) are considered. Specifically, two data sets which correspond to macroeconomic data are considered, referred to as Macro 1 and Macro 2 by Giannone et al. (2021a). In the rest of this paper these data sets will be referred to in the same manner.

The first of these is a data set first used by Stock & Watson (2002) in order to forecast US industrial activity using principal components. The dependent variable is the monthly growth rate of industrial production in the United States and the set of possible explanatory variables consists of 130 macroeconomic indicators. All the data have been transformed to obtain stationarity. This data set is available at FRED-MD and is regularly updated, the most recent version contains data up to April 2023. For the purpose of comparability, the same sample period as in Giannone et al. (2021a) is used, which ranges from February 1960 to December 2014. We obtain the data set from the replication material of Giannone et al. (2021a).

The second data set considered is a data set on the cross-country determinants of long-term economic growth, a large range of which have been collected in the data set constructed by Barro & Lee (1994). The data set contains data for 90 countries, the dependent variable is average GDP growth over the period 1960-1985 and the 60 possible explanatory variables correspond to pre-1960 values of varying measures of institutional, geographic and socio-economic characteristics. We obtain this data set from the replication material of Giannone et al. (2021a), who in turn obtain it from the replication material of Belloni, Chernozhukov & Hansen (2011).

3 Methodology

In this paper, the basic model proposed by Giannone et al. (2021a) is used and expanded upon. In section 3.1 the model proposed in Giannone et al. (2021a) will be laid out. In section 3.2 different ways to adapt this basic framework in order to incorporate macroeconomic theory and intuition will be explained and discussed.

3.1 Model Giannone et al. (2021a)

Giannone et al. (2021a) consider the following model to predict response variable y_t:

$$y_t = u_t'\phi + x_t'\beta + \varepsilon_t, \tag{1}$$

where u_t and x_t are vectors of regressors with dimension l and k respectively, generally it holds that $l \ll k$, and variance is normalized to 1. ε_t denotes an i.i.d. normally distributed error term with mean zero and variance equal to σ^2. Here u_t represents the set of variables that always must be included in the model, thus ϕ is never exactly equal to zero. x_t represents the set of variables which may be useful as predictors, but are not necessarily so, thus some elements of β may be zero.

The following priors are specified for the distribution of the unknown coefficients ϕ, β, σ^2:

$$p(\sigma^2) \propto \frac{1}{\sigma^2}, \tag{2}$$

$$\phi \sim \text{flat}, \tag{3}$$

$$\beta_i|\sigma^2, \gamma^2, q \sim \begin{cases} \mathcal{N}(0, \sigma^2\gamma^2) & \text{with pr. } q, \\ 0 & \text{with pr. } 1-q, \end{cases} \qquad i = 1, ..., k. \tag{4}$$

The priors for σ^2 and ϕ are relatively standard and uninformative. Each element of β is normally distributed with the same variance with probability q or zero with probability 1 - q. The hyperparameter γ^2 determines the level of shrinkage which is applied to the parameters included in β. Giannone et al. (2021a) also propose a different formulation of the prior distribution of β_i, namely $\beta_i|\sigma^2, \gamma^2, q \sim \mathcal{N}(0, \sigma^2\gamma^2 v_i)$, $i = 1, ..., k$, with $v_i \sim$ Bernoulli(q). This formulation highlights the relation between this model and alternative specifications adopted in earlier literature.

Giannone et al. (2021a) state that for example, setting $q = 1$ corresponds to the Bayesian ridge regression and replacing the Bernoulli distribution of v_i with an exponential, shifted exponential, half-Cauchy or truncated Gamma density allows us to get the Bayesian lasso, lava, horseshoe, and elastic net methods respectively[4].

To determine a prior on the hyperparameters q and γ^2 the following mapping is defined: $R^2(\gamma^2, q) = \frac{qk\gamma^2\bar{v}_x}{qk\gamma^2\bar{v}_x + 1}$. $\bar{v}_x$ is the predictors' average sample variance, in this scenario equal to 1 due to the standardization of regressors x. The independent priors on q and R^2 are specified as follows:

$$q \sim \text{Beta}(a, b), \tag{5}$$

$$R^2 \sim \text{Beta}(A, B). \tag{6}$$

For q we set $a = b = 1$, which corresponds to the uniform distribution. The R^2 is also assigned a Beta distribution with $A = B = 1$, which again corresponds to the uniform distribution. The posterior distributions of the different parameters are calculated using the Markov Chain Monte Carlo algorithm proposed by Giannone et al. (2021a). This algorithm is described in detail in the Appendix of Giannone et al. (2021a).

3.2 Incorporating economic theory and intuition

There are different methods to incorporate macroeconomic theory or intuition into the basic model of Giannone et al. (2021a). The two methods which will be discussed and analyzed in this paper are (i) changing the prior distributions of model parameters and (ii) forcing theoretically relevant parameters to always be included in the model. In the rest of this section these two approaches will be described in more detail.

3.2.1 Prior distribution of β

In their paper, Giannone et al. (2021a) note that a misspecification of the distribution of non-zero coefficients may lead to a poor performance of their approach:

> The drawback of this approach, however, is that it might perform poorly if our parametric assumption is not a good approximation of the distribution of the non-zero coefficients

Although in a supplement to their paper, Giannone et al. (2021b) perform a simulation experiment on the effectiveness of their approach when the true DGP differs from the normal distribution, they do not explore any alternative prior distributions for β applied to the actual data. Imposing a normal distribution as a prior on β might have the effect of concentrating the posterior distribution of β around zero and not allowing for any large values to 'escape' the shrinkage. Considering that there is usually a high degree of collinearity between the regressors in macroeconomic data sets as mentioned by Hendry (2018) and Giannone et al. (2021a), it might thus be interesting to impose a prior distribution with fatter tails on β. This allows for

[4]See Park & Casella (2008), Chernozhukov et al. (2017), Carvalho et al. (2010) and Li & Lin (2010) respectively.

larger coefficient values to escape, which subsequently could lead to a denser specification if these larger coefficients are still able to explain the data well. A suitable prior distribution to allow for fatter tails is the t-distribution. An efficient method to incorporate this adapted prior into the MCMC algorithm is the so-called latent variable representation of the t-distribution, discussed by Van Erp et al. (2019) Chan (2017) and Griffin & Brown (2005) among others.

Specifically, following Fava & Lopes (2021) the prior distribution of $\beta_i|\sigma^2, \gamma^2, q$ originally proposed by Giannone et al. (2021a) and discussed in Section 3.1 is changed to the following:

$$\beta_i|\sigma^2, \gamma^2, \lambda_i^2, q \sim \begin{cases} \mathcal{N}(0, \sigma^2\gamma^2\lambda_i^2) & \text{with pr. } q, \\ 0 & \text{with pr. } 1-q, \end{cases} \qquad i = 1, ..., k. \tag{7}$$

Where the prior distribution of λ_i^2 is set to an Inverse-Gamma distribution such that:

$$\lambda_i^2 \sim IG(\frac{\nu}{2}, \frac{\nu}{2}) \tag{8}$$

It can then be shown that:

$$\beta_i|\sigma^2, \gamma^2, q \sim \begin{cases} t_\nu(0, \sigma^2\gamma^2) & \text{with pr. } q, \\ 0 & \text{with pr. } 1-q, \end{cases} \qquad i = 1, ..., k. \tag{9}$$

Following Fava & Lopes (2021), we redefine $\bar{v}_x$ as:

$$\bar{v}_x \equiv E[\sigma_{i,i}]\frac{\nu}{\nu - 2} \tag{10}$$

With this redefinition, the conditional posterior distributions of R^2, q, ϕ, z and σ^2 are unchanged (Fava & Lopes (2021)). The conditional distribution of β_i after incorporating these changes is as follows:

$$\frac{\beta_i}{\sqrt{\lambda_i^2}} := \beta_i^*|Y, \phi, \sigma^2, R^2, q, z \sim \begin{cases} t_\nu(0, \sigma^2\gamma^2) & \text{with pr. } q, \\ 0 & \text{with pr. } 1-q, \end{cases} \qquad i = 1, ..., k. \tag{11}$$

Lastly, the conditional distribution of λ_i^2 is given by:

$$\lambda_i^2|v, \beta_i, \sigma^2, R^2 \sim IG(\frac{\nu+1}{2}, \frac{\nu + \beta_i^2/\sigma^2\gamma^2}{2}) \tag{12}$$

We extend the algorithm proposed in the Appendix of Giannone et al. (2021a) using the adaptations proposed above to set a t-distribution as prior distribution for β. Following Fava & Lopes (2021), we do not try to learn an optimal value for ν, since the goal of this research is to evaluate the effect of setting a different prior distribution on the conclusions regarding the sparsity or density of the data rather than to find an optimal predictive model. Thus, as in Fava & Lopes (2021) we present results for ν=4, 10, 30 and 100. An interesting avenue for further research would be to leave ν as unknown and determine the optimal degrees of freedom for the distribution of β through the data.

3.2.2 Theoretically relevant variables

Within the model of Giannone et al. (2021a) it is relatively simple to 'force' theoretically relevant variables to always be included in the model. That is, by moving certain variables from x_t to u_t these variables will always be included in the model. This allows us to investigate the effects of always including these variables on the conclusions regarding the sparsity or density of the data. As was discussed in Section 3.1, the parameter paired to the variables included in u_t is ϕ. The prior on ϕ in the model of Giannone et al. (2021a) is flat. For completeness, we investigate the effects of changing the prior on ϕ in the scenario where more variables are included in u_t. Specifically the prior for ϕ is defined as follows:

$$\phi \sim \mathcal{N}(1, \sigma^2) \tag{13}$$

We use this specification to stay close to the original model of Giannone et al. (2021a). We choose a mean of 1 instead of zero for this specification, since we assume that the variables included in u_t are meaningful and thus should not be shrunk towards zero. However, we find that using a normal prior instead of a flat prior on ϕ does not meaningfully change any of the results, similarly so when the mean is set to zero. Therefore, the results using a normal prior on ϕ are not included in this paper. For completeness and for possible future reference, we include the derivation for the posterior distribution of ϕ using the above prior in Appendix A.

In this paper two different methods to determine which theoretically relevant variables should always be included in the model are proposed. One method which is based on macroeconomic theory and one method which uses insights and intuition from the field of Bayesian macroeconometrics. Both will be described below, starting with the method based on macroeconomic theory.

Hendry & Johansen (2015), in the spirit of Haavelmo (1944), pose that a macroeconomic model should always include certain explanatory variables which are based on economic theory and should additionally include a set of explanatory variables which might be zero, and of which the decision of inclusion is purely based on data. Following this line of reasoning, for both data sets we determine the variables which should be moved from x_t to u_t based on theoretical literature on the respective topics of these data sets.

We start by determining the theoretically relevant variables for the Macro 1 data set[5]. In this data the US industrial production is considered as the dependent variable, and thus US industrial production is also our variable of interest with respect to theoretical literature. Shi et al. (2018) pose that the slope of the yield curve is a potentially important explanatory variable with respect to explaining real economic growth and in predicting recessions[6]. They propose using the following economic phenomena to explain real macroeconomic growth[7]: inflation, monetary policy interest rate and yield curve spread. Following Shi et al. (2018) these general macroeconomic effects can be proxied by the core consumer price index, effective federal funds rate, and the difference between the 10-year rate and the federal funds rate respectively. Using

[5]see Section 2

[6]See also Stock & Watson (1989), Estrella & Hardouvelis (1991), Plosser & Rouwenhorst (1994)

[7]Here US industrial production is considered to be a proxy for real macroeconomic growth

the findings of Shi et al. (2018), we thus choose to move the variables corresponding to the core consumer price index, the effective federal funds rate and the difference between the 10-year treasury note yield and the effective federal funds rate from x_t to u_t. These three variables and their corresponding codes and column numbers in the Fred-MD index are shown in Table 1.

Table 1 This table shows ID's, FRED Codes and descriptions for the variables of the Macro 1 data set which are added to u_t based on the results derived by among others Shi et al. (2018)

ID[8]	FRED Code	Description[9]
83	FEDFUNDS	Effective Federal Funds rate
97	T10YFFM	Difference between the 10-year rate and the Federal Funds rate
111	CPIAUCSL	Total consumer price index

To determine which variables should be added to u_t for the Macro 2 data set, we use the results obtained by Barro & Lee (1994). They find that a combination of (i) variables which concern human capital, (ii) variables related to investment and government spending, and (iii) variables related to political stability and corruption have a significant effect on cross-country GDP growth. Building on the findings of Barro & Lee (1994), we choose to move the variables (i) Total year of secondary education, Life expectancy, (ii) Ratio of total investment to GDP, Ratio of total government expenditure to GDP, (iii) Black market premium and Political instability from x_t to u_t. In Table 2 an overview of these variables, their place in the data set and a short description is given.

Table 2 This table shows ID's, Codes and Descriptions for the variables of the Macro 2 data set which are added to u_t based on the results derived by Barro & Lee (1994).

ID	Code	Description[10]
3	bmp1l	Black market premium
17	lifee065	Life expectancy
21	invsh41	Total investment to GDP ratio
26	govsh41	Goverment expenditure to GDP ratio
43	pinstab1	Political instability
54	syr65	Total years of secondary education (male and female)

Another method, using intuition from Bayesian macroeconometrics, to determine which variables should always be included in the model is based on the intuition behind prior selection for GGSE models, as described by Del Negro (2011). Del Negro (2011) notes that a common practice to obtain prior information in a more literal way is to look at the data in the period before the sample period. Taking the intuition behind this approach we want to include the

[8]These ID's do not perfectly match the ID's denoted in the Appendix of McCracken & Ng (2016) as some columns are removed due to lack of data.

[9]Based on the Appendix of McCracken & Ng (2016).

[10]Data on codes and descriptions is obtained from a readme on the Barro & Lee (1994) data set provided by the NBER, see `http://www2.nber.org/pub/barro.lee/readme.txt`.

variables which were overall most likely to be included in the original model. Specifically, we look at the posterior probabilities of inclusion of the predictors obtained through the results of the original model. If a variable has a large posterior probability of inclusion, it was included into the model a large amount of time. If a variable is included often, it is intuitive to say that these variables, based on the data, should be informative. Using this line of reasoning, we will select which variables to move from x_t to u_t based on the posterior probability of inclusion of that particular variable in the results of the original model of Giannone et al. (2021a). To get these posterior probabilities of inclusion, which are also shown in the heatmap of Figure 3, we evaluate $\bar{z}_i$, the average of the values of z_{ij} over the total number of draws of predictor i. For each predictor and each draw, z_{ij} is defined as follows:

$$z_{ij} = \begin{cases} 1 \text{ if variable } i \text{ is included in the model in draw } j, \\ 0 \text{ otherwise} \end{cases} \qquad i = 1, ..., k, \; j = 1, ..., M - N. \quad (14)$$

Thus $\bar{z}$ has the intuitive interpretation as the percentage of times a certain variable is included in the model, for example a value of $\bar{z}$ of 0.85 would mean that a variable was included in the model in 85% of the total number of runs of the model. To determine which variables should be included as elements of u_t, a benchmark for $\bar{z}_i$ is determined. If $\bar{z}_i$ is above this benchmark, variable i will be added to u_t. In this research we evaluate four benchmarks for $\bar{z}_i$, namely 0.80, 0.90, 0.95 and 0.99. The latter three of these benchmarks correspond to the significance levels generally considered in the literature, namely 10%, 5% and 1% respectively. As the data of the Macro 2 data set only has one variable which has a value for $\bar{z}$ above 0.95 and none above 0.99, we also analyze the results using a benchmark for $\bar{z}_i$ of 0.80. Table 3 and Table 4 show for the Macro 1 and Macro 2 data sets respectively which variables are included in u_t for the different benchmarks.

