# Bayesian forecasting with highly correlated predictors

Dimitris Korobilis
University of Glasgow

July 2012

**Abstract**

This paper considers Bayesian variable selection in regressions with a large number of possibly highly correlated macroeconomic predictors. I show that by acknowledging the correlation structure in the predictors can improve forecasts over existing popular Bayesian variable selection algorithms.

**Keywords:** Bayesian semiparametric selection; Dirichlet process prior; correlated predictors; clustered coefficients

**JEL Classification:** C11, C14, C32, C52, C53

# 1 Introduction

Many empirical problems in economics involve regressions where many predictors (possibly more than the number of available observations) are available, of which only a limited set is relevant for forecasting and policy analysis. An integrated way to deal with such demanding statistical inference is to use Bayesian simulation algorithms to estimate posterior probabilities of importance of each economic predictor based on evidence in the data. These algorithms perform variable selection (i.e. selecting the predictors with probability higher than 0.5) as well as model averaging (i.e. using all available predictors scaled by their respective probability). A popular application of Bayesian variable selection and model averaging is in the problem of identifying determinants of economic growth (Fernandez, Ley and Steel, 2001). Other studies try to determine which macroeconomic fundamentals help predict exchange rates (Wright, 2008), inflation (Koop and Korobilis, 2012), or which stock market characteristics drive stock returns (Cremers, 2002).

The purpose of this paper is to evaluate variable selection and model averaging, in the presence of many highly correlated predictors in forecasting regression models. In particular, I consider 183 quarterly macroeconomic predictors for forecasting output and inflation, in a setting similar to the one used by authors such as Stock and Watson (1999, 2002). Such datasets have many variables which are disaggregates of major macroeconomic series, such as employment and industrial production in different production sectors, or the various components of GDP. There can be high correlation within a set of disaggregated series, but also between different sets of series[1].

Given this particular structure of the data, in this note I examine the properties of the semiparametric variable selection prior proposed by Dunson et al. (2008) which allows for simultaneous selection of important predictors and soft clustering of predictors having similar impact on the variable of interest. This prior is a generalization of the typical "spike and slab" priors used for Bayesian variable selection and model averaging in the statistics literature; see George, Sun and Ni (2008) and Korobilis (2012) for recent applications in economics. In an exercise involving forecasting short-run (up to four quarters) inflation and output with more predictors than observations, I find that the semiparametric variable selection prior improves over the more traditional spike and slab prior, and is superior to principal components analysis for this particular problem.

The paper is structured as follows: Section 2 presents the model; Section 3 describes the dataset and forecasting results; Section 4 concludes.

[1] In the dataset used in this paper, the correlation coefficient of employment in durable goods and employment in nondurable goods manufacturing is 0.81, while employment in durable goods and total industrial production have correlation of 0.84.

# 2 Methodology

## 2.1 Spike and slab priors for variable selection

The majority of empirical macroeconomic forecasting models involve estimating dynamic regressions of the form

$$y_{t+h} = \gamma + \sum_{i=1}^{p} \varphi_i y_{t-(i-1)} + x_t \beta + \varepsilon_{t+h}, \tag{1}$$

where $y_{t+h}$ is the variable of interest which we want to forecast, $y_{t-i+1}$ are the $p$ own lags of $y$ for $i = 1, ..., p$, $x_t$ is a $(K \times 1)$ vector of exogenous predictors, and $\varepsilon_{t+h}$ is a Gaussian forecast error with zero mean and variance $\sigma^2$. In the remainder of this paper I assume that the intercept and two lags are always included in the forecasting model. For that reason, the regression coefficients $\theta = (\gamma, \varphi_1, \varphi_2)$ as well as the variance $\sigma^2$ admit noninformative priors of the form

$$\begin{aligned} \theta &\sim N\left(0_{3\times 1}, 100 I_3\right) \\ \sigma^2 &\sim iGamma\left(0.01, 0.01\right). \end{aligned}$$

When $K$ becomes "large", Stock and Watson (2002) suggest to use shrinkage based on replacing $x_t$ by its first few principal components, while other authors (Cremers, 2002; Koop and Potter, 2004) stress the benefit of selecting the best, according to some criterion, variables/predictors. Among several Bayesian algorithms developed, a popular method for variable selection is the spike and slab prior for the coefficients $\beta$, which was formalized by Mitchell and Beauchamp (1988) and is of the form

$$\beta_j \sim \pi \delta_0\left(\beta\right) + \left(1 - \pi\right) N\left(0, \tau^2\right), \tag{2}$$

where $\delta_a\left(v\right)$ is the Dirac delta function for random variable $v$ which places all probability mass on the point $a$. Thus, the prior for $\beta_j$, $j = 1, ..., K$, is a mixture of a point mass at zero (the spike) and a locally uninformative (depending on how large the value of $\tau^2$ is) Gaussian prior. The probabilities $\pi$ are random variables updated by the data and they determine whether the prior of $\beta_j$ is restricted to be zero, or whether it comes from the unrestricted Gaussian density with variance $\tau^2$. As is the case with other popular model selection and averaging priors (for instance the $g$-prior; see Koop and Potter, 2004), this prior does not explicitly model the correlation structure in the data when determining which variables are restricted to enter the regression. In fact, in many cases authors orthogonalize their predictors $x_t$ in order to speed-up convergence of the posterior sampling algorithm, thus ignoring completely correlations.

