Appendix

A Hyper-Parameters of Horseshoe Prior

The horseshoe prior of the initial regression coefficients is $\beta_{j,0} \sim N(0, \tau_0 \tau_j)$ with $\tau_0 \sim IB(0.5, 0.5)$ and $\tau_j \sim IB(0.5, 0.5)$ for $j = 1, ..., K$. Following Makalic and Schmidt (2016), the inverted beta distributions are represented as hierarchical inverse gamma ones by introducing auxiliary variables::

$$\tau_0 \sim IB(0.5, 0.5) \Longleftrightarrow \tau_0 | \kappa_0 \sim IG\left(0.5, \frac{1}{\kappa_0}\right), \quad \kappa_0 \sim IG(0.5, 1)$$
$$\tau_j \sim IB(0.5, 0.5) \Longleftrightarrow \tau_j | \kappa_j \sim IG\left(0.5, \frac{1}{\kappa_j}\right), \quad \kappa_j \sim IG(0.5, 1)$$

with the following posteriors:

$$\begin{aligned}
&\tau_0 | \kappa_0, \beta_0, \tau_1, ..., \tau_K \sim IG\left(\frac{1+K}{2}, \frac{1}{\kappa_0} + \frac{1}{2}\sum_{j=1}^{K} \frac{1}{\tau_j}\beta_{j,0}^2\right), \\
&\kappa_0 | \tau_0 \sim IG\left(1, 1 + \frac{1}{\tau_0}\right), \\
&\tau_j | \beta_0, \tau_0, \kappa_j \sim IG\left(1, \frac{1}{\kappa_j} + \frac{1}{2\tau_0}\beta_{j,0}^2\right), \\
&\kappa_j | \tau_j \sim IG\left(1, 1 + \frac{1}{\tau_j}\right).
\end{aligned}$$

B Estimating SV Model

The sampler of Kastner and Fruhwirth-Schnatter (2014) is adapted to estimate the SV model of Equation (3). The log linearization strategy of Omori et al. (2007) is applied to approximate the logarithm of a $\chi^2(1)$-distributed variable by a mixture of normal distributions. A key ingredient of Kastner and Fruhwirth-Schnatter (2014) is applying the ASIS strategy of Yu and Meng (2011) to boost the sampling efficiency of the long-run mean and the variance parameter of the log volatility process. The details of the method can be found in Kastner and Fruhwirth-Schnatter (2014) and are not repeated here to save space.

The main difference in this paper from Kastner and Fruhwirth-Schnatter (2014) is the prior of the variance parameter in the log volatility process. Instead of setting a fixed value for the scale parameter s_h in the gamma prior $\sigma_h^2 \sim G(0.5, 2s_h)$, this paper specifies a prior $s_h \sim IB(0.5, 0.5)$ to determine s_h in a data driven way. The conditional posterior of s_h can be obtained by applying the hierarchical inverse gamma representation in Makalic and Schmidt (2016): $s_h|a_h, \sigma_h^2 \sim IG\left(1, \frac{1}{a_h} + \frac{\sigma_h^2}{2}\right)$ where a_h is an auxiliary variable with the prior $a_h \sim IG(0.5, 1)$ and the posterior $a_h|s_h \sim IG\left(1, 1 + \frac{1}{s_h}\right)$.

C Sampling the Latent Variable z_t^*

The target is to sample from the posterior $p(z_t^*|y, x, \Theta, z_{-t}^*) \propto p(z_t^*|z_{-t}^*, \rho)p(y|x, z^*, \Theta)$ following the notations from Section 3.1.

Given the AR specification of z_t^* in Equation (5), the prior part can be written as $p(z_t^*|z_{-t}^*, \rho) \propto p(z_t^*|z_{t-1}^*, \rho)p(z_{t+1}^*|z_t^*, \rho)$ for $t = 2, ..., n-1$. One can derive $z_t^*|z_{-t}^*, \rho \sim N(b_{z,t}, B_{z,t})$ where $B_{z,t}$ is a K-by-K diagonal matrix with the j^{th} diagonal entry $\frac{1}{1+\rho_j^2}$ and $b_{z,t}$ is a K-by-1 vector with the j^{th} entry $\frac{\rho_j}{1+\rho_j^2}(z_{j,t-1} + z_{j,t+1})$ for $j = 1, ..., K$ and $t = 2, ..., n-1$. When $t = 1$, one has $p(z_1^*|z_{-1}^*) \propto p(z_1^*)p(z_2^*|z_1^*)$ where $z_{j,1}^* \sim N\left(0, \frac{1}{1-\rho_j^2}\right)$ for $j = 1, ..., K$. It is straightforward to derive $z_1^*|z_{-1}^* \sim N(b_{z,1}, B_{z,1})$ with $B_{z,1} = I_K$ and $b_{z,1} = \rho \odot z_2^*$, where $\odot$ denotes the Hadamard product. When $t = n$, the posterior is simply $p(z_n^*|z_{-n}^*) \propto p(z_n^*|z_{n-1}^*)$ which is a normal distribution $N(\rho \odot z_{n-1}^*, I_K)$.