Table 3 This table shows ID's, FRED Codes, descriptions and values for $\bar{z}$ for the variables of the Macro 1 data set which are added to u_t based on the determined benchmarks for $\bar{z}$. *, **, *** and **** show that a value is larger than the benchmark of 0.80, 0.90, 0.95 and 0.99 respectively.

ID	CODE	Description	$\bar{z}$
39	NDMANEMP	All employees - non-durable goods	0.9933****
61	NAPMNOI	ISM: New Orders Index	0.9695***
125	CES2000000008	Avg Hourly earnings: Construction	0.9102**
109	NAPMPRI	ISM Manufacturing: Prices Index	0.9012**
32	CLAIMSx	Initial Claims	0.8532*
92	TB3SMFFM	3-Month Treasury C Minus FEDFUNDS	0.8157*

Table 4 This table shows ID's, Codes, and values for $\bar{z}$ for the variables of the Macro 2 data set which are added to u_t based on the determined benchmarks for $\bar{z}$. $\cdot, \cdot\cdot$ show that a value is larger than the benchmark of 0.80 or 0.95 respectively.

ID	Code	Description	$\bar{z}$
1	bmp1l	Black market premium	0.9759$\cdot\cdot$
7	p65	Total enrolment primary education	0.8100$\cdot$
22	gde1	Ratio of government spending to defense expenditure	0.8009$\cdot$

4 Results

The results obtained will be laid out in two parts. First, the results related to the prior and posterior distribution and values of q and γ obtained by Giannone et al. (2021a) will be replicated for the data sets Macro 1 and Macro 2, discussed in Section 2. These results are shown in Section 4.1. Then, the results obtained by applying the different methods discussed in Section 3.2 are shown in Section 4.2

4.1 Replication

In Figure 1 the prior and posterior distribution of q and $log(\gamma)$ for the Macro 1 and Macro 2 data sets is shown. These results are obtained using the model specifications as discussed in Section 3.1 and are thus a replication of the results obtained by Giannone et al. (2021a). By construction, q and $log(\gamma)$ are negatively correlated a priori. This reflects the prior belief that shrinkage and sparsity are substitutes when it comes to dealing with dimensionality. In Figure 1 it becomes clear that the posterior joint distribution of q and γ is much more concentrated than the prior distribution and that the negative correlation which was implied a priori is also present a posteriori and even seems to be more strongly present. Furthermore, it is interesting to note that the posterior distribution of the Macro 1 data set is concentrated around lower values for q than the Macro 2 data set. These results corroborate the conclusions drawn by Giannone et al. (2021a).

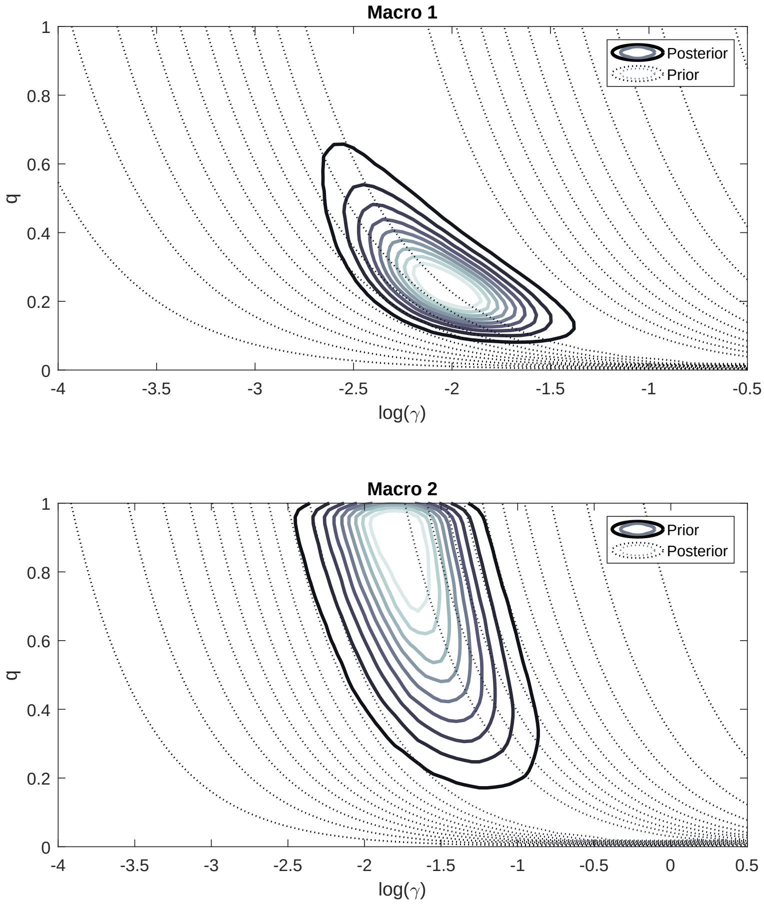


Figure 1. Contours of the prior and posterior distribution of q and $log(\gamma)$.

The difference between the posterior distributions of q of the Macro 1 and Macro 2 data sets becomes even more clear when looking at Figure 2. The curve of the posterior distribution of q for the Macro 1 data set has a nice bell shape which peaks at values for q between 0.2 and 0.3. On the other hand, the curve of the posterior distribution for the Macro 2 data set trends steadily upwards and peaks at $q = 1$. Thus, it seems that a denser model in which (almost) all variables are included is more suitable when explaining the Macro 2 data, while the Macro 1 data is better explained by a smaller – though still sizable – subset of the explanatory variables. In the case of the Macro 1 data set, where optimally a subset of the explanatory variables is included, it is thus particularly interesting whether it can be determined which variables should actually be included. Again, these results corroborate the analysis of Giannone et al. (2021a).

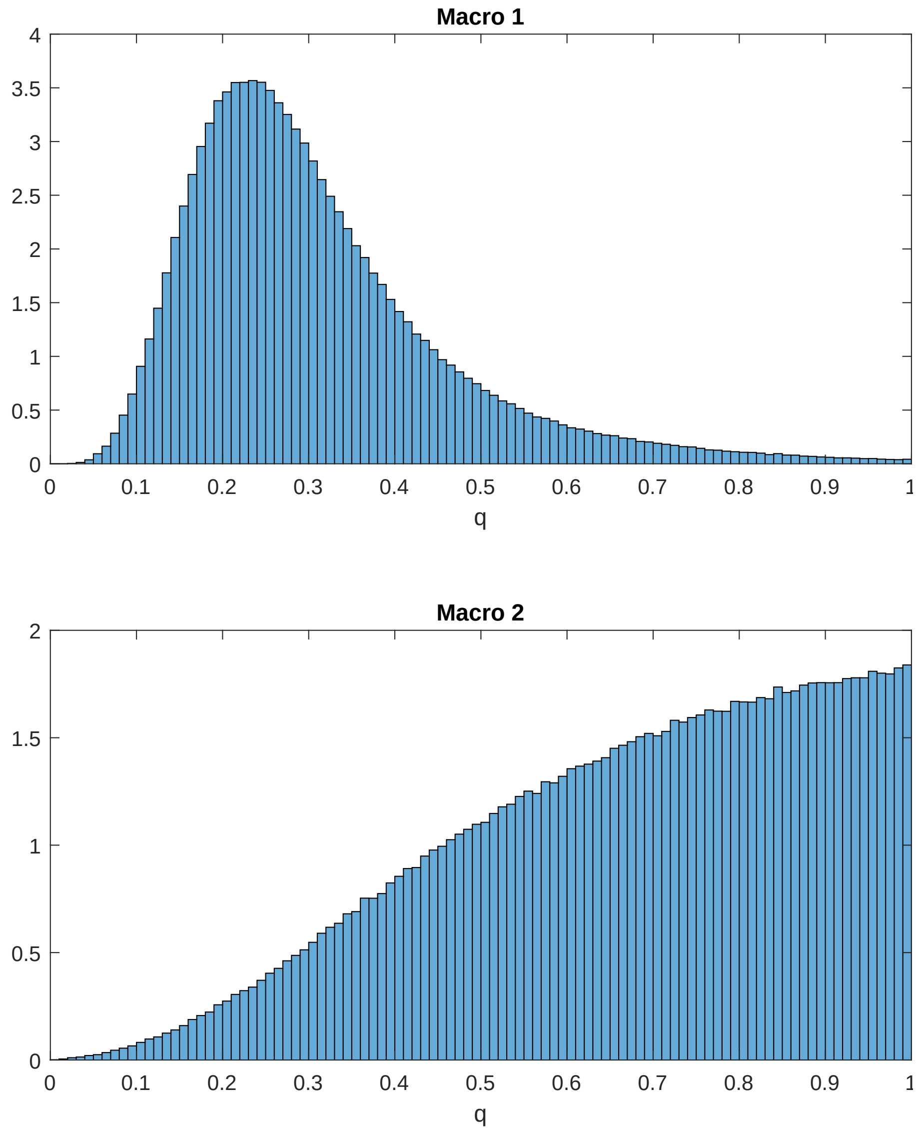


Figure 2. Posterior density of q.

In Figure 3 the probabilities of inclusion for each predictor are plotted. A lighter colour means that a specific predictor is included relatively less often in the model. Figure 3 shows that there is a lot of uncertainty about which variables should actually be included in the model. For example, in the Macro 1 case there is not even a single variable which is (almost) never included, as no predictor shows a white bar, which would suggest a posterior probability of inclusion close to zero. Some predictors do seem to be included most of the times, shown by the darker bars Figure 3. For the Macro 2 data, the posterior q's are all relatively large and no clear pattern arises on predictors which are included significantly more often, except for the first variable. As was stressed by Giannone et al. (2021a), these figures show that the level of model uncertainty is very large, which makes the use of model averaging methods over model selection methods even more relevant in predictive exercises.

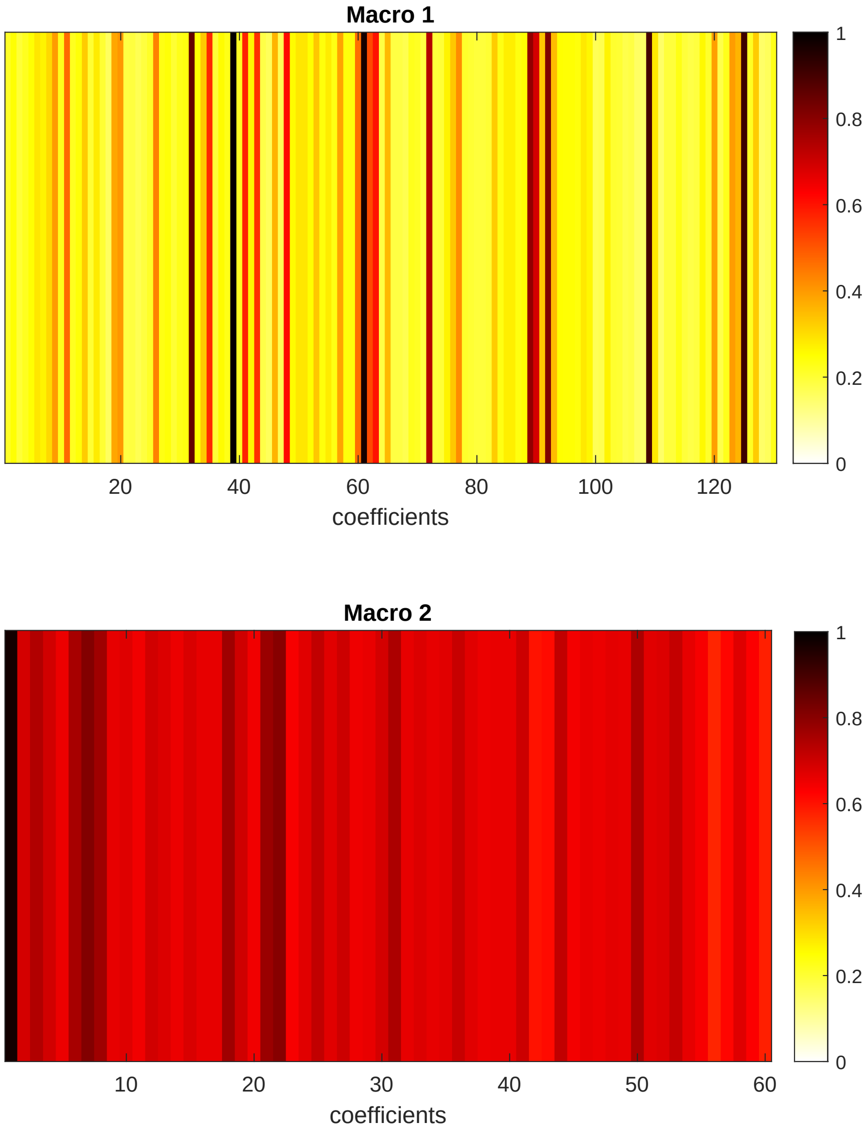


Figure 3. Heat map of the probabilities of inclusion of each predictor.

Finally, Figure 4 shows the probability of inclusion of each predictor conditional on every possible value for q. Again, darker colours signify a larger probability of inclusion, the white-and-black dotted line corresponds to the posterior mode of q. The graphs in Figure 4 show quite clearly that assuming a priori that a model is sparse – and thus that q is low – induces sparser results a posteriori. It is interesting to note that even with very harsh assumptions on the sparsity of a model, there is still relatively much uncertainty with regards to which variables should be included, which is again in line with the conclusions drawn by Giannone et al. (2021a).

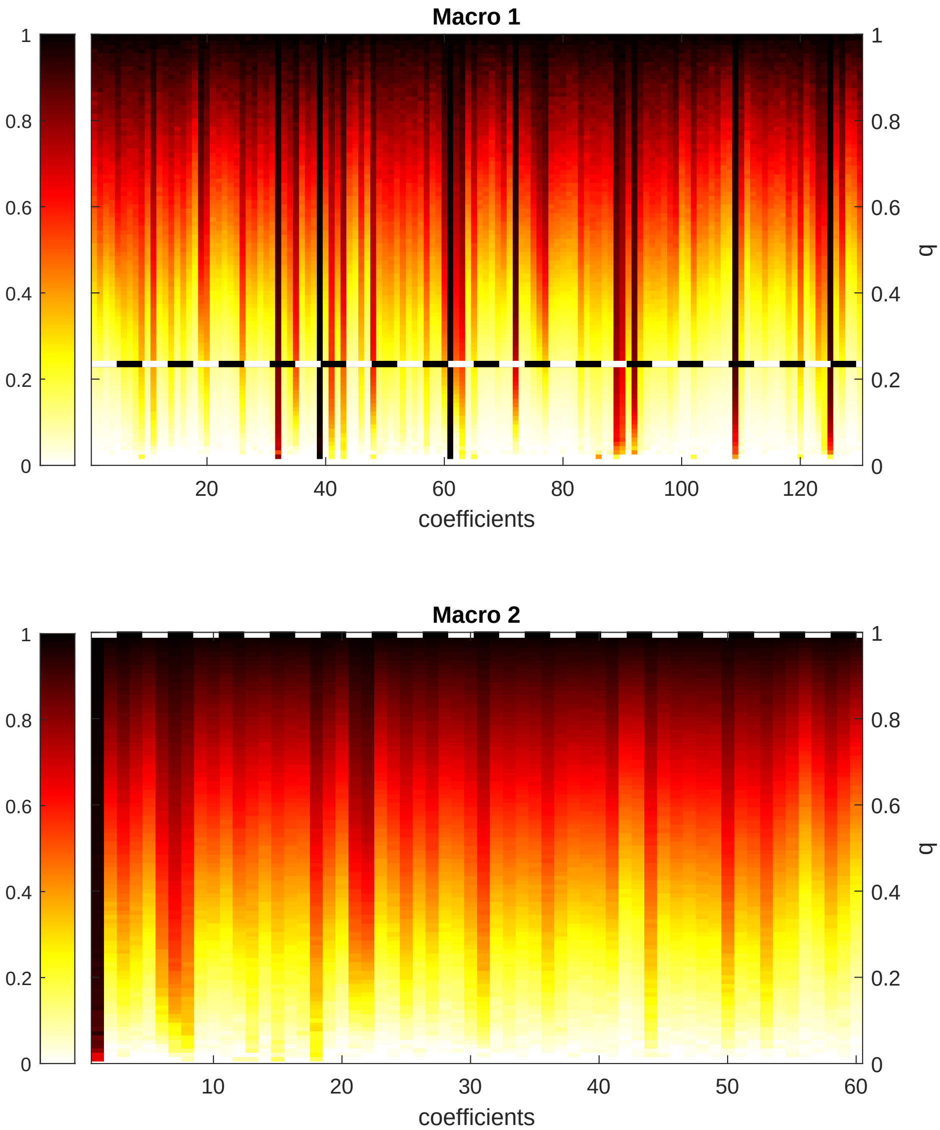


Figure 4. Heat map of the probabilities of inclusion of each predictor, conditional on q. The horizontal dashed line denotes the posterior mode of q.