## 2.2 Semiparametric spike and slab prior

Given the considerations above, and the structure of the datasets customarily used by macroeconomists, the simple spike and slab prior can be reformulated in order to account for correlations in the data. An interesting extension has been proposed by Dunson et al. (2008); see also MacLehose et al. (2007). In these papers, the coefficients $\beta$ admit a prior of the form

$$\beta_j \sim \pi\delta_0(\beta) + (1-\pi)G \tag{3}$$

$$G \sim DP(\alpha G_0) \tag{4}$$

$$G_0 \sim N\left(0, \tau^2\right). \tag{5}$$

In this formulation $G$ is a nonparametric density which follows a Dirichlet process with base measure $G_0$ and concentration parameter $\alpha$. Usually $G_0$ is chosen to be a well-known density, for instance the Gaussian, making the prior an infinite mixture of the densities $G_0$. Hence, priors like this are "pseudo-nonparametric", since a parametric mixture of distributions is used to approximate the unknown density $G$. In this case the base measure $G_0$ is Gaussian with zero mean and variance $\tau^2$, which is the typical conjugate prior distribution used on linear regression coefficients. Hence, this prior implies that each coefficient $\beta_j$ will either be restricted to 0 with probability $\pi$, or with probability $(1-\pi)$ will come from a mixture of Gaussian densities.

Thus, this prior allows for calculation of Bayesian posterior probabilities of the hypothesis $H_{0j} : \beta_j = 0$ against $H_{1j} : \beta_j \neq 0$, while clustering the $j$'s for the non-null predictors. The clustering effect comes as a property of the Dirichlet process: $\beta_j$'s coming from the same Gaussian mixture component, will share the same mean and variance. As an example, consider coefficients $\beta_j$, $j = 1, ..., 6$ which are distributed according to $(\beta_1, \beta_3) \sim N\left(0, 10^6\right)$, $(\beta_2, \beta_4) \sim N(0, 0.1)$ and $(\beta_5, \beta_6) \sim \delta_0$. In this specific example $(\beta_1, \beta_3)$ are clustered together and come from a Gaussian with variance $10^6$, hence the posterior mean/median of these coefficients is close to the value of the LS estimator. The second cluster consists of coefficients $(\beta_2, \beta_4)$ which have prior variance 0.1, hence their posterior median will be equivalent to a ridge regression estimator. Finally, $(\beta_5, \beta_6)$ are restricted to be zero, so that $x_{5,t}$ and $x_{6,t}$ are completely irrelevant for forecasting $y_{t+h}$. Hence, this example shows that this prior is a hybrid of variable selection (coefficients restricted to be zero) and at the same time shrinkage (coefficients shrunk towards, but not equal to, zero).

For the prior hyperparameters $\alpha, \pi, \tau$ which show up in the hierarchical prior in equations (3)-(5), I define further prior distributions in order to let the data determine their values. These hyperprior distributions are

$$\tau^2 \sim iGamma\,(0.01, 0.01) \quad (6)$$
$$\alpha \sim Gamma\,(1, 2) \quad (7)$$
$$\pi \sim Beta\,(1, 1)\,, \quad (8)$$

and the chosen hyperparameters are fairly uninformative. Estimation of the regression coefficients using the prior in equations (3)-(8) is implemented using Markov Chain Monte Carlo methods which are described analytically in the Technical Appendix. After monitoring for convergence, the Gibbs sampler is run for 150,000 iterations after an initial burn-in period of 50,000 iterations.

# 3 Empirical Results

## 3.1 Forecast evaluation

I consider short-term forecasts, i.e. $h = 1, 2, 3, 4$ horizons ahead, of inflation (Consumer Price Index: All Items) and output (Real Gross Domestic Product) using 183 predictors[2]. All data used are quarterly, seasonally adjusted and are observed for the period 1959.Q1-2011.Q2. The Data Appendix contains a full description of all variables and the relevant stationarity transformations used. 50% of the available sample is used as the first estimation period, forecasts are calculated, then one observation is added at the end of the initial sample and estimation and forecasting is repeated. This recursive forecasting procedure is followed until the whole sample is exhausted.

Following standard practice, I use the model with no predictors (i.e. an autoregressive model with 2 lags and an intercept, estimated using diffuse priors) as a benchmark model. Additionally, the regression model (1) with the 183 predictors is estimated using the semiparametric variable selection prior (3)-(8), and the traditional spike and slab prior consisting of equations (2), (6) and (8). Lastly, I provide forecasts from the regression model (1) where the 183 variables in $x_t$ are replaced by the first principal component[3] and a diffuse prior is used on all regression coefficients (so that posterior and predictive means/medians are equivalent to the OLS point estimates).

I use a large set of alternative measures of out-of-sample predictive ability. Let $N$ denote the number of observations in the out-of-sample evaluation period, and denote the forecast errors of the benchmark AR(2) model $M_0$ as $\epsilon_i^0$, and of model $M_j$

[2] When forecasting inflation, output becomes a predictor and vice-versa.

[3] The final conclusions of this paper are not affected if a larger number of principal components is considered. The first principal component gives the lowest mean absolute error in most instances, although models with a larger number of principal components also achieve a larger value of the predictive likelihood.

as $\epsilon_i^j$, for $i = 1, ..., N$ and $j = 1, ..., G$. Define $MSE^j = N^{-1} \sum_{i=1}^{N} \left(\epsilon_i^j\right)^2$ (similarly for $MSE^0$), $d_i = \epsilon_i^j - \epsilon_i^0$, and $\overline{d} = N^{-1} \sum_{i=1}^{N} d_i$. Additionally denote by $\widetilde{p}\left(y_{t+h}|y_t, x_t\right)$ the predictive likelihood, i.e. the value of the predictive density $p\left(y_{t+h}|y_t, x_t\right)$ evaluated at the realized value of $y_{t+h}$. The out-of-sample statistics for model $M_j$ are computed as

$$
\begin{aligned}
R^2 &= 1 - \frac{MSE^j}{MSE^0}, \\
\Delta MAE &= \frac{1}{N} \sum_{i=1}^{N} \left( \left|\epsilon_i^0\right| - \left|\epsilon_i^j\right| \right), \\
\Delta RMSE &= \sqrt{MSE^0} - \sqrt{MSE^j}, \\
MSE - T &= \sqrt{(N-1)/N} \times \left[ \frac{\overline{d}}{\widehat{se}\left(\overline{d}\right)} \right], \\
APL &= \frac{1}{N} \sum_{i=1}^{N} \widetilde{p}_i \left(y_{t+h}|y_t, x_t\right).
\end{aligned}
$$

For all of the statistics, but the $MSE - T$, higher values indicate better performance of model $M_j$ relative to the benchmark AR(2) model. For the MSE-T statistic, the lower the values, the better the performance of model $M_j$ relative to $M_0$.