4.2 Extension

In this section the results are shown for the approaches laid out in Section 3.2. Specifically, in Section 4.2.1 the results corresponding to the changes made to the prior of β proposed in Section 3.2.1 are discussed. In Section 4.2.2 the results of always including theoretically relevant variables as explained in Section 3.2.2 will be shown. In order to compare the results of these different implementations and the original results of Giannone et al. (2021a), the figures shown in Section 4.1 are recreated for the newly applied methods. The figures relevant for discussion and interpretation are shown in the paper, the full set of resulting figures is shown in Appendices B and C.

4.2.1 Prior distribution of β

The contours of the prior and posterior density of q and $\log(\gamma)$ for the model in which the prior distribution of β is a t-distribution are shown in Figure 5 for the Macro 1 data set and in

Figure 6 for the Macro 2 data set. Note that in Figure 5 the results for degrees of freedom ν equal to 4, 10, 30 and 100 are shown while in Figure 6 results for degrees of freedom ν equal to 4, 30, 100 and 2000 are shown. This is due to the fact that the results for the Macro 2 data set with a prior t-distribution show extreme sparsity for lower degrees of freedom and thus the figures for ν equal to 4, 10 and 30 are the same. As this does not lead to a very interesting analysis, the results for ν equal to 2000 are shown for comparison with the original results of Giannone et al. (2021a) with a normal prior distribution on β, as the t-distribution tends to a normal distribution when the degrees of freedom become large.

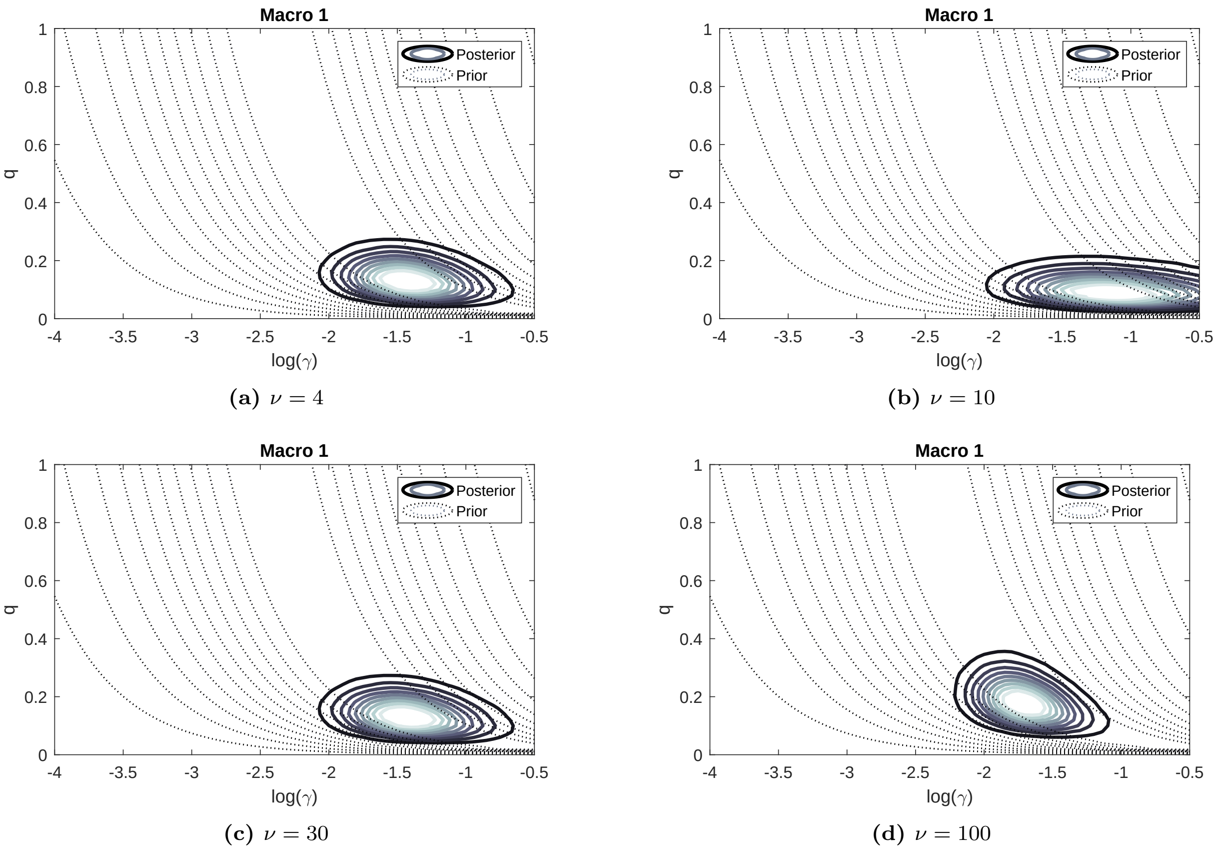
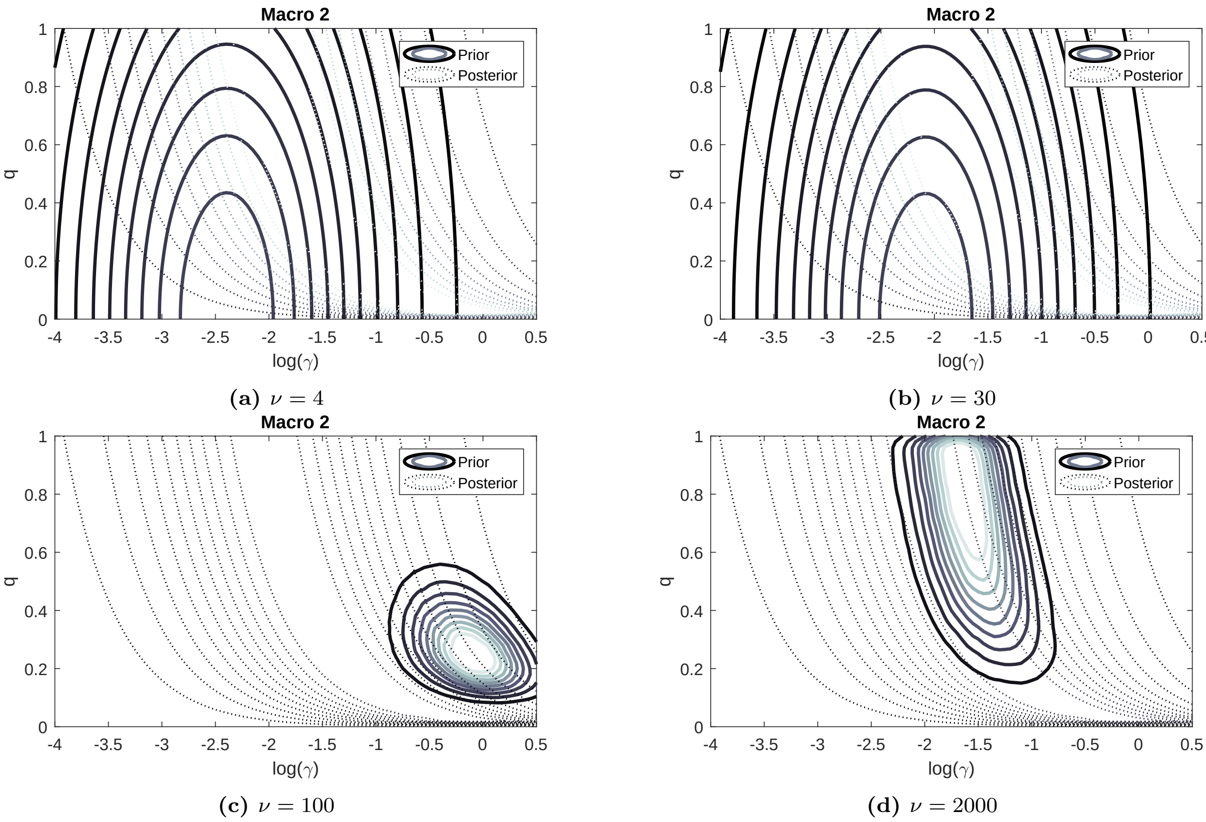


Figure 5. Contours of the prior and posterior density of q and $\log(\gamma)$ for the model in which a prior with t-distribution with degrees of freedom ν equal to 4, 10, 30 and 100 in Subfigure 5a, 5b, 5c and 5d respectively.

In Figure 5 we see that setting a prior t-distribution on β leads to a sparser representation when compared to the results shown in Figure 1. Furthermore, the negative relation between the probability of inclusion q and the level of shrinkage γ becomes less negative. Finally, the plots are more concentrated towards the bottom right side of the figure. We can see from Figure 5 that setting the t-distribution as a prior on β leads to a sparser representation and relatively less shrinkage when compared to the original model of Giannone et al. (2021a). When comparing the subfigures, it becomes clear that as ν becomes larger, the model becomes less sparse, the negative relation between q and γ becomes more pronounced and more shrinkage is applied. In short, as the degrees of freedom of the prior t-distribution increase, the figures start tending towards the original results shown in Figure 1, which is to be expected, since the t-distribution tends towards the normal distribution when the degrees of freedom become larger. Looking at the results for the Macro 2 data set, shown in Figure 6, we see an even more extreme effect of setting a prior t-distribution on β. In the original model of Giannone et al. (2021a), the Macro 2 data set had a relatively dense representation, which can still be seen in Figure 6d. However, for smaller values of ν the data tends to extreme sparsity, in the sense that no variables are selected into the model at all. This becomes abundantly clear in Figure 10. The results for ν equal to 100, shown in Figure 6c are also very interesting, as the results for the Macro 2 data set for this value of ν resemble the figures of the Macro 1 data set shown in Figure 5. The posterior density of q is concentrated around 0.2 and there is clear negative relationship between q and $\log(\gamma)$.



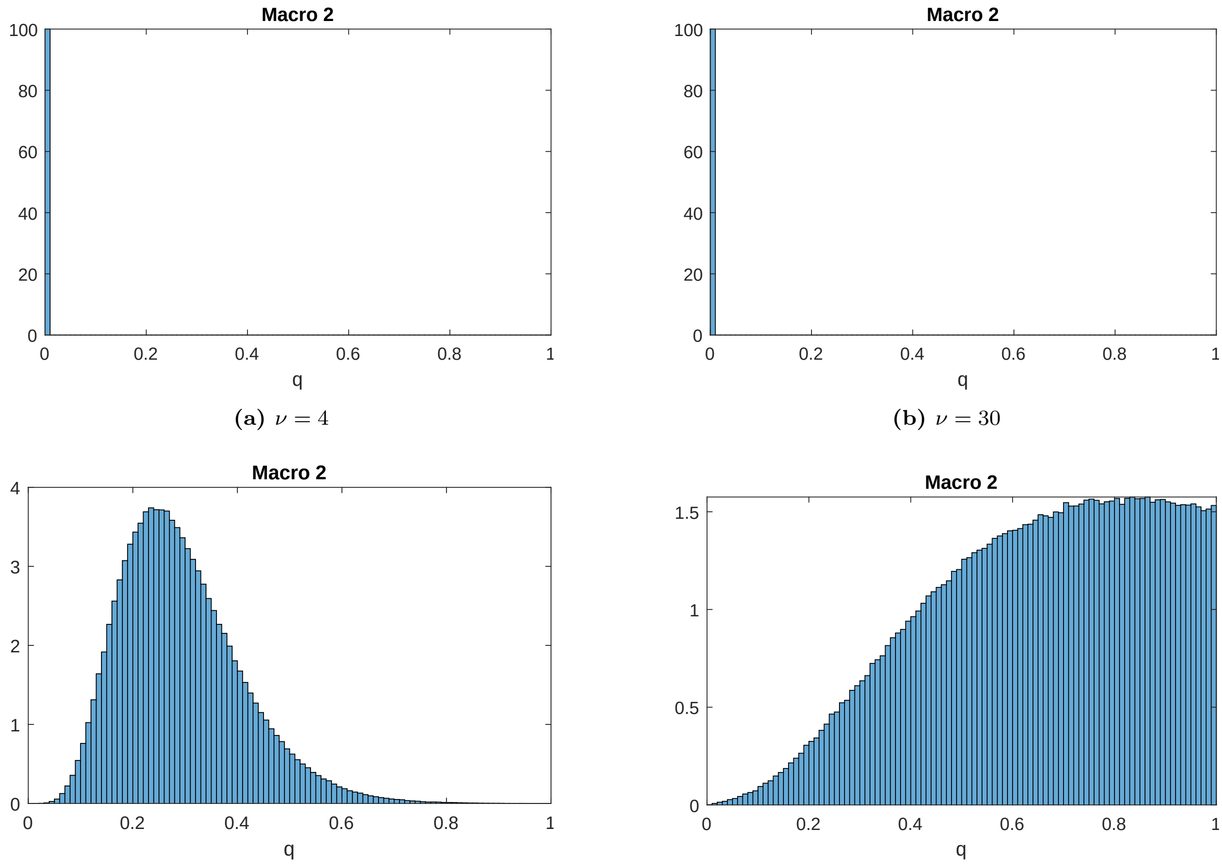
(a) $\nu = 4$ **(b)** $\nu = 30$ **(c)** $\nu = 100$ **(d)** $\nu = 2000$

Figure 6. Contours of the prior and posterior density of q and $\log(\gamma)$ for the model in which a prior with t-distribution with degrees of freedom ν equal to 4, 30, 100 and 2000 in Subfigure 6a, 6b, 6c and 6drespectively.

Looking at Figure 8 we can confirm the observations made based on the prior and posterior densities shown in Figure 5. We see that the bell shape of the posterior distribution which was also observed in the results of Giannone et al. (2021a) in Figure 2 is still present, but that it is steeper. Although the posterior distribution of q with a prior t-distribution on β looks similar to the case with a normal prior on β, it becomes clear that the posterior densities of q shown in Figure 5 are more concentrated in general and are concentrated around smaller values of q than the posterior density of the original Giannone et al. (2021a) model shown in Figure 2. As became clear from Figure 5 the posterior densities tend to move more towards the original figure of Giannone et al. (2021a) when ν becomes larger. That is, as ν increases, the graphs concentrate around larger values of q and have larger tails.