The Bayesian semiparametric selection and the spike and slab priors provide probabilities of each variable being included in the "true" model. Comparison of these probabilities for each of the 183 variables would be interesting, however it is not implemented here for the sake of brevity. Table 1 shows the values of forecast metrics presented above, coming from the three shrinkage methods, namely the Bayesian semiparametric selection (BSS), the spike and slab (SnS) and the principal component analysis (PCA). The results suggest that semiparametric variable selection does outperform in most instances parametric variable selection in terms of forecast error ($R^2, DMAE, DRMSE, MSE - T$). When the whole predictive distribution is considered (predictive likelihood, $APL$) the more parsimonious parametric variable selection is superior. Using the semiparametric prior to account for possible correlations in the data is beneficial when forecasting the mean, however this comes at the cost of having to sample more parameters and hence increasing the variance of the predictive density.

One way to reduce the larger variance of the predictive density is to use more informative priors to sample $\tau^2, \alpha$ and $\pi$. Additionally, restrictions could be imposed on the number of mixture components sampled. When using the Dirichlet process an unknown number of mixtures is assumed, leading the algorithm to sample as many as 28 mixture components for the prior in equations (3)-(5), regardless that most of them contain no elements. A simple restriction which will make the variable selection algorithm more efficient is to restrict the maximum number of components that can be sampled.

Although the parametric spike and slab prior does not perform better than the benchmark AR(2) model for CPI inflation, both variable selection algorithms are performing better than the principal component forecasts. It is quite surprising that principal component forecasts are performing so poorly. A potential explanation is that for most of the evaluation period the number of predictors (183) are more than the number of observations (102 initial observations up to 205 final observations), hence the principal component estimates are not consistent estimates of the true factors. Examining this issue is beyond the purpose of this short note.

| | Results for CPI | | | Results for GDP | | |
|---|---|---|---|---|---|---|
| | BSS | SnS | PCA | BSS | SnS | PCA |
| | | $h = 1$ | | | $h = 1$ | |
| $R^2$ | 0.5712 | -0.3000 | -0.6657 | 0.2064 | 0.2164 | 0.0742 |
| $DMAE$ | 0.0853 | -0.0370 | -0.1425 | 0.0174 | 0.0089 | -0.0001 |
| $DRMSE$ | 0.1928 | -0.0783 | -0.1624 | 0.0655 | 0.0683 | 0.0226 |
| $MSE-T$ | 0.4993 | 0.8744 | 1.3946 | 0.5342 | 0.4921 | 0.4520 |
| $APL$ | 0.2904 | 0.3476 | 0.3397 | 0.3076 | 0.3665 | 0.2470 |
| | | $h = 2$ | | | $h = 2$ | |
| $R^2$ | 0.4554 | -0.0683 | -0.7520 | 0.1496 | 0.0536 | 0.0193 |
| $DMAE$ | 0.0473 | -0.1281 | -0.1956 | 0.0102 | -0.0113 | 0.0027 |
| $DRMSE$ | 0.1640 | -0.0210 | -0.2028 | 0.0501 | 0.0175 | 0.0063 |
| $MSE-T$ | 0.6591 | 0.7470 | 1.4938 | 0.3666 | 0.3040 | 0.1079 |
| $APL$ | 0.2926 | 0.3551 | 0.3339 | 0.3068 | 0.3652 | 0.2488 |
| | | $h = 3$ | | | $h = 3$ | |
| $R^2$ | 0.3996 | -0.3675 | -1.0306 | 0.1149 | -0.0417 | -0.0805 |
| $DMAE$ | 0.0134 | -0.1813 | -0.2174 | 0.0008 | -0.0234 | -0.0193 |
| $DRMSE$ | 0.1242 | -0.0929 | -0.2340 | 0.0442 | -0.0154 | -0.0295 |
| $MSE-T$ | 0.9057 | 1.1140 | 1.5045 | 0.1849 | 0.4247 | 0.1159 |
| $APL$ | 0.2942 | 0.3453 | 0.3304 | 0.3014 | 0.3285 | 0.2436 |
| | | $h = 4$ | | | $h = 4$ | |
| $R^2$ | 0.3166 | -0.6415 | -1.2486 | -0.0692 | -0.3319 | -0.1328 |
| $DMAE$ | -0.0101 | -0.2278 | -0.2399 | -0.0609 | -0.1014 | -0.0367 |
| $DRMSE$ | 0.0906 | -0.1474 | -0.2602 | -0.0290 | -0.1311 | -0.0546 |
| $MSE-T$ | 0.9318 | 1.4330 | 1.6545 | -0.2625 | 0.2310 | -0.2514 |
| $APL$ | 0.2926 | 0.3582 | 0.3361 | 0.2808 | 0.2895 | 0.2339 |

Table 1: Forecasting results

## 4 Conclusions

This paper presents a Bayesian prior which allows for shrinkage of coefficients in regressions with many highly correlated predictors, by selecting or restricting coefficients in groups. In a forecasting exercise involving short-term predictions of price inflation and output, this Bayesian algorithm gives considerably better results than a Bayesian prior which does not account for the correlation in exogenous predictors. Additionally, forecasts are superior to a benchmark AR(2) model, and principal component shrinkage.