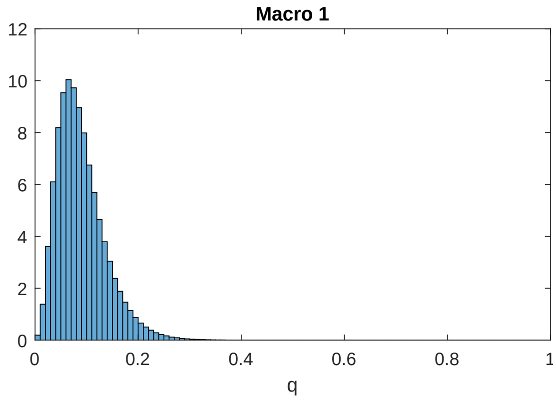
The posterior densities of q of the Macro 2 data set are shown in Figure 7. As for the Macro 1 data set, the findings drawn from Figure 6 translate well into this figure. Thus, logically, the histogram of the posterior distribution of q for lower values of ν is simply a straight bar at 0.

Again, we see that the results for ν equal to 100 shown in Subfigure 7c show resemblance to the figures of the Macro 1 data set in shape and concentration, while the results for ν equal to 2000 in Subfigure 7d resemble the results of the original model of Giannone et al. (2021a).

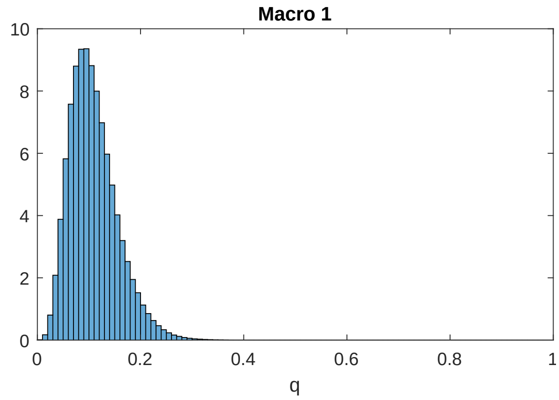


(a) $\nu = 4$ **(b)** $\nu = 30$ **(c)** $\nu = 100$ **(d)** $\nu = 2000$

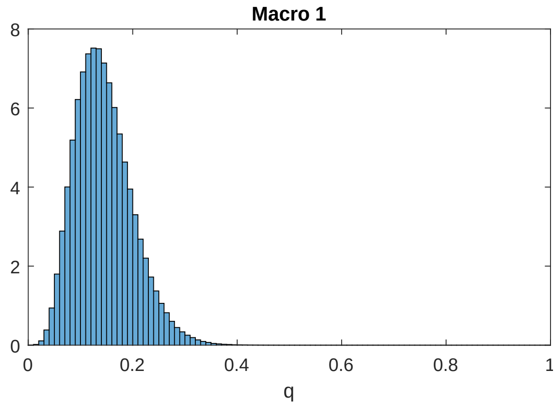
Figure 7. Posterior density of q for the model in which a prior with t-distribution with degrees of freedom ν equal to 4, 30, 100 and 2000 in Subfigure 7a, 7b,7c and 7d.



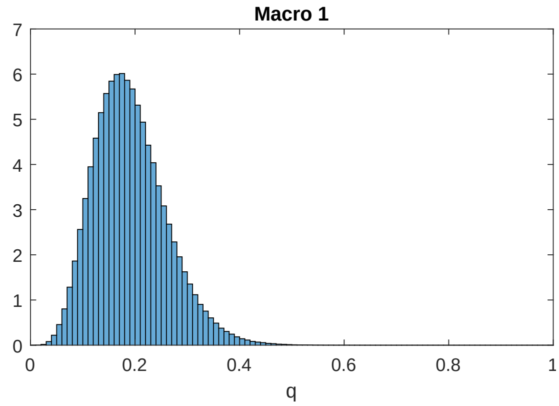
(a) $\nu = 4$



(b) $\nu = 10$



(c) $\nu = 30$



(d) $\nu = 100$

Figure 8. Posterior density of q for the model in which a prior with t-distribution with degrees of freedom ν equal to 4, 10, 30 and 100 in Subfigure 8a, 8b, 8c and 8d respectively.

In Figure 9 and 10 the heatmaps of the probabilities of inclusion of each predictor are shown for the Macro 1 and Macro 2 data set respectively. Starting with the results of the Macro 1 data set shown in Figure 9, it becomes clear that changing the prior distribution of β from a normal distribution to a t-distribution does not only impact the level and concentration of sparsity, but it also significantly impacts the pattern of sparsity. The heatmap for the original model in Figure 3 shows almost no white or light-yellow bars. Consequently, Giannone et al. (2021a) conclude that there is a lot of uncertainty about which regressors should actually be included in the model and accordingly they stress the importance of (weighted) model averaging to combat this uncertainty. Looking at Figure 9, it can clearly be seen that the heatmap is significantly more lightly coloured than in the case of the original Giannone et al. (2021a) model and thus there is less uncertainty about which regressors should be included in the model. This does not mean that for example the model with a prior t-distribution with ν equal to 4 immediately allows for a truly sparse representation with specifically selected variables, but model uncertainty does not seem to be as pronounced for this model specification compared to the original Giannone et al. (2021a) model. As was seen in the previous figures, the heatmaps tend more towards the heatmaps of the original model as ν becomes larger.

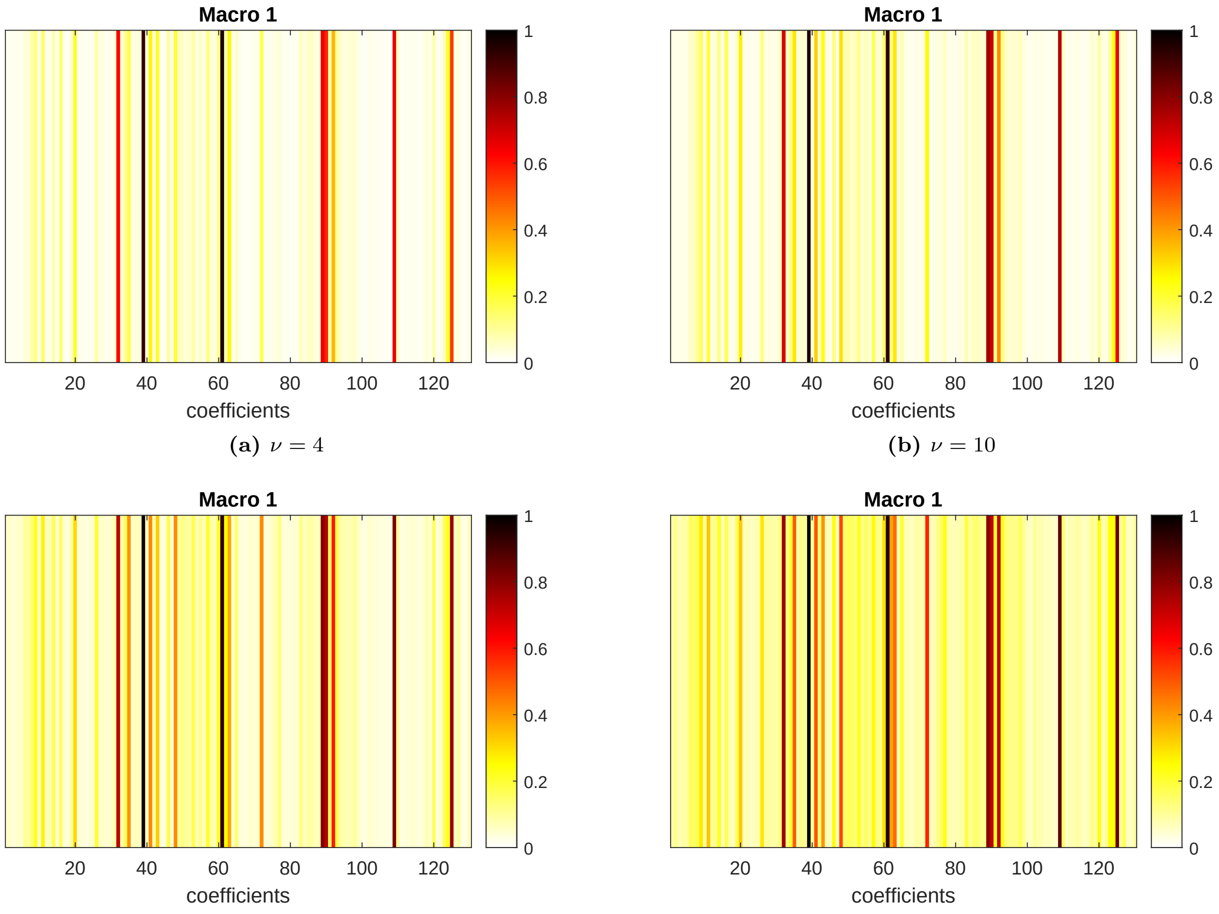
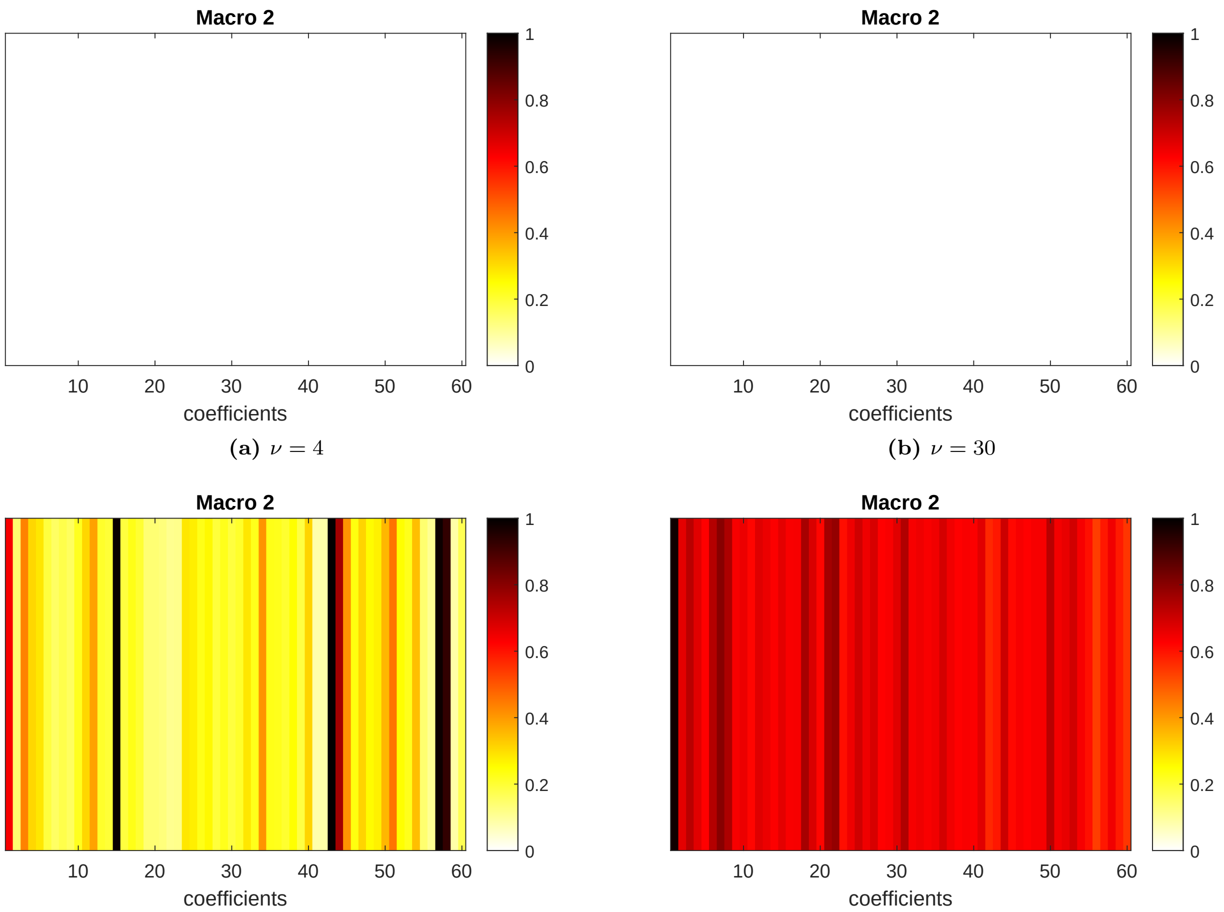


Figure 9. Heat map of the probabilities of inclusion of each predictor for the Macro 1 data set with a prior with t-distribution with degrees of freedom ν equal to 4, 10, 30 and 100 in Subfigure 9a, 9b, 9c and 9d respectively.

The first feature to be noticed in Figure 10 is of course the fact that Subfigures 10a and 10b are completely white. This is probably the clearest indication that no variables are included at all when a prior t-distribution with lower degrees of freedom is considered for β in the Macro 2 data set. Subfigure 10c shows the heatmap for ν equal to 100. As was seen in the previous figures, the model for the Macro 2 data set for ν equal to 100 shows relatively similar behaviour to the Macro 1 models. It can clearly be seen that some variables are included almost every single time, but that most variables are included only around 30% of the time, which is apparent from the large number of yellow bars and the occasional dark bar in the heatmap. Thus, for ν equal to 100, there is a lot of uncertainty with respect to which variables should be included in the model, and thus model averaging would be a suitable technique to deal with this issue. Finally, Subfigure 10d shows the heatmap for ν equal to 2000. This heatmap, as is to be expected, looks very similar to Figure 3. All variables are included most of the time, but there are no specific variables which are always selected. The conditional heatmaps for the Macro 1 and Macro 2 data sets with a prior t-distribution, similar to Figure 4 do not provide any further insights and are thus not discussed. These can be found in Appendix B.



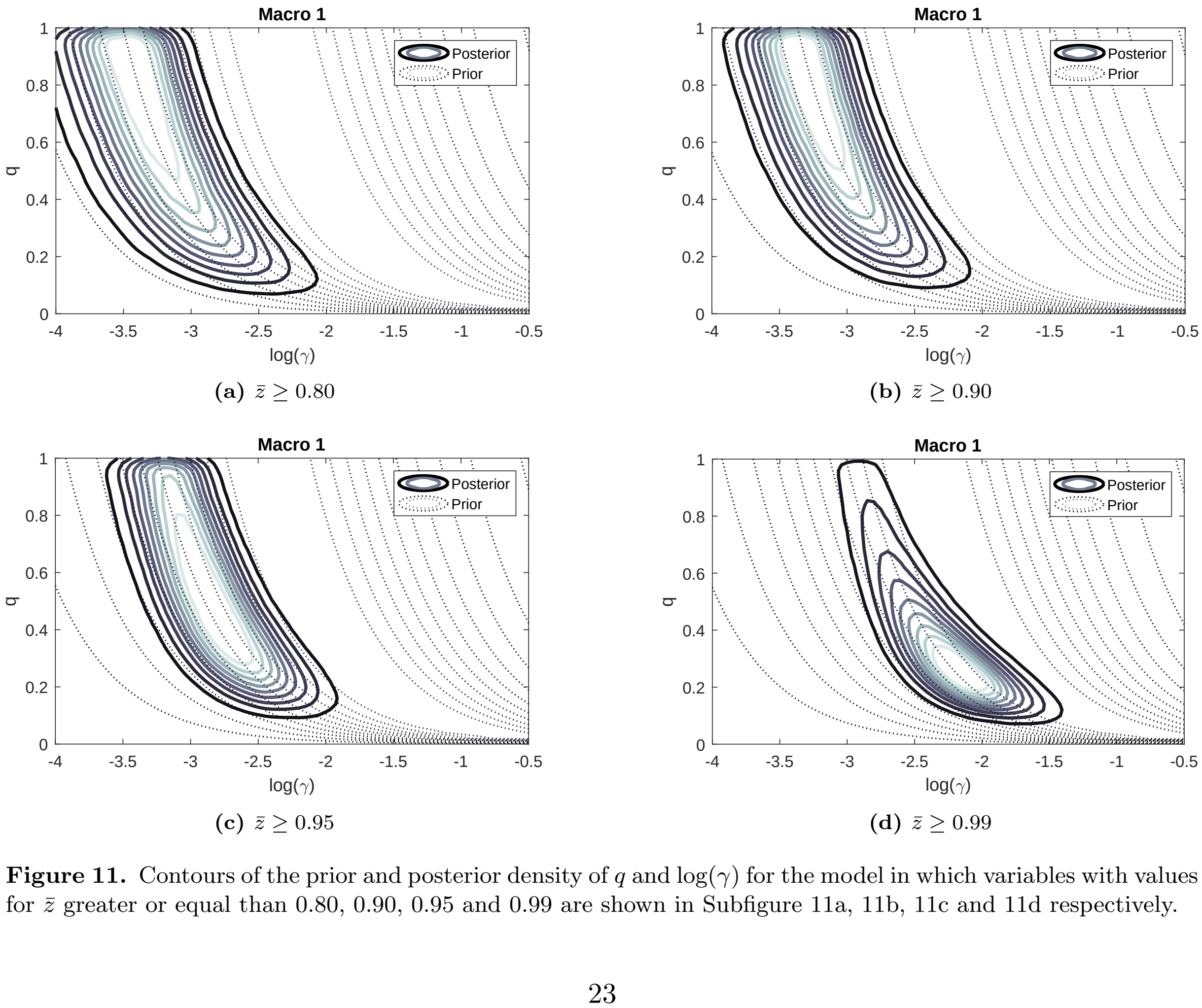
(a) $\nu = 4$ (b) $\nu = 30$ (c) $\nu = 100$ (d) $\nu = 2000$

Figure 10. Heat map of the probabilities of inclusion of each predictor for the Macro 2 data set with a prior with t-distribution with degrees of freedom ν equal to 4, 30, 100 and 2000 in Subfigure 10a, 10b, 10c,and 10d.