## References


[1] Cremers, M. (2002). Stock return predictability: A Bayesian model selection perspective. *Review of Financial Studies* 15, 1223-1249.

[2] Dunson, D. B., Herring, A. H. and Engel, S. M. (2008). Bayesian selection and clustering of polymorphisms in functionally related genes. *Journal of the American Statistical Association*, 103, 534-546.

[3] Fernandez, C., Ley, E. and Steel, M. F. J. (2001). Benchmark priors for Bayesian model averaging. *Journal of Econometrics*, 100, 381-427.

[4] George, E. I., Sun, D. and S. Ni. (2008). Bayesian stochastic search for VAR model restrictions. *Journal of Econometrics*, 142, 553-580.

[5] Koop, G. and Korobilis, D. (2012). Forecasting inflation using dynamic model averaging. *International Economic Review*, 53, 867-886.

[6] Korobilis, D. (2012). VAR forecasting using Bayesian variable selection. *Journal of Applied Econometrics*, forthcoming, doi: 10.1002/jae.1271.

[7] MacLehose, R. F., Dunson, D. B., Herring, A. H. and Hoppin, J. A. (2007). Bayesian Methods for Highly Correlated Exposure Data. *Epidemiology* 18, 199-207.

[8] Mitchell, T. J. and Beauchamp, J. J. (1988). Bayesian variable selection in linear regression. *Journal of the American Statistical Association*, 83, 1023-1032.

[9] Stock, J. H. and Watson, M. W. (1999). Forecasting inflation. *Journal of Monetary Economics* 44, 293-335.

[10] Stock, J. H. and Watson, M. W. (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics*, 20, 147-162.

[11] Wright, J. H. (2008). Bayesian model averaging and exchange rate forecasts. *Journal of Econometrics* 146, 329-341.

# A Technical Appendix

The model is of the form

$$y_t = x_t\beta + \varepsilon_t,$$

with the usual assumptions of normality and heteroskedasticity[4]. Here $\beta$ is of dimension $(K \times 1)$ and I make the assumption that all $K$ elements are subject to the semiparametric selection prior. In the empirical section I have also an intercept $\gamma$ and lag coefficients $\varphi$ which are always unrestricted. These admit noninformative priors as in the main text but I ignore them here, because the posterior for $\beta$ is quite messy (notationally), so adding also $\gamma$ and $\varphi$ would make the formulas below more awkward to read. In practice it is straightforward to augment the formulas presented below in order to draw altogether $(\gamma, \varphi, \beta)$ from a multivariate normal.

I rewrite the priors used in the main passage compactly for convenience. For the regression coefficients $\beta$ I use a nonparametric multiple shrinkage prior of the form

$$\begin{aligned}
\beta_j &\sim \pi\delta_0(\beta) + (1-\pi)G && \text{(A.1)}\\
G &\sim DP(\alpha G_0) && \text{(A.2)}\\
G_0 &\sim N\left(\underline{\mu}, \tau^2\right) && \text{(A.3)}\\
\tau^2 &\sim iGamma(\underline{a}_1, \underline{a}_2) && \text{(A.4)}\\
\alpha &\sim Gamma\left(\underline{\rho}_1, \underline{\rho}_2\right) && \text{(A.5)}\\
\pi &\sim Beta(\underline{c}, \underline{d}), && \text{(A.6)}
\end{aligned}$$

where in this paper $\underline{\mu} = 0$. For the error variance $\sigma^2$ I use a noninformative inverse-gamma prior of the form

$$\sigma^2 \sim iGamma(\underline{\nu}_1, \underline{\nu}_2), \qquad \text{(A.7)}$$

where the "noninformativeness" comes when $\underline{\nu}_1, \underline{\nu}_2 \rightarrow 0$. When using Dirichlet process priors it is always helpful to derive the simple stick breaking representation of the coefficient $\beta_j$ conditional on $\beta_{-j}$ (and marginalized over the uncertain nonparametric density $G$)[5]. This is of the form

$$\left(\beta_j|\beta_{-j}\right) \sim \frac{\alpha(1-\pi)}{\alpha + K - p_{\beta_1} - 1} N\left(\underline{\mu}, \tau^2\right) + \pi\delta_0(\beta) + \sum_{l=2}^{k_\beta} \frac{p_{\beta_l}(1-\pi)}{\alpha + K - p_{\beta_1} - 1}\delta_{\beta_l}(\beta) \qquad \text{(A.8)}$$

[4]These assumptions need not hold. For the experienced Bayesian it is straightforward to derive the conditional posteriors with, say, Markov Switching dynamics, stochastic volatility, and Student-t errors.

[5]To establish some notation, $\beta_{-j}$ denotes the vector $\beta$ with its $j$-th element removed. In the following, $\delta_x(y)$ denotes the Dirac-delta function for random variable $x$ which gives a point mass at $y$. Lastly, for a vector $z_t$ define $Z$ to be the matrix of all stacked $z_t$, for example for $x_t$ we have $X = (x_1, ..., x_T)$.

where $k_\beta$ is the number of atoms in the above equation (number of mixture components plus the $\delta_\beta(0)$ component), and $p_{\beta_n}$ is the number of elements of the vector $\beta$ which which are equal to $\delta_{\beta_l}(\beta)$, $n = 1, 2, ..., k_\beta$, where it holds that $\delta_{\beta_1}(\beta) = \delta_0(\beta)$. Additionally, for notational convenience define the prior weights as

$$\begin{aligned} w_0 &= \frac{\alpha(1-\pi)}{\alpha + K - p_{\beta_1} - 1} \\ w_1 &= \pi \\ w_l &= \frac{p_{\beta_l}(1-\pi)}{\alpha + K - p_{\beta_1} - 1}, \; l = 2, ..., k_\beta. \end{aligned}$$