In conclusion, it becomes clear that changing the prior distribution of β from a normal distribution to a t-distribution has large effects on the inference with respect to the sparsity or density of the data and with respect to the patterns in this sparsity. For the Macro 1 data set these effects are relatively nuanced, it becomes clear that the model with a prior t-distribution on β leads to sparser specifications, a less pronounced negative relationship between the probability of inclusion and the level of shrinkage and in particular to a clearer pattern of sparsity. It also becomes clear that as the degrees of freedom ν increase, the results start tending towards the original results of Giannone et al. (2021a). In contrast to the relatively nuanced effects on the results for the Macro 1 data set, the effects of changing the prior distribution of β on the results of the Macro 2 are extreme. The results of Giannone et al. (2021a) show that the Macro 2 data set tends towards the densest representation and thus that all variables should be included in the model for this data set. However, when the prior distribution of β is set as a t-distribution, for values of ν up to at least 30, the Macro 2 data tends to the sparsest representation possible: no variables included at all. For values of ν around 100 the Macro 2 data set still has a rather sparse representation, specifically when compared to the original results of Giannone et al. (2021a). Thus, it must be concluded that the specification of the prior distribution of β has a very large impact on the conclusions which can be drawn with respect to the sparsity or density of these particular data sets.

4.2.2 Theoretically relevant variables

In Figure 11 the plots of the prior and posterior distribution for different benchmarks of $\bar{z}$ are shown, where intuitively in Subfigure 11a a larger number of variables is always included than in for example Subfigure 11d. As discussed before and shown in Figure 1, in the original model of Giannone et al. (2021a), where no variables are switched to u_t, the Macro 1 data set has a relatively sparse representation. It is very interesting to see that when more than 1 variable is switched to u_t, as in Subfigures 11a, 11b and 11c, the model becomes significantly denser, with larger values for q and smaller values for γ. Only Subfigure 11d shows a reasonable resemblance to the figure of the original model. Thus, always including variables based on the criterium of a large posterior probability of inclusion leads to a denser representation. The results for the Macro 2 data set were not impacted much by always including some of the variables, the relevant figures are all shown in Appendix C. As the Macro 2 data set already has a very dense representation in the original model of Giannone et al. (2021a), it is difficult to say if in the Macro 2 data set always including certain variables based on the $\bar{z}$ benchmark also leads to a denser representation. An explanation for this interesting effect may lie in the negative relationship between q and γ. Due to this relationship, the variables with large posterior values of q would also have relatively low variance and thus would be shrunk quite heavily. Moving these variables to u_t and thus allowing them to escape this heavy shrinkage could force more shrinkage for other variables, allowing for more variables to be included and thus for a denser representation of the data.



(a) $\bar{z} \geq 0.80$

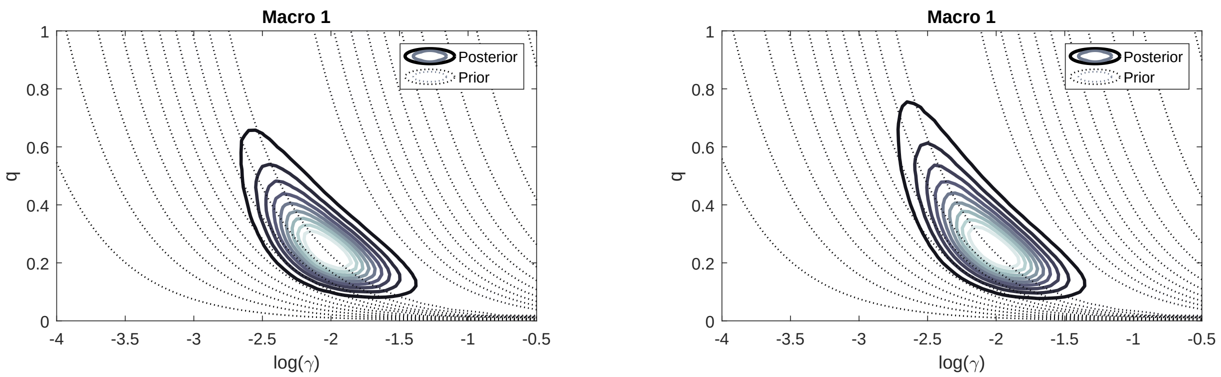
(b) $\bar{z} \geq 0.90$

(c) $\bar{z} \geq 0.95$

(d) $\bar{z} \geq 0.99$

Figure 11. Contours of the prior and posterior density of q and $\log(\gamma)$ for the model in which variables with values for $\bar{z}$ greater or equal than 0.80, 0.90, 0.95 and 0.99 are shown in Subfigure 11a, 11b, 11c and 11d respectively.

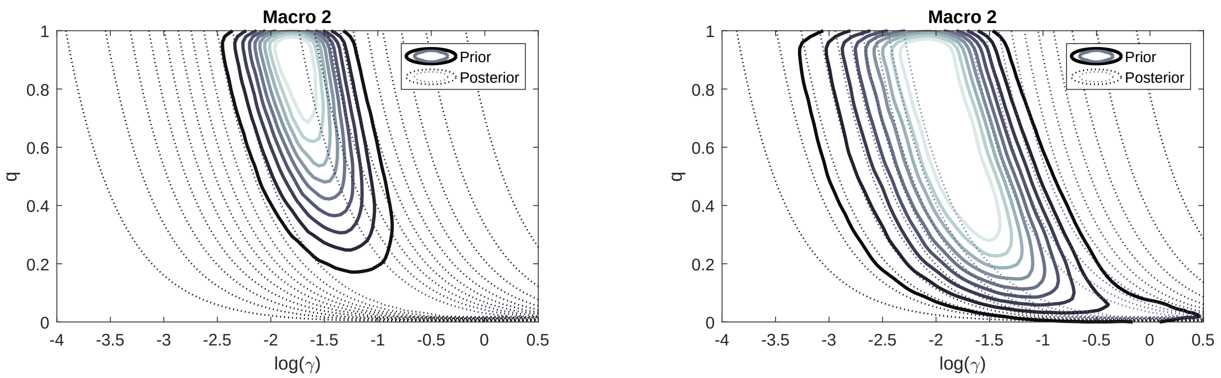
We have seen that always including certain variables into the model based on a benchmark for $\bar{z}$ has great effects on the density of the model for the Macro 1 data set, which raises the question whether this is a consequence of 'fixing' these variables as a whole or that variable choice is influential in this respect. To evaluate this, we make a comparison to the model in which we always include certain variables based on theoretical macroeconomic results. The plot of the prior and posterior distribution of q and γ is shown in Subfigure 12b, with the original plot of the Giannone et al. (2021a) model in Subfigure 12a for comparison. Specifically, keeping in mind the very large effects shown in Figure 11, it is very interesting to see that always including the variables based on theory leads to only minor changes in the posterior distribution when compared to the original model. The effects of always including variables based on theoretical arguments on the prior and posterior distribution of q and γ are more pronounced for the Macro 2 data set, as can be seen in Figure 13. Where the density plot of the original model is relatively concentrated around large values for 1, the plot in Subfigure 13b is drawn out over basically the full spectrum of values for q. This suggests that always including these variables seems to allow for sparser specifications, but that there is no real consistency in the level of sparsity or density.



(a) Original Model

(b) Model always including theoretical variables

Figure 12. Contours of the prior and posterior density of q and $\log(\gamma)$ for the original model of Giannone et al. (2021a) and the model in which the theoretically relevant variables are always included are shown in Subfigure 12a and 12b respectively



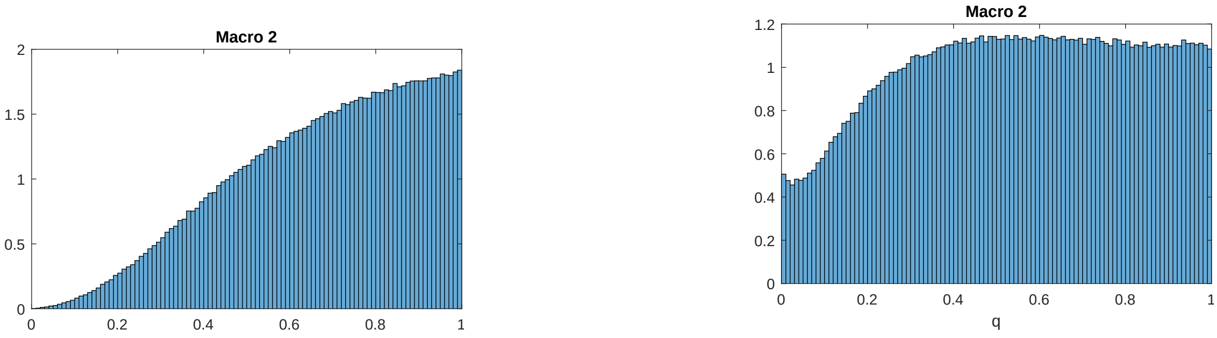
(a) Original Model

(b) Model always including theoretical variables

Figure 13. Contours of the prior and posterior density of q and $\log(\gamma)$ for the original model of Giannone et al. (2021a) and the model in which the theoretically relevant variables are always included are shown in Subfigure 13a and 13b respectively.

The odd behaviour of posterior distribution of q for the Macro 2 data when including theoretically relevant variables is clearly visible in Subfigure 14b. Firstly, the posterior density of q is relatively large for all values of q, specifically when compared to for example Subfigure 14a. The posterior density then increases with q up to a value of q of around 0.4, after which it levels out. It is hard to say what this means for the sparsity or density for this particular model, other than that all specifications which are at least somewhat dense seem to be suitable.

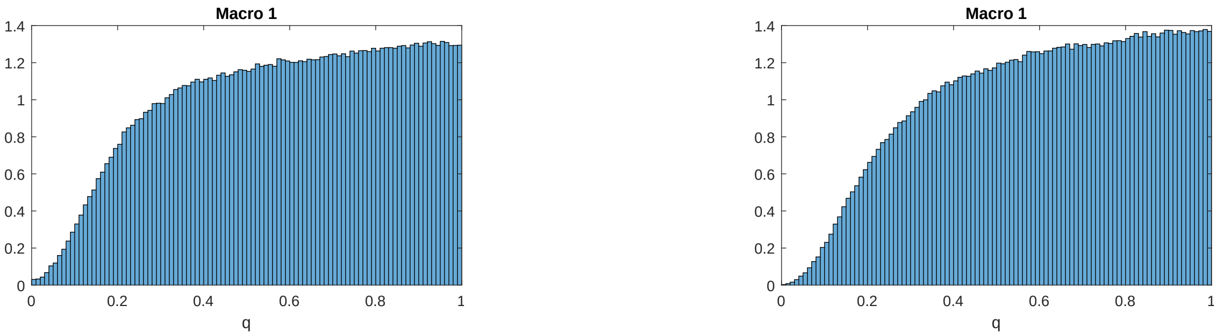
Figure 15 very clearly shows that always including variables based on the $\bar{z}$ benchmark leads to a very dense representation of the Macro 1 data set. For example Subfigures 15a and 15b are more reminiscent of the plot of the posterior distribution of q of the Macro 2 data set, which is decidedly dense.



(a) Original Model

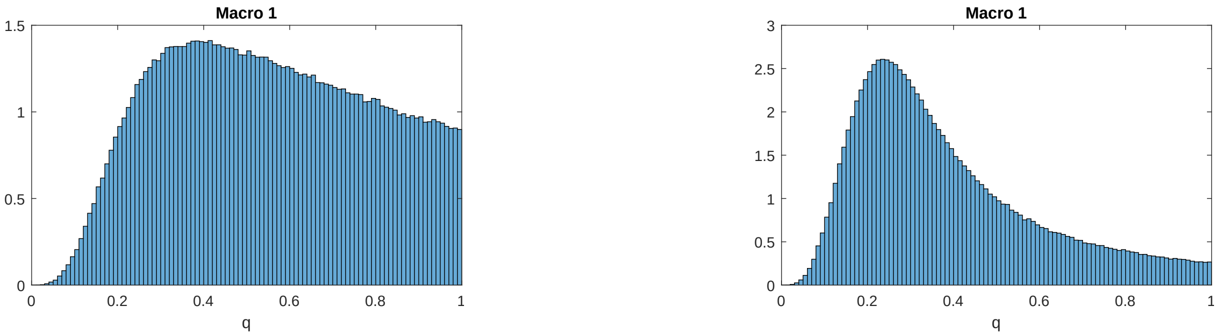
(b) Model always including theoretical variables

Figure 14. Posterior density of q for the original model of Giannone et al. (2021a) and the model in which the theoretically relevant variables are always included are shown in subfigure 14a and 14b respectively.



(a) $\bar{z} \geq 0.80$

(b) $\bar{z} \geq 0.90$

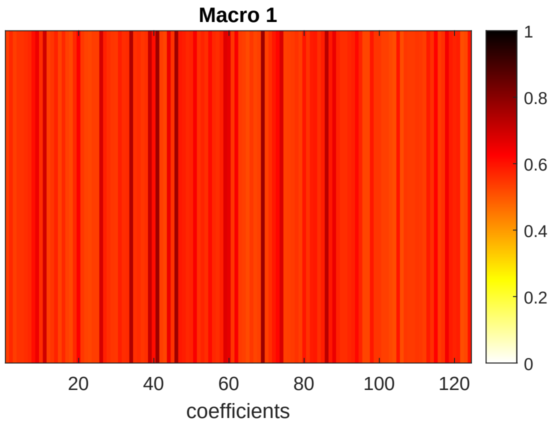


(c) $\bar{z} \geq 0.95$

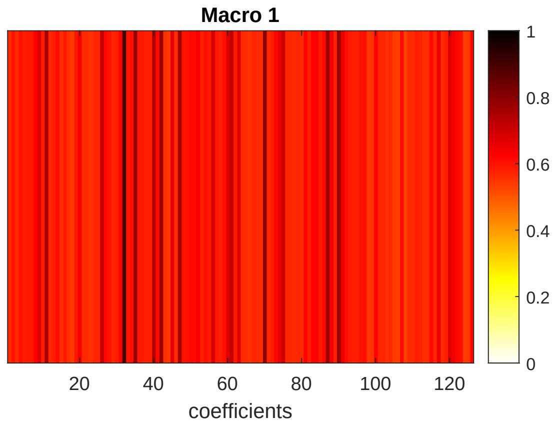
(d) $\bar{z} \geq 0.99$

Figure 15. Posterior density of q for the model in which variables with values for $\bar{z}$ greater or equal than 0.80, 0.90, 0.95 and 0.99 are shown in Subfigure 15a, 15b, 15c and 15d respectively.

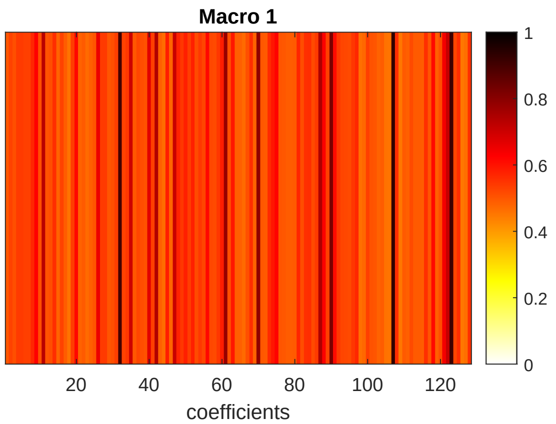
The heatmap of the probabilities of inclusion shown in Figure 16 further substantiates the findings regarding the Macro 1 data set with variables always included based on benchmark $\bar{z}$. For a lower benchmark, and thus for larger numbers of variables always included, the posterior probabilities of inclusion are relatively large for all variables and no clear pattern appears. As the benchmark becomes stricter, as in Subfigure 16d, the results more closely resemble the results obtained for the original model of Giannone et al. (2021a).



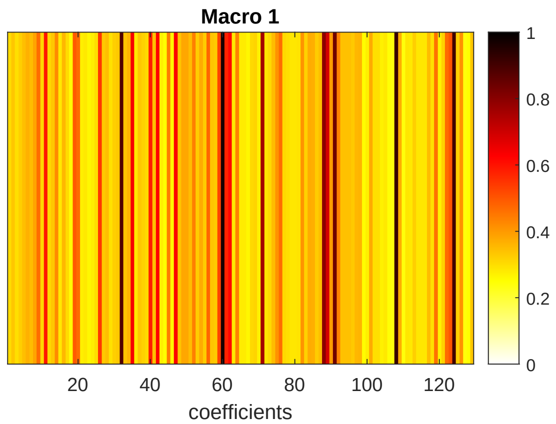
(a) $\bar{z} \geq 0.80$



(b) $\bar{z} \geq 0.90$



(c) $\bar{z} \geq 0.95$



(d) $\bar{z} \geq 0.99$

Figure 16. Heat map of the probabilities of inclusion of each predictor for the Macro 1 data set with values for $\bar{z}$ greater or equal than 0.80, 0.90, 0.95 and 0.99 are shown in Subfigure 16a, 16b, 16c and 16d respectively.

In conclusion, always including certain variables based on a benchmark of the average posterior probability of inclusion of a variable $\bar{z}$ does not very much affect the results obtained for the Macro 2 data set. However, the effect on the results of the Macro 1 data set is quite pronounced. Including only the one variable with the largest posterior value for $\bar{z}$ has a limited effect and does not affect the posterior of q or γ overly much when compared to the original model of Giannone et al. (2021a). However, when a lower benchmark for $\bar{z}$ is used, the Macro 1 data set strongly tends towards a very dense rather than sparse representation. Model uncertainty remains an issue in all cases.

The significant effect that always including these variables has on the conclusions regarding the density of the Macro 1 data set, makes the conclusions which result from always including certain variables selected based on macroeconomic theory even more interesting. Namely, when these variables – selected based on theory — are always included, the figures and conclusions concerning the Macro 1 data set hardly change when compared to the original model. Always including these theoretical variables in the Macro 2 model does rather affect the results compared to the original model, leading to a posterior distribution which is dense in the broadest sense of the word, where the histogram of the posterior distribution of q is practically flat for values ranging from 0.4 up to 1.

5 Conclusion

In this we replicate the results of Giannone et al. (2021a) and we confirm their original findings regarding both the density or sparsity of two macroeconomic data sets and the high degree of uncertainty with respect to variable selection, warranting the use of model averaging techniques in predictive exercises. We find that the prior distribution which is chosen for β in the framework of Giannone et al. (2021a) has a very large impact on the conclusions regarding the sparsity or density of these data sets. For example, setting a t-distribution with relatively low values for ν leads to models which show significantly more sparsity than the results found by Giannone et al. (2021a). Furthermore, we find that always including specific variables in the model by moving them from x_t to u_t has effects on the results depending on the decision rule used for variable selection. Choosing variables based on the posterior probability of inclusion of those specific variables leads to a dense representation, with the level of density increasing with the number of variables included. However, choosing variables to always include based on theoretical macroeconomic arguments has a negligible impact on the conclusions regarding sparsity or density for one of the data sets, while having indefinite effects on the other data set. Thus, always including certain variables into the model does not lead to more sparsity, and can rather induce a denser representation of the data.

Possible avenues for further research are evaluating the effects of setting other, theoretically sound, priors on β on the conclusions regarding sparsity or density of the data. Furthermore, not fixing the degrees of freedom ν of the prior t-distribution, but rather learning this value through the data would be an interesting topic to investigate. The effects of making changes in the prior distributions of the other parameters and hyperparameters and evaluating the effect of doing so on the conclusions are also noteworthy avenues for future research. Specifically defining a prior on ϕ which is not flat or normal in combination with moving certain variables to u_t could be interesting. Also, testing the effects of other decision rules for variable selection in this procedure and evaluating the differences in results is an interesting topic for future research.

Giannone et al. (2021a) concluded that there was an 'Illusion of sparsity'. In this paper we show that their results are sensitive to the relatively arbitrary prior choice on parameter β, casting doubt on the conclusions Giannone et al. (2021a) draw on the general sparsity or density of these macroeconomic data sets. Further research is necessary to reach a final and conclusive outcome in this discussion.

References

Barro, R. J. & Lee, J.-W. (1994). Sources of economic growth. In *Carnegie-rochester conference series on public policy* (Vol. 40, pp. 1–46).

Belloni, A., Chernozhukov, V. & Hansen, C. (2011). Inference for high-dimensional sparse econometric models. *arXiv preprint arXiv:1201.0220*.

Belloni, A., Chernozhukov, V. & Wang, L. (2011). Square-root lasso: pivotal recovery of sparse signals via conic programming. *Biometrika*, *98*(4), 791–806.

Betancourt, M. (2021). *Prior Modeling.* Retrieved 11-05-2023, from `https://betanalpha.github.io/assets/case_studies/prior_modeling.html`

Canova, F. (2007). *Methods for applied macroeconomic research* (Vol. 13). Princeton university press.

Carvalho, C. M., Polson, N. G. & Scott, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika*, *97*(2), 465–480.

Chan, J. C. (2017). Notes on bayesian macroeconometrics. *Manuscript available at http://joshuachan. org*.

Chernozhukov, V., Hansen, C. & Liao, Y. (2017). A lava attack on the recovery of sums of dense and sparse signals.

Christiano, L. J., Eichenbaum, M. & Evans, C. L. (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of political Economy*, *113*(1), 1–45.

Del Negro, M. (2011). Bayesian macroeconometrics.

Del Negro, M. & Schorfheide, F. (2004). Priors from general equilibrium models for vars. *International Economic Review*, *45*(2), 643–673.

Doan, T., Litterman, R. & Sims, C. (1984). Forecasting and conditional projection using realistic prior distributions. *Econometric reviews*, *3*(1), 1–100.

Estrella, A. & Hardouvelis, G. A. (1991). The term structure as a predictor of real economic activity. *The journal of Finance*, *46*(2), 555–576.

Fava, B. & Lopes, H. F. (2021). The illusion of the illusion of sparsity: An exercise in prior sensitivity. *Brazilian Journal of Probability and Statistics*, *35*(4), 699–720.

Fernandez, C., Ley, E. & Steel, M. F. (2001). Model uncertainty in cross-country growth regressions. *Journal of applied Econometrics*, *16*(5), 563–576.

Gelman, A., Simpson, D. & Betancourt, M. (2017). The prior can often only be understood in the context of the likelihood. *Entropy*, *19*(10), 555.

Giannone, D., Lenza, M. & Primiceri, G. E. (2019). Priors for the long run. *Journal of the American Statistical Association*, *114*(526), 565–580.

Giannone, D., Lenza, M. & Primiceri, G. E. (2021a). Economic predictions with big data: The illusion of sparsity. *Econometrica*, *89*(5), 2409–2437.

Giannone, D., Lenza, M. & Primiceri, G. E. (2021b). Supplement to "economic predictions with big data: The illusion of sparsity". *Econometrica*, *89*(5), 2409–2437.

Griffin, J. E. & Brown, P. J. (2005). Alternative prior distributions for variable selection with very many more variables than observations.

Gruber, L. & Kastner, G. (2022). Forecasting macroeconomic data with bayesian vars: Sparse or dense? it depends! *arXiv preprint arXiv:2206.04902*.

Haavelmo, T. (1944). The probability approach in econometrics. *Econometrica: Journal of the Econometric Society*, iii–115.

Hendry, D. F. (2018). Deciding between alternative approaches in macroeconomics. *International Journal of Forecasting*, *34*(1), 119–135.

Hendry, D. F. & Johansen, S. (2015). Model discovery and trygve haavelmo's legacy. *Econometric Theory*, *31*(1), 93–114.

Hoerl, A. E. & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55–67.

Ingram, B. F. & Whiteman, C. H. (1994). Supplanting the 'minnesota'prior: Forecasting macroeconomic time series using real business cycle model priors. *Journal of Monetary Economics*, *34*(3), 497–510.

Karlsson, S. (2013). Forecasting with bayesian vector autoregression. *Handbook of economic forecasting*, *2*, 791–897.

King, R. G., Plosser, C. I. & Rebelo, S. T. (1988). Production, growth and business cycles: Ii. new directions. *Journal of Monetary Economics*, *21*(2-3), 309–341.

Koop, G. (2017). Bayesian methods for empirical macroeconomics with big data. *Review of Economic Analysis*, *9*(1), 33–56.

Li, Q. & Lin, N. (2010). The bayesian elastic net.

Litterman, R. B. (1980). *Bayesian procedure for forecasting with vector autoregressions.* Massachusetts Institute of Technology.

Litterman, R. B. (1986). Forecasting with bayesian vector autoregressions—five years of experience. *Journal of Business & Economic Statistics*, *4*(1), 25–38.

Litterman, R. B. et al. (1979). *Techniques of forecasting using vector autoregressions* (Tech. Rep.).

McCracken, M. W. & Ng, S. (2016). Fred-md: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, *34*(4), 574–589.

Park, T. & Casella, G. (2008). The bayesian lasso. *Journal of the American Statistical Association*, *103*(482), 681–686.

Piironen, J. & Vehtari, A. (2017). Sparsity information and regularization in the horseshoe and other shrinkage priors.

Plosser, C. I. & Rouwenhorst, K. G. (1994). International term structures and real economic growth. *Journal of monetary economics*, *33*(1), 133–155.

Raman, S., Fuchs, T. J., Wild, P. J., Dahl, E. & Roth, V. (2009). The bayesian group-lasso for analyzing contingency tables. In *Proceedings of the 26th annual international conference on machine learning* (pp. 881–888).

Shi, S., Phillips, P. C. & Hurn, S. (2018). Change detection and the causal impact of the yield curve. *Journal of Time Series Analysis*, *39*(6), 966–987.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica: journal of the Econometric Society*, 1–48.

Stock, J. H. & Watson, M. W. (1989). Interpreting the evidence on money-income causality. *Journal of Econometrics*, *40*(1), 161–181.

Stock, J. H. & Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American statistical association*, *97*(460), 1167–1179.

Team, S. D. (2017). Prior choice recommendations.

Theil, H. & Goldberger, A. S. (1961). On pure and mixed statistical estimation in economics. *International Economic Review*, *2*(1), 65–78.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288.

Tibshirani, R. J. (2011). *The solution path of the generalized lasso*. Stanford University.

Van Erp, S., Oberski, D. L. & Mulder, J. (2019). Shrinkage priors for bayesian penalized regression. *Journal of Mathematical Psychology*, *89*, 31–50.

Xu, X. & Ghosh, M. (2015). Bayesian variable selection and estimation for group lasso.

Yuan, M. & Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *68*(1), 49–67.

Zou, H. & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, *67*(2), 301–320.

Appendix A Derivation of posterior distribution ϕ with normal prior

Building on the Appendix of Giannone et al. (2021a), the posterior distribution of the unknown objects in the model with this new prior on ϕ is given by
$p(\phi, \beta, \sigma^2, R^2, z, q|Y, U, X)$

$\propto p(Y|U, X, \phi, \beta, \sigma^2, R^2, z, q) \cdot p(\phi, \beta, \sigma^2, R^2, z, q)$

$\propto p(Y|U, X, \phi, \beta, \sigma^2) \cdot p(\beta|\sigma^2, R^2, z, q) \cdot p(z|q, \sigma^2, R^2) \cdot p(\phi|\sigma^2) \cdot p(q) \cdot p(\sigma^2) \cdot p(R^2)$

$\propto \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{T}{2}} e^{-\frac{1}{2\sigma^2}(Y-U\phi-X\beta)'(Y-U\phi-X\beta)} \cdot \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{k}{2}} e^{-\frac{1}{2\sigma^2}(\phi-\mathbf{1})'(\phi-\mathbf{1})} \cdot \prod_{i=1}^{k} \left[\left(\frac{1}{2\pi\sigma^2\gamma^2}\right)^{\frac{1}{2}} e^{-\frac{\beta_i^2}{2\sigma^2\gamma^2}} \right]^{z_i} [\delta(\beta_i)]^{1-z_i}$
$\cdot \prod_{i=1}^{k} q^{z_i}(1-q)^{1-z_i} \cdot q^{a-1}(1-q)^{b-1} \cdot \left(\frac{1}{\sigma^2}\right) \cdot (R^2)^{A-1}(1-R^2)^{B-1},$

with $\gamma^2 = \frac{1}{k\bar{v}_x q} \cdot \frac{R^2}{1-R^2}$, and $\delta(\cdot)$ is the Dirac-delta function. We sample from the posterior of $(\phi, \beta, \sigma^2, R^2, z, q)$with a Gibbs sampling algorithm with blocks (i) (R^2, q), (ii) ϕ and (iii) (z, β, σ^2). The conditional posteriors of (R^2, q) and (z, β, σ^2) do not change as a result of the new prior on ϕ and are thus the same as described in the appendix of Giannone et al. (2021a).