*Gibbs sampling algorithm for Bayesian clustering and selection:*

- Given $k_\beta$ number of mixture components, sample $\theta = \left(\theta_1, ..., \theta_{k_\beta}\right)$ from

  $$(\theta|-) \sim N(E_\theta, V_\theta),$$

  with $E_\theta = V_\theta\left(T^{-1}M + \sigma^{-2}X'_\pi\widehat{Y}\right)$ and $V_\beta = \left(T^{-1} + \sigma^{-2}X'_\pi X_\pi\right)^{-1}$, where $T = \tau^2 I_{k_\beta}$ and $M = \mu\mathbf{1}_{k_\beta}$. Here $X'_\pi$ denotes the matrix $X$ with the columns corresponding to coefficients belonging to $\theta_1$ being replaced with zeros (or equivalently, with these columns removed). Hence the remaining columns correspond to unrestricted coefficients which belong to one of the remaining $k_\beta - 1$ mixture components.

- Sample $\beta_j$ conditional on $\beta_{-j}$, data, and other model parameters for $j = 1, ..., K$ from

  $$\left(\beta_j|\beta_{-j}, -\right) \sim \overline{w}_0 N\left(E_\beta, V_\beta\right) + \sum_{l=1}^{k_\beta} \overline{w}_l \theta_l,$$

  so that with probability $\overline{w}_l$ we assign $\beta_j$ equal to the atom of mixture component $l$ (i.e. $\beta_j = \theta_l$), while with probability $\overline{w}_0$ we assign $\beta_j$ to a new $N\left(E_\beta, V_\beta\right)$ component. In the expression above it holds that

  $$\begin{aligned} E_\beta &= V_\beta\left(\tau^{-2}\underline{\mu} + \sigma^{-2}X'\widetilde{Y}\right) \\ V_\beta &= \left(\tau^{-2} + \sigma^{-2}X'X\right)^{-1}, \end{aligned}$$

  and that

  $$\begin{aligned} \overline{w}_0 &\propto \frac{w_0 N\left(0; \underline{\mu}, \tau^2\right) \prod_{i=1}^T N\left(\widetilde{y}_t; 0, \sigma^2\right)}{N\left(0; E_\beta, V_\beta\right)} \\ \overline{w}_l &\propto w_l N\left(0; \underline{\mu}, \tau^2\right) \prod_{i=1}^T N\left(\widetilde{y}_t; x_{t,l}\theta_l, \sigma^2\right), \; l = 1, ..., k_\beta, \end{aligned}$$

where $\widetilde{y}_t = y_t - \sum_{j' \neq j} x_{t,j'} \beta_{j'} = y_t - (x_\pi)_t \theta + x_{j',t} \beta_{j'}$ for $j, j\prime = 1, ..., K$, $(x_\pi)_t$ is the $t$-th observation of the matrix $X_\pi$ constructed in step 1, and $N(a; b, c)$ denotes the normal density with mean $b$ and variance $c$, evaluated at point $a$.

- Introduce an indicator variable $S_\beta = l$ if the coefficient $\beta_j$ belongs to cluster $l$, where $j = 1, ..., K$ and $l = 1, ..., k_\beta$, in which case it holds that $\beta_j = \theta_l$. In addition, set $S_\beta = 0$ if $\beta_j \neq \theta_l$, that is when $\beta_j$ does not belong to a preassigned cluster and a new cluster is introduced for this coefficient. Then the conditional posterior of $S_\beta$ is

$$(S_\beta | -) \sim Multinomial \left( 0, 1, ..., k_\beta; \overline{w}_0, \overline{w}_1, ..., \overline{w}_{k_\beta} \right).$$

- Sample the restriction probability $\pi$ from the coniditional distribution

$$(\pi | -) \sim Beta \left( \underline{c} + \sum_{j=1}^{K} I(S_\beta = 1), d + \sum_{j=1}^{K} I(S_\beta \neq 1) \right)$$

- Sample the latent variable $\eta$ from the posterior conditional

$$(\eta | -) \sim Beta \left( a + 1, K - \sum_{j=1}^{K} I(S_\beta = 1) \right).$$

- Sample the Dirichlet process precision coefficient $\alpha$ from the conditional posterior

$$\begin{aligned}(\alpha | -) \sim\ & \pi_\eta Gamma \left( \underline{\rho}_1 + k_\beta - n_{S_\beta = 1}, \underline{\rho}_2 - \log \eta \right) + \\ & (1 - \pi_\eta) \, Gamma \left( \underline{\rho}_1 + k_\beta - n_{S_\beta = 1} - 1, \underline{\rho}_2 - \log \eta \right)\end{aligned}$$

where the weight $\pi_\eta$ is given by

$$\frac{\pi_\eta}{1 - \pi_\eta} = \frac{\underline{\rho}_1 + k_\beta - n_{S_\beta = 1} - 1}{\left( K - \sum_{j=1}^{K} I(S_\beta = 1) \right) \left( \underline{\rho}_2 - \log \eta \right)},$$

and $n_{S_\beta = 1} = 1$ if $\sum_{j=1}^{K} I(S_\beta = 1) > 0$, and it is 0 otherwise (i.e. when no coefficient $\beta_j$ is restricted).

- Sample the variance $\tau^2$ coefficient from the conditional density

$$\left( \tau^2 | - \right) \sim iGamma \left( \underline{a}_1 + \frac{1}{2} (k_\beta - 1), \underline{a}_2^{-1} + \frac{1}{2} \sum_{l=2}^{k_\beta} \left( \theta_l - \underline{\mu} \mathbf{1} \right)^2 \right).$$

## B Data Appendix

The dataset is from Robert G. King and Mark W. Watson (2012), "Inflation and Unit Labor Cost", unpublished manuscript, and can be found on the link (as of May 2012): http://www.princeton.edu/ ~mwatson/ ddisk/ gerz_25_jan_2012.zip. The data series have been downloaded by these authors from St. Louis FRED, and all series span the period 1959.Q1-2011.Q2.

All variables are transformed to be approximate stationary. In particular, if $z_{i,t}$ is the original untransformed series, the transformation codes are (column Tcode below): 1 - no transformation (levels), $x_{i,t} = z_{i,t}$; 2 - first difference, $x_{i,t} = z_{i,t} - z_{i,t-1}$ ; 4 - logarithm, $x_{i,t} = \ln z_{i,t}$; 5 - first difference of logarithm, $x_{i,t} = \ln(z_{i,t}/z_{i,t-1})$; 6 - second difference of logarithm, $x_{i,t} = \ln(z_{i,t}/z_{i,t-1}) - \ln(z_{i,t-1}/z_{i,t-2})$.

| No | Mnemonic | Long Desc. | Tcode |
|---|---|---|---|
| 1 | INDPRO | Industrial Production: Total index | 5 |
| 2 | IPFINAL | Industrial Production: Final Products (Market Group) | 5 |
| 3 | IPCONGD | Industrial Production: Consumer goods | 5 |
| 4 | IPMAT | Industrial Production: Materials | 5 |
| 5 | IPDMAT | Industrial Production: Durable Materials | 5 |
| 6 | IPNMAT | Industrial Production: nondurable Materials | 5 |
| 7 | MCUMFN | Capacity Utilization: Manufacturing | 1 |
| 8 | IPDCONGD | Industrial Production: Durable Consumer Goods | 5 |
| 9 | IP.B51110.S | Industrial Production: Automotive products | 5 |
| 10 | IPNCONGD | Industrial Production: Nondurable Consumer Goods | 5 |
| 11 | IPBUSEQ | Industrial Production: Business Equipment | 5 |
| 12 | IP.B51220.S | Industrial Production: Consumer Energy Products | 5 |
| 13 | MANEMP | All Employees: Manufacturing | 5 |
| 14 | PAYEMS | Total Nonfarm Payrolls: All Employees | 5 |
| 15 | SRVPRD | All Employees: Service-Providing Industries | 5 |
| 16 | USGOOD | All Employees: Goods-Producing Industries | 5 |
| 17 | USGOVT | All Employees: Government | 5 |
| 18 | USPRIV | All Employees: Total Private Industries | 5 |
| 19 | CES9091000001 | All Employees: Federal | 5 |
| 20 | CES9092000001 | All Employees: State government | 5 |
| 21 | CES9093000001 | All Employees: Local government | 5 |
| 22 | DMANEMP | All Employees: Durable Goods Manufacturing | 5 |
| 23 | NDMANEMP | All Employees: Nondurable Goods Manufacturing | 5 |
| 24 | USCONS | All Employees: Construction | 5 |
| 25 | USEHS | All Employees: Education & Health Services | 5 |
| 26 | USFIRE | All Employees: Financial Activities | 5 |
| 27 | USINFO | All Employees: Information Services | 5 |
| 28 | USLAH | All Employees: Leisure & Hospitality | 5 |
| 29 | USMINE | All Employees: Natural Resources & Mining | 5 |
| 30 | USPBS | All Employees: Professional & Business Services | 5 |
| 31 | USSERV | All Employees: Other Services | 5 |
| 32 | USTPU | All Employees: Trade, Transportation & Utilities | 5 |
| 33 | USTRADE | All Employees: Retail Trade | 5 |
| 34 | USWTRADE | All Employees: Wholesale Trade | 5 |