The conditional posterior of ϕ is now given by

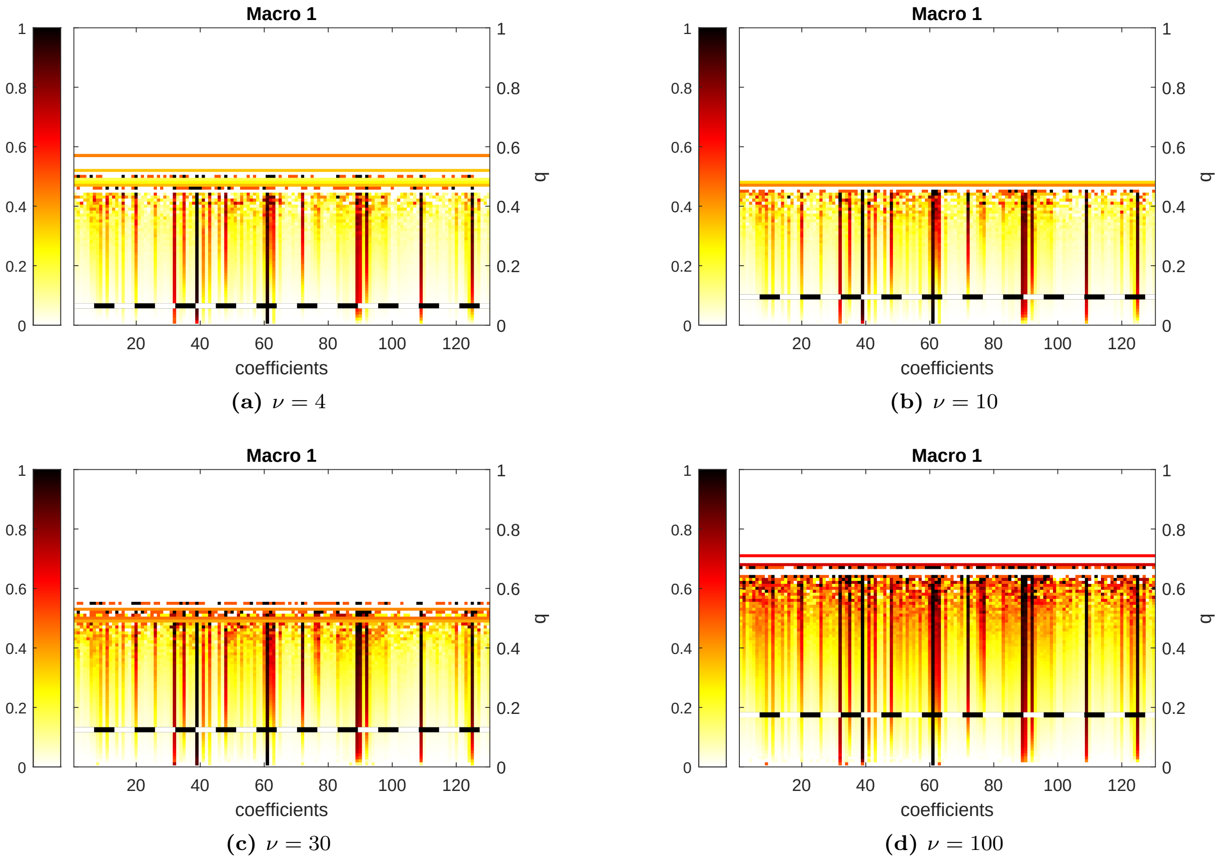
$p(\phi|Y, U, X, z, \beta, R^2, q, \sigma) \propto e^{-\frac{1}{2\sigma^2}(Y-U\phi-X\beta)'(Y-U\phi-X\beta)} \cdot e^{-\frac{1}{2\sigma^2}(\phi-\mathbf{1})'(\phi-\mathbf{1})}$,

which implies that

$\phi|Y, U, X, z, \beta, \gamma, q, \sigma \sim \mathcal{N}((U'U + \mathcal{I})^{-1}(U'Y - U'X\beta + \mathcal{I}), \sigma^2(U'U + \mathcal{I})^{-1})$

As was shown by Giannone et al. (2021a), the posterior of $\phi|Y, U, X, z, \beta, \gamma, q, \sigma$ with a flat prior on ϕ can be sampled from the normal distribution. The conjugate prior for the normal likelihood is a normal distribution, thus it is no surprise that the posterior of $\phi|Y, U, X, z, \beta, \gamma, q, \sigma$ using a normal distribution as prior on ϕ can also be sampled from a normal distribution. In practice the difference between sampling from these two posterior distributions is marginal, having no significant effect on the results.

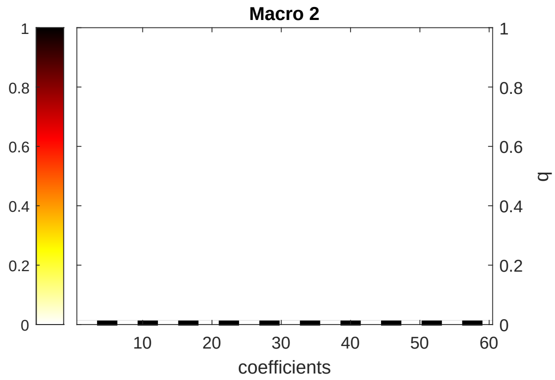
Appendix B Figures with results using a t-distribution as prior on β



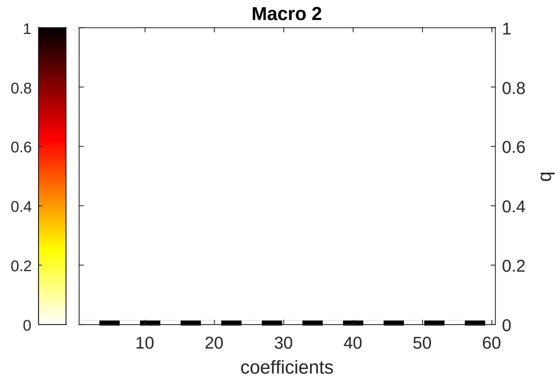
(a) $\nu = 4$ (b) $\nu = 10$

(c) $\nu = 30$ (d) $\nu = 100$

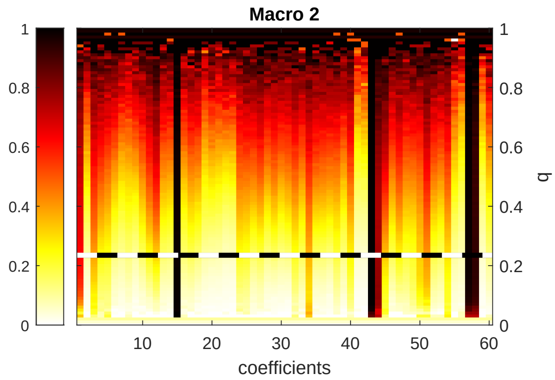
Figure 17. Heat map of the probabilities of inclusion of each predictor, conditional on q. The horizontal dashed line denotes the posterior mode. Results for the Macro 1 data set with a prior with t-distribution with degrees of freedom ν equal to 4, 10, 30 and 100 in subfigure 17a, 17b, 17c and 17d respectively.



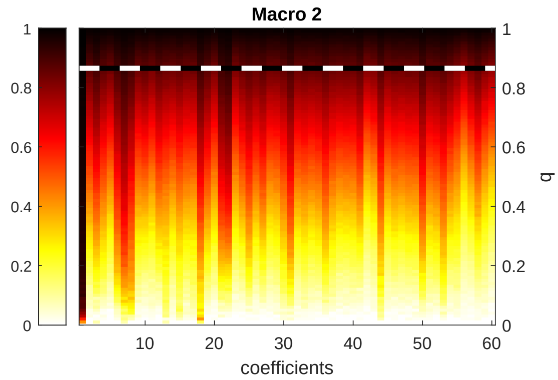
(a) $\nu = 4$



(b) $\nu = 30$



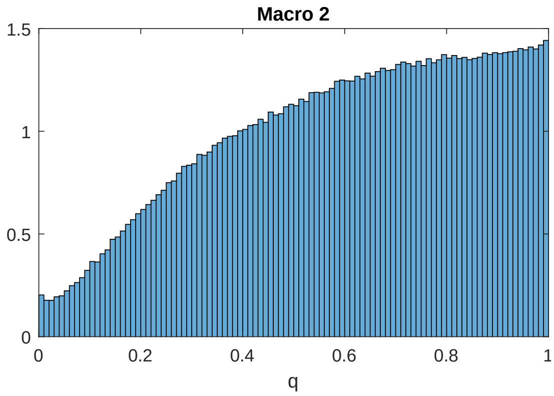
(c) $\nu = 100$



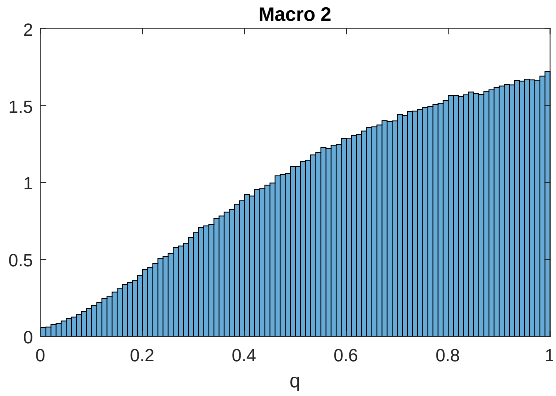
(d) $\nu = 2000$

Figure 18. Heat map of the probabilities of inclusion of each predictor, conditional on q. The horizontal dashed line denotes the posterior mode. Results for the Macro 2 data set with a prior with t-distribution with degrees of freedom ν equal to 4, 30, 100 and 2000 in subfigure 18a,18b, 18c and 18d respectively.

Appendix C Figures with results when theoretically relevant variables are always included

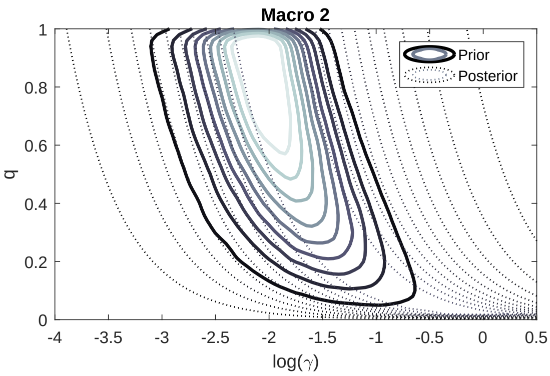


(a) $\bar{z} \geq 0.80$

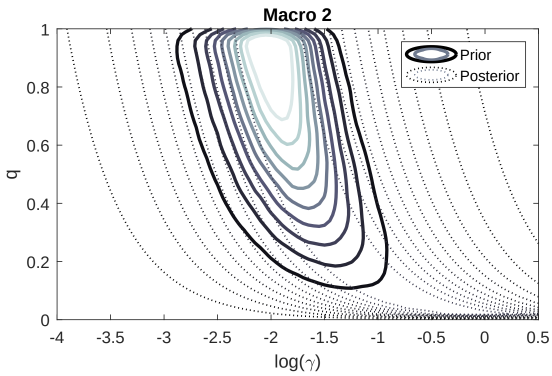


(b) $\bar{z} \geq 0.95$

Figure 19. Posterior density of q for the model in which variables with values for $\bar{z}$ greater or equal than 0.80 and 0.95 are shown in subfigure 19a and 19b respectively.

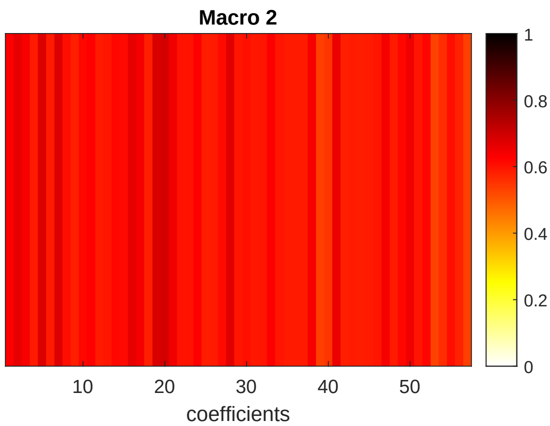


(a) $\bar{z} \geq 0.80$

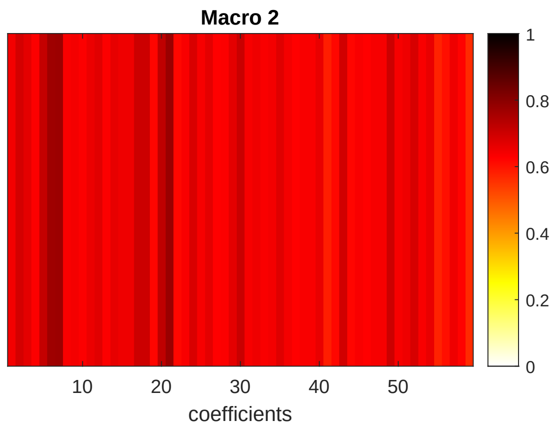


(b) $\bar{z} \geq 0.95$

Figure 20. Contours of the prior and posterior density of q and $\log(\gamma)$ for the model in which variables with values for $\bar{z}$ greater or equal than 0.80 and 0.95 are shown in subfigure 20a and 20b respectively.

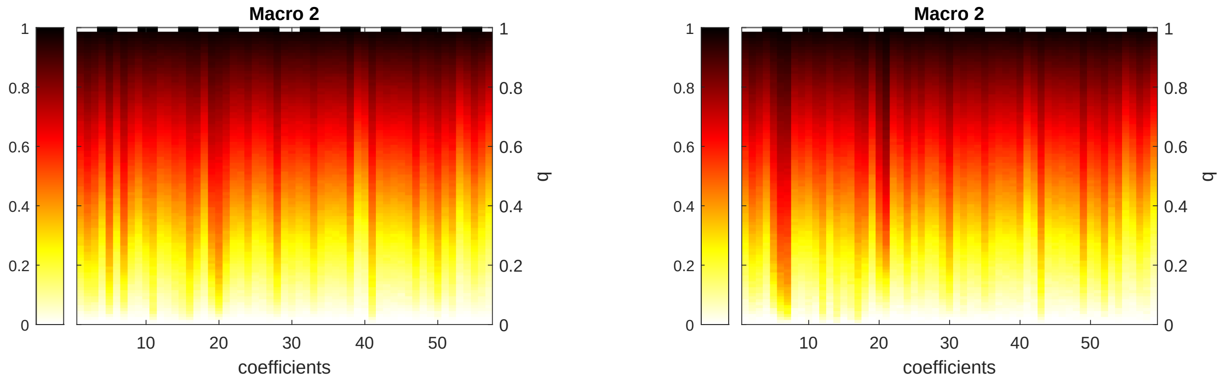


(a) $\bar{z} \geq 0.80$



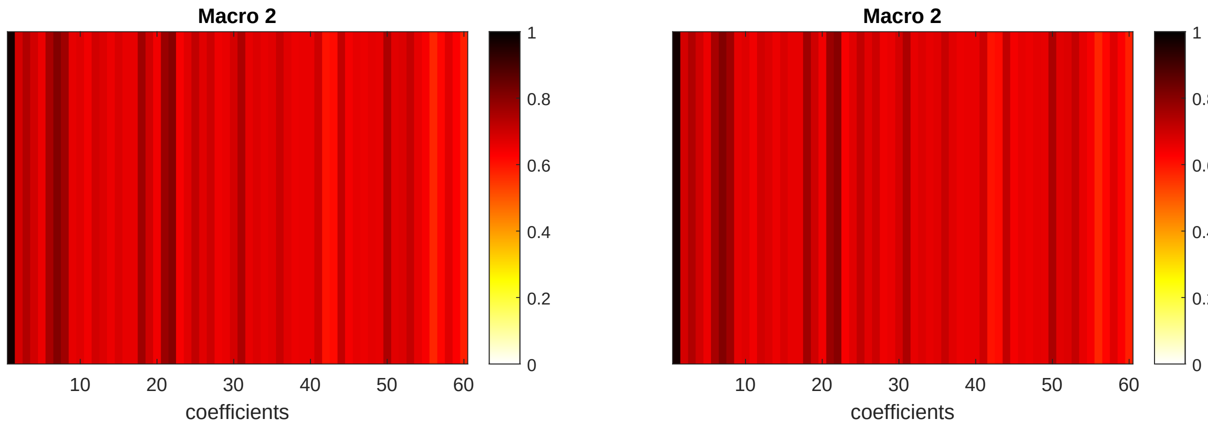
(b) $\bar{z} \geq 0.95$

Figure 21. Heat map of the probabilities of inclusion of each predictor for the Macro 1 data set with values for $\bar{z}$ greater or equal than 0.80 and 0.95 are shown in subfigure 21a and 21b respectively.