| | | | |
|---|---|---|---|
| 35 | CE160V | Emp Total (Household Survey) | 5 |
| 36 | CLF16OV | Civilian Labor Force | 5 |
| 37 | LNS11300000 | LaborForce Participation Rate (16 Over) SA | 2 |
| 38 | UNRATE | Unemployment Rate | 2 |
| 39 | URATE_ST | Unrate Short Term (< 27 weeks) | 2 |
| 40 | URATE_LT | Unrate Long Term (>= 27 weeks) | 2 |
| 41 | LNS14000012 | Unemployment Rate - 16-19 yrs | 2 |
| 42 | LNS14000025 | Unemployment Rate - 20 yrs. & over, Men | 2 |
| 43 | LNS14000026 | Unemployment Rate - 20 yrs. & over, Women | 2 |
| 44 | UEMPLT5 | Number Unemployed for Less than 5 Weeks | 5 |
| 45 | UEMP5TO14 | Number Unemployed for 5-14 Weeks | 5 |
| 46 | UEMP15T26 | Civilians Unemployed for 15-26 Weeks | 5 |
| 47 | UEMP27OV | Number Unemployed for 27 Weeks & over | 5 |
| 48 | LNS12032194 | Employment Level - Part-Time for Economic Reasons, All Industries | 5 |
| 49 | AWHMAN | Average Weekly Hours: Manufacturing | 1 |
| 50 | AWOTMAN | Average Weekly Hours: Overtime: Manufacturing | 2 |
| 51 | A0M046 | Index of Help-Wanted Advertising in Newspapers | 1 |
| 52 | HOUST | Housing Starts: Total: New Privately Owned Housing Units Started | 5 |
| 53 | HOUST5F | Privately Owned Housing Starts: 5-Unit Structures or More | 5 |
| 54 | HOUSTMW | Housing Starts in Midwest Census Region | 5 |
| 55 | HOUSTNE | Housing Starts in Northeast Census Region | 5 |
| 56 | HOUSTS | Housing Starts in South Census Region | 5 |
| 57 | HOUSTW | Housing Starts in West Census Region | 5 |
| 58 | PERMIT | New Private Housing Units Authorized by Building Permit | 5 |
| 59 | A0M007 | Mfrs' new orders durable goods industries (bil. chain 2000 $) | 5 |
| 60 | A0M008 | Mfrs' new orders, consumer goods and materials (mil. 1982 $) | 5 |
| 61 | A1M092 | Mfrs' unfilled orders durable goods indus. (bil. chain 2000 $) | 5 |
| 62 | A0M032 | Index of supplier deliveries - vendor performance (pct.) | 1 |
| 63 | A0M027 | Mfrs' new orders, nondefense capital goods (mil. 1982 $) | 5 |
| 64 | A0M070 | Manufacturing and trade inventories (bil. Chain 2005 $) | 5 |
| 65 | A0M057 | Manufacturing and trade sales (mil. Chain 2005 $) | 5 |
| 66 | A0M059 | Sales of retail stores (mil. Chain 2000 $) | 5 |
| 67 | PPIACO | Producer Price Index: All Commodities | 6 |
| 68 | WPU0561 | Producer Price Index: Crude Petroleum | 5 |
| 69 | PPIFGS | Producer Price Index: Finished Goods | 6 |
| 70 | PPIFCF | Producer Price Index: Finished Consumer Foods | 6 |
| 71 | PPIFCG | Producer Price Index: Finished Consumer Goods | 6 |
| 72 | PPIIDC | Producer Price Index: Industrial Commodities | 6 |
| 73 | PPIITM | Producer Price Index: Intermediate Materials: Supplies & Components | 6 |
| 74 | PSCCOM | Spot Market Price Index: BLS & CRB: All Commodities (1967=100) | 5 |
| 75 | PMCP | NAPM Commodity Prices Index (%) | 1 |
| 76 | CPIAUCSL | Consumer Price Index: All Items | 6 |
| 77 | CPILFESL | Consumer Price Index: All Items Less Food & Energy | 6 |
| 78 | CES2000000008 | Average Hourly Earnings: Construction | 5 |
| 79 | CES3000000008 | Average Hourly Earnings: Manufacturing | 5 |
| 80 | AHETPI | Average Hourly Earnings: Total Private Industries | 5 |
| 81 | AAA | Moody's Seasoned Aaa Corporate Bond Yield | 2 |
| 82 | BAA | Moody's Seasoned Baa Corporate Bond Yield | 2 |
| 83 | FEDFUNDS | Effective Federal Funds Rate | 2 |
| 84 | CPF3M | 3-Month AA Financial Commercial Paper Rate | 2 |
| 85 | CP90_Tbill | CP3FM-TB3MS | 1 |

| | | | |
|---|---|---|---|
| 86 | GS1 | 1-Year Treasury Constant Maturity Rate | 2 |
| 87 | GS10 | 10-Year Treasury Constant Maturity Rate | 2 |
| 88 | MORTG | 30-Year Conventional Mortgage Rate | 2 |
| 89 | TB3MS | 3-Month Treasury Bill: Secondary Market Rate | 2 |
| 90 | TB6MS | 6-Month Treasury Bill: Secondary Market Rate | 2 |
| 91 | MED3 | 3-Month Eurodollar Deposit Rate (London) | 2 |
| 92 | MED3_TB3M | MED3-TB3MS (Version of TED Spread) | 1 |
| 93 | AAA_GS10 | AAA-GS10 Spread | 1 |
| 94 | BAA_GS10 | BAA-GS10 Spread | 1 |
| 95 | MRTG_GS10 | MORTG-GS10 Spread | 1 |
| 96 | TB6M_TB3M | TB6M-TB3M Spread | 1 |
| 97 | GS1_TB3M | GS1-TB3M Spread | 1 |
| 98 | GS10_TB3M | GS10-TB3M Spread | 1 |
| 99 | BOGAMBSL | Board of Governors Monetary Base | 5 |
| 100 | BOGNONBR | Non-Borrowed Reserves of Depository Institutions | 5 |
| 101 | BUSLOANS | Commercial and Industrial Loans at All Commercial Banks | 5 |
| 102 | CONSUMER | Consumer (Individual) Loans at All Commercial Banks | 5 |
| 103 | IMFSL | Institutional Money Funds | 5 |
| 104 | M1SL | M1 Money Stock | 5 |
| 105 | M2SL | M2 Money Stock | 5 |
| 106 | MZMSL | MZM Money Stock | 5 |
| 107 | NONBORTAF | Non-Borrowed Reserves of Dep. Institutions + Term Auction Credit | 5 |
| 108 | NONREVSL | Total Nonrevolving Credit Outstanding | 5 |
| 109 | REALLN | Real Estate Loans at All Commercial Banks | 5 |
| 110 | TRARR | Board of Governors Total Reserves | 5 |
| 111 | TOTALSL | Total Consumer Credit Outstanding | 5 |
| 112 | FSPCOM | S&P'S Common Stock Price Index: Composite (1941-43=10) | 5 |
| 113 | FSDJ | Common Stock Prices: Dow Jones Industrial Average | 5 |
| 114 | MVOL | VXO/VIX Index | 1 |
| 115 | TWEXMMTH | FRB Nominal Major Currencies Dollar Index | 5 |
| 116 | EXSZUS | Foreign Exchange Rate: Switzerland (Swiss Franc per U.S. $) | 5 |
| 117 | EXJPUS | Foreign Exchange Rate: Japan (¥ per U.S. $) | 5 |
| 118 | EXUSUK | Foreign Exchange Rate: United Kingdom (cents per £) | 5 |
| 119 | EXCAUS | Foreign Exchange Rate: Canada (Canadian $ per U.S.$) | 5 |
| 120 | U0M083 | Consumer expectations (Copyright, University of Michigan) | 1 |
| 121 | DPIC96 | Real Disposable Personal Income | 5 |
| 122 | FPIC96 | Real Private Fixed Investment | 5 |
| 123 | GCEC96 | Real Government Consumption Expenditures & Gross Investment | 5 |
| 124 | GDPC96 | Real Gross Domestic Product | 5 |
| 125 | GPDIC96 | Real Gross Private Domestic Investment, | 5 |
| 126 | PCECC96 | Real Personal Consumption Expenditures | 5 |
| 127 | NRIPDC96 | Real Nonresidential Investment: Equipment & Software | 5 |
| 128 | EXPGSC96 | Real Exports of Goods & Services | 5 |
| 129 | GRECPT | Government Current Receipts (Nominal) | 5 |
| 130 | FGCEC96 | Real Federal Consumption Expenditures & Gross Investment | 5 |
| 131 | IMPGSC96 | Real Imports of Goods & Services | 5 |
| 132 | PCDGCC96 | Real Personal Consumption Expenditures: Durable Goods | 5 |
| 133 | PCESVC96 | Real Personal Consumption Expenditures: Services | 5 |
| 134 | PCNDGC96 | Real Personal Consumption Expenditures: Nondurable Goods | 5 |
| 135 | PNFIC96 | Real Private Nonresidential Fixed Investment, 3 Decimal | 5 |