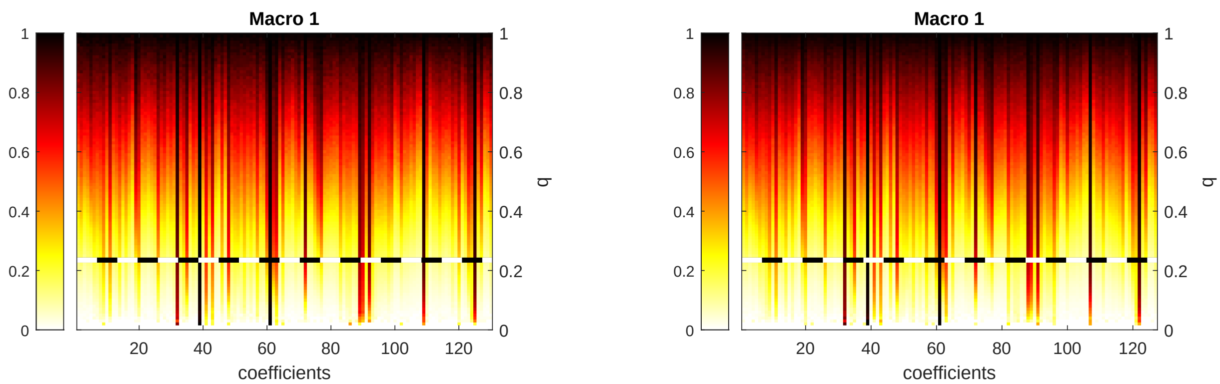
(a) $\bar{z} \geq 0.80$ **(b)** $\bar{z} \geq 0.95$

Figure 22. Heat map of the probabilities of inclusion of each predictor, conditional on q. The horizontal dashed line denotes the posterior mode. Results for the Macro 2 data set with values for $\bar{z}$ greater or equal than 0.80 and 0.95 are shown in subfigure 22a and 22b respectively.



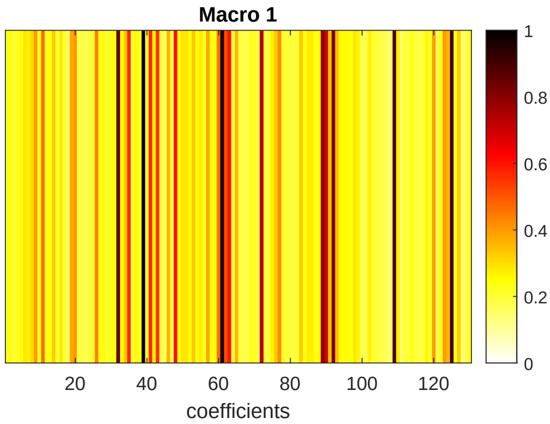
(a) Original Model **(b)** Model always including theoretical variables

Figure 23. Heat map of the probabilities of inclusion of each predictor for the original model of Giannone et al. (2021a) and the model in which the theoretically relevant variables are always included are shown in subfigure 23a and 23b respectively.

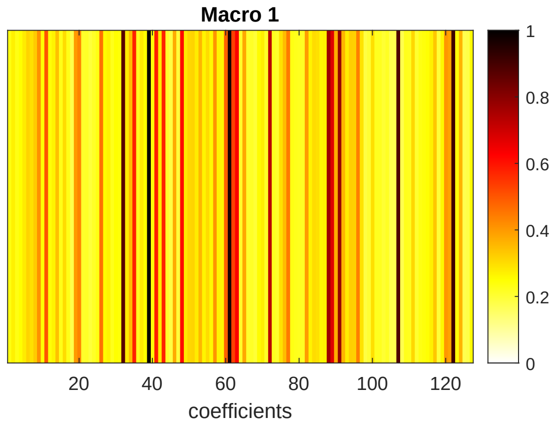


(a) Original Model **(b)** Model always including theoretical variables

Figure 24. Heat map of the probabilities of inclusion of each predictor for the original model of Giannone et al. (2021a) and the model in which the theoretically relevant variables are always included are shown in subfigure 24a and 24b respectively.

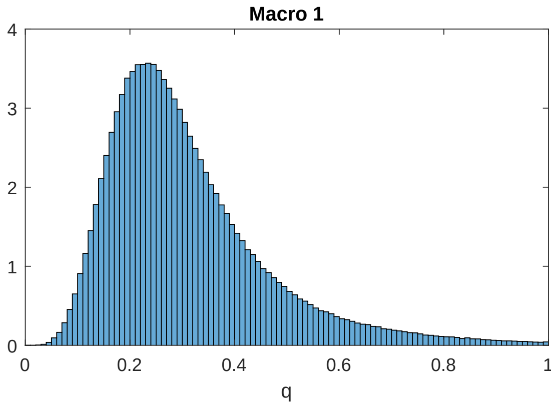


(a) Original Model

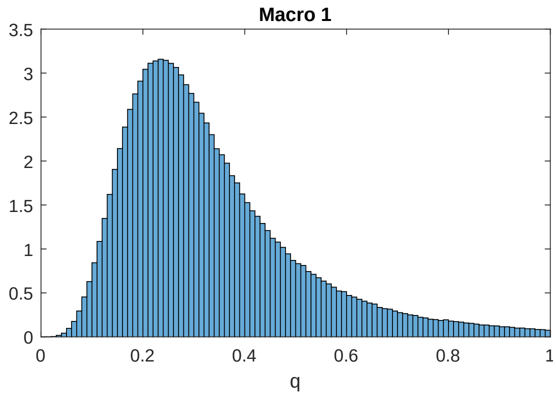


(b) Model always including theoretical variables

Figure 25. Heat map of the probabilities of inclusion of each predictor for the original model of Giannone et al. (2021a) and the model in which the theoretically relevant variables are always included are shown in subfigure 25a and 25b respectively.

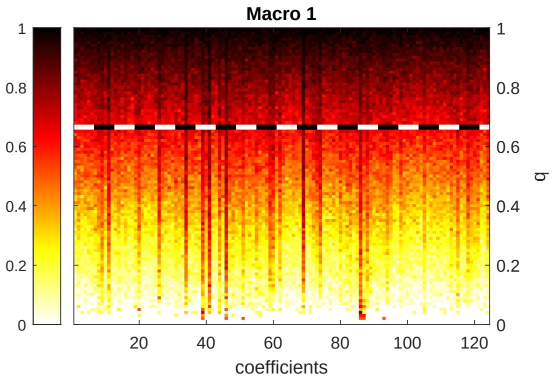


(a) Original Model

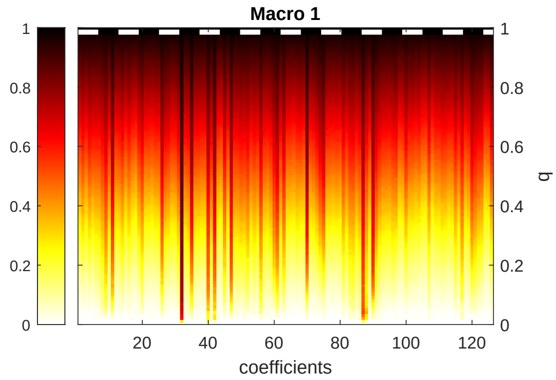


(b) Model always including theoretical variables

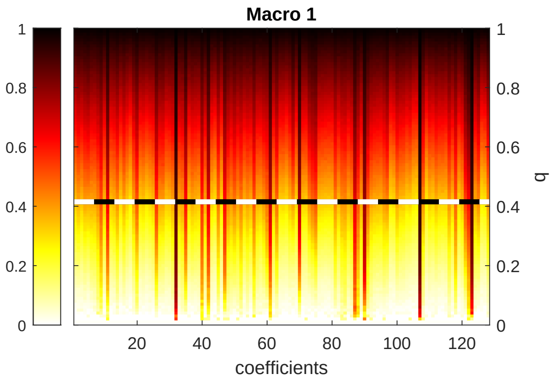
Figure 26. Posterior density of q for the original model of Giannone et al. (2021a) and the model in which the theoretically relevant variables are always included are shown in subfigure 26a and 26b respectively.



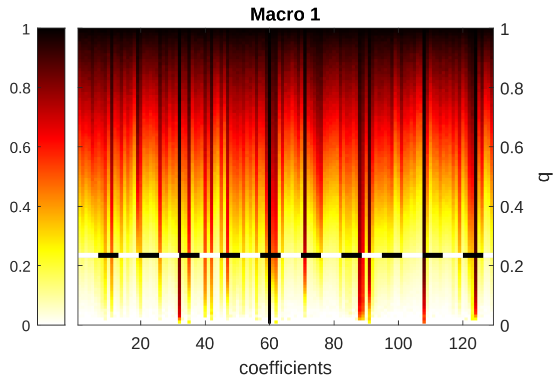
(a) $\bar{z} \geq 0.80$



(b) $\bar{z} \geq 0.90$

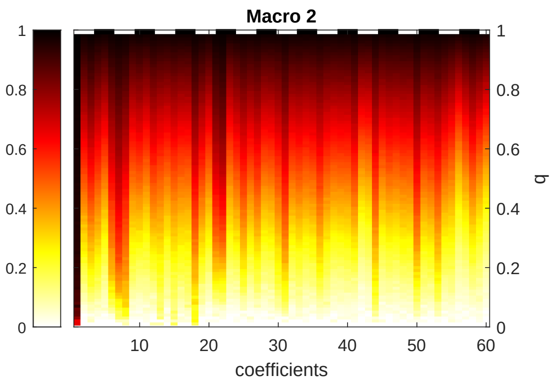


(c) $\bar{z} \geq 0.95$

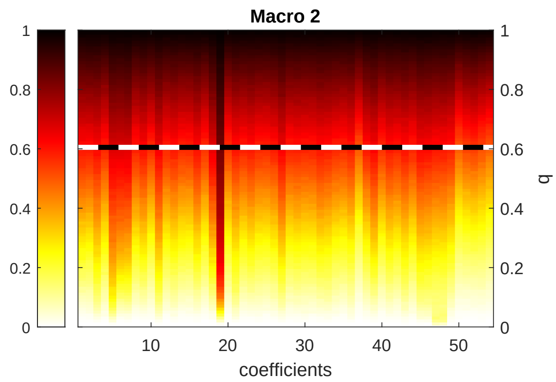


(d) $\bar{z} \geq 0.99$

Figure 27. Heat map of the probabilities of inclusion of each predictor, conditional on q. The horizontal dashed line denotes the posterior mode. Results for the Macro 1 data set with values for $\bar{z}$ greater or equal than 0.80, 0.90, 0.95 and 0.99 are shown in subfigure 27a, 27b, 27c and 27d respectively.



(a) Original Model



(b) Model always including theoretical variables

Figure 28. Heat map of the probabilities of inclusion of each predictor for the original model of Giannone et al. (2021a) and the model in which the theoretically relevant variables are always included are shown in subfigure 28a and 28b respectively.

Appendix D Description of Matlab code

All coding and all runs are done in Matlab version R2020b on Windows.

D.1 Data

The raw data of the Macro 1 and Macro 2 data set is obtained from the replication material of Giannone et al. (2021a) and transformed using the files *generate_freddataGLP.m* and *GenCvMacro2.m* for the Macro 1 and Macro 2 data set respectively. The input files for these codes are csv file *2016-04.csv* and excel file *GrowthChern.xlsx* and output files are *FredMDlargeHor1.mat* and *GrowthData.mat* respectively. These Matlab data files are used to obtain the estimation results.

D.2 Estimation

All code used to obtain estimation results loads the *FredMDlargeHor1.mat* and *GrowthData.mat* files explained in section D.1 for results relevant to the Macro 1 and Macro 2 data set respectively.

To obtain the replication results of the work of Giannone et al. (2021a) for the results pertaining to the Macro 1 and Macro 2 data sets, *macro1.m* and *macro2.m* are run, which results in the output files *macro1_PosteriorDraws_474613bb.mat* and *Macro2_PosteriorDraws_474613bb.mat* respectively. These codes make use of the algorithm of *SpikeSlabGLP.m*

To obtain the extension results in which certain variables are moved from x_t to u_t the files *macro1_fix_variables.m* and *macro2_fixvariables.m* are run. These codes make use of the algorithm of *SpikeslabGLP.m*

macro1_fix_variables.m is run five times, each time a different set of variables is moved from x_t to u_t and the results are saved under a different name accordingly. This results in output files *macro1_PosteriorDraws_474613bb_fixvariables_80.mat, ..._90.mat, ..._95.mat, ..._99.mat* and *..._theory.mat* for the benchmarks of $\bar{z}$ equal to 80, 90, 95 and 99 and the theoretically selected variables respectively.

macro2_fixvariables.m is run three times, again each time a different set of variables is moved from x_t to u_t and the results are saved under a different name accordingly. This results in output files *Macro2_PosteriorDraws_474613bb_fixvariables_80.mat, ..._95.mat* and *..._theory.mat* for the benchmarks of $\bar{z}$ equal to 80, 95 and the theoretically selected variables respectively.

To obtain the extension results in which the prior of β is changed to the t-distribution, the files *macro1_tdistribution.m* and *macro2_tdistribution.m* are run. These codes make use of the algorithm of *SpikeslabGLP_tDistribution.m.*

macro1_tdistribution.m is run four times, where each time a different value for ν is set and the results are saved under a different name accordingly. This results in output files *macro1_PosteriorDraws_474613bb_tDistribution_nu4.mat, ..._nu10.mat, ..._nu30.mat* and *..._nu100.mat* for values of ν equal to 4, 10, 30 and 100 respectively.

macro2_tdistribution.m is run five times, where each time a different value for ν is set and the results are saved under a different name accordingly. This results in output files *macro2_PosteriorDraws_474613bb_tDistribution_nu4.mat, ..._10.mat, ..._30.mat, ..._100.mat* and *..._2000.mat* for values of ν equal to 4, 10, 30, 100 and 2000 respectively.

D.3 Figures

We use four types of figure in this paper. For the figures used in the replication the following codes are used: *Figure41_PriorPosterior.m, Figure42_qPosterior.m, Figure43_heatmap.m* and *Figure44_ConditionalHeatmaps.m*, which correspond to the plots of the prior and posterior density of q and $\log(\gamma)$, the posterior density of q, the heatmap of the probability of inclusion of variables and the heatmap of the probability of inclusion of variables conditional on q respectively. As an input the estimation results of the replication discussed above are used.

The figures for the extension in which certain variables are moved from x_t to u_t are generated by running the files *Figure41_PriorPosterior_fixVariables.m, Figure42_qPosterior_fixVariables.m, Figure43_heatmap_fixVariables.m* and *Figure44_ConditionalHeatmaps_fixVariables.m.*For the figures of Macro 1 and Macro 2 these files are are run respectively five and three times using the different input files discussed in section D.2.

The figures for the extension in which the prior of β is changed to the t-distribution are generated by running the files *Figure41_PriorPosterior_tDistribution.m, Figure42_qPosterior_tDistribution.m, Figure43_heatmap_tDistribution.m* and *Figure44_ConditionalHeatmaps_tDistribution.m.* For the figures of Macro 1 and Macro 2 these files are run respectively four and five times using the different input files discussed in section D.2.