| | | | |
|---|---|---|---|
| 136 | PRFIC96 | Real Private Residential Fixed Investment, 3 Decimal | 5 |
| 137 | SLCEC96 | Real State & Local Consumption Expenditures & Gross Investment | 5 |
| 138 | CBIC96 | Real Change in Private Inventories, 3 Decimal | 5 |
| 139 | CBIC96_GDP | Ch. Inv/GDP | 1 |
| 140 | OUTBS | Business Sector: Output | 5 |
| 141 | OUTNFB | Nonfarm Business Sector: Output | 5 |
| 142 | HOABS | Business Sector: Hours of All Persons | 5 |
| 143 | HOANBS | Nonfarm Business Sector: Hours of All Persons | 5 |
| 144 | PRS85006013 | Nonfarm Business Sector: Employment | 5 |
| 145 | PCEPILFE | PCE: Chain-type Price Index Less Food & Energy | 6 |
| 146 | PCEPI | PCE: Chain-type Price Index | 6 |
| 147 | PCED_G | PCE: Goods | 6 |
| 148 | PCED_DG | PCE: Durable Goods | 6 |
| 149 | PCED_NDG | PCE: Nondurable Goods | 6 |
| 150 | PCED_S | PCE: Services | 6 |
| 151 | PCED_SC | PCE: Household Consumption Expenditures (for Services) | 6 |
| 152 | PCED_MV | PCE: Motor Vehicles and Parts | 6 |
| 153 | PCED_DHE | PCE: Furnishings and Durable Household Equipment | 6 |
| 154 | PCED_REC | PCE: Recreational Goods and Vehicles | 6 |
| 155 | PCED_ODG | PCE: Other Durable Goods | 6 |
| 156 | PCED_FB | PCE: Food and Beverages Purchased for Off-Premises Cons. | 6 |
| 157 | PCED_APP | PCE: Clothing and Footwear | 6 |
| 158 | PCED_GAS | PCE: Gasoline and Other Energy Goods | 6 |
| 159 | PCED_ONG | PCE: Other Nondurable Goods | 6 |
| 160 | PCED_HU | PCE: Housing and Utilities | 6 |
| 161 | PCED_HC | PCE: Health Care | 6 |
| 162 | PCED_TRA | PCE: Transportation Services | 6 |
| 163 | PCED_RECS | PCE: Recreation Services | 6 |
| 164 | PCED_FS | PCE: Food Services and Sccommodations | 6 |
| 165 | PCED_INS | PCE: Financial Services and Insurance | 6 |
| 166 | PCED_OS | PCE: Other Services | 6 |
| 167 | GDPCTPI | Gross Domestic Product: Chain-type Price Index | 6 |
| 168 | GPDICTPI | Gross Private Domestic Investment: Chain-type Price Index | 6 |
| 169 | IPDBS | Business Sector: Implicit Price Deflator | 6 |
| 170 | COMPRNFB | Nonfarm Business Sector: Real Compensation Per Hour | 5 |
| 171 | RCPHBS | Business Sector: Real Compensation Per Hour | 5 |
| 172 | OPHNFB | Nonfarm Business Sector: Output Per Hour of All Persons | 5 |
| 173 | OPHPBS | Business Sector: Output Per Hour of All Persons | 5 |
| 174 | ULCBS | Business Sector: Unit Labor Cost | 5 |
| 175 | ULCNFB | Nonfarm Business Sector: Unit Labor Cost | 5 |
| 176 | UNLPNBS | Nonfarm Business Sector: Unit Nonlabor Payments | 5 |
| 177 | TTABSHNO | Total Tangible Assets - Balance Sheet of Households & Nonprofits | 5 |
| 178 | TNWBSHNO | Total Net Worth - Balance Sheet of Households & Nonprofits | 5 |
| 179 | NWORTH_PDI | Networth Relative to Personal Disp Income | 1 |
| 180 | TTABSHNO_ XEANSHNO | TTABSHNO-REANSHNO | 5 |
| 181 | REABSHNO | Real Estate - Assets - Balance Sheet of Households & Nonprofits | 5 |
| 182 | TFAABSHNO | Total Financial Assets - Balance Sheet of Households & Nonprofits | 5 |
| 183 | TLBSHNO | Total Liabilities - Balance Sheet of Households and Nonprofits | 5 |
| 184 | Liab_PDI | Liabilities Relative to Person Disp Income | 5 |