Denote $y^t = \{y_1, ..., y_t\}$ and $y^{t,n} = \{y_t, ..., y_n\}$. The GCK algorithm is applied to compute the components of the likelihood $p(y|x, z^*, \Theta)$ that is relevant to sampling z_t^*:

$$\begin{aligned} p(y|x, z^*, \Theta) &\propto p(y^{t,n}|y^{t-1}, x, z^*, \Theta) \\ &\propto r_t^{-\frac{1}{2}} \det(Q_t)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}m_t'\Omega_t m_t + \mu_t' m_t + \frac{1}{2}\phi_t' Q_t^{-1}\phi_t - \frac{(y_t - v_t)^2}{2r_t}\right) \end{aligned} \tag{C1}$$

where $m_t = E(\beta_t|y^t, x, z^*, \Theta)$ and $M_t = V(\beta_t|y^t, x, z^*, \Theta)$ are computed by a Kalman filter, $r_t = q_t + x_t' M_{t-1} x_t$, $q_t = \sigma_t^2 + x_t' W_t x_t$, $W_t = \text{diag}(d_t \odot v^2)$, $Q_t = I_K + T_t' \Omega_t T_t$, $T_t T_t' = M_t$,

$\phi_t = T_t'(\mu_t - \Omega_t m_t)$ and $v_t = x_t' m_{t-1}$. The quantities of μ_t and Ω_t are computed by a backward recursion:

$$\Omega_n = 0, \quad \mu_n = 0$$
$$\Omega_{t-1} = A_t'(\Omega_t - \Omega_t C_t D_t^{-1} C_t' \Omega_t) A_t + \frac{x_t x_t'}{q_t}$$
$$\mu_{t-1} = A_t'(I_K - \Omega_t C_t D_t^{-1} C_t')(\mu_t - \Omega_t b_t y_t) + \frac{x_t y_t}{q_t} \tag{C2}$$

where $b_t = \frac{W_t x_t}{q_t}$, $A_t = I_K - b_t x_t'$, $C_t' C_t = W_t - \frac{W_t x_t x_t' W_t}{q_t}$ and $D_t = I_K + C_t' \Omega_t C_t$. The details of derivation of the GCK algorithm can be found in the original paper of Gerlach et al. (2000) and are not repeated here to save space. The Kalman filter is a standard technique for linear Gaussian state space systems for which a concise description can be found in Gerlach et al. (2000) and a textbook treatment can be found in Hamilton (1994).

D Computing the Model Likelihood Function

The likelihood function $p(y|x, z^*, \Theta)$ of the LMI model (Equation (2) and (4)) can be computed by running a Kalman filter to integrate out β_t. Specifically, decompose the likelihood as $p(y|x, z^*, \Theta) = p(y_1|x, z^*, \Theta) \prod_{t=1}^{n-1} p(y_{t+1}|y^t, x, z^*, \Theta)$ where $y^t = \{y_1, ..., y_t\}$.

To compute $p(y_{t+1}|y^t, x, z^*, \Theta)$, first consider the distribution $p(y_{t+1}|\beta_t, y^t, x, z^*, \Theta)$. By substituting $\beta_{t+1} = \beta_t + \eta_{t+1}$ into the equation $y_{t+1} = x_{t+1}'\beta_{t+1} + \epsilon_{t+1}$, it is straightforward to show $y_{t+1}|\beta_t, y^t, x, z^*, \Theta \sim N(x_{t+1}'\beta_t, \sigma_{t+1}^2 + x_{t+1}' W_{t+1} x_{t+1})$ where $W_{t+1} = \text{diag}(d_{t+1} \odot v^2)$. Next write $p(y_{t+1}|y^t, x, z^*, \Theta) = \int p(y_{t+1}|\beta_t, y^t, x, z^*, \Theta) p(\beta_t|y^t, x, z^*, \Theta) d\beta_t$. The distribution $p(\beta_t|y^t, x, z^*, \theta_z)$ is normal where the mean $m_t = E(\beta_t|y^t, x, z^*, \Theta)$ and the covariance matrix $M_t = V(\beta_t|y^t, x, z^*, \Theta)$ can be computed through a Kalman filter. It can be shown that $p(y_{t+1}|y^t, x, z^*, \Theta)$ is also normal with the mean $x_{t+1}' m_t$ and the variance $x_{t+1}' M_t x_{t+1} + \sigma_{t+1}^2 + x_{t+1}' W_{t+1} x_{t+1}$.

The initial component $p(y_1|x, z^*, \Theta) = \int p(y_1|\beta_0, x, z^*, \Theta) p(\beta_0|x, z^*, \Theta) d\beta_0$ is computed by inserting the prior $p(\beta_0|v_0^2) = N(0, \mathrm{diag}(v_0^2))$, where $v_0^2 = \tau_0\tau$, and is normal with the mean of zero and the variance $x_1'\mathrm{diag}(v_0^2)x_1 + \sigma_1^2 + x_1'W_1x_1$.

E ASIS for the LMI Model

The ASIS is applied to two blocks of parameters. The first block involves the time-varying coefficients β and the parameters v and β_0 and is based on the following reparameterization of Equation (2):

$$
\begin{aligned}
&y_t = x_t'\beta_0 + (x_t \odot \beta_t^*)'v + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_t^2), \\
&\beta_t^* = \beta_{t-1}^* + \eta_t^*, \quad \eta_t^* \sim N(0, \mathrm{diag}(d_t)), \quad \beta_0^* = 0
\end{aligned} \tag{E1}
$$

where $\beta_t^* = \mathrm{diag}(v)^{-1}(\beta_t - \beta_0)$. Note that β_0 and v become the fixed coefficients of a linear regression model conditional on β_t^* and σ_t^2 in Equation (E1). Such a reparameterization of a TVP model was pioneered in Fruhwirth-Schnatter and Wagner (2010). The specific steps are as follows:

1. In each MCMC sweep after running the Gibbs sampler of Section 3.1, compute $\beta_t^* = \mathrm{diag}(v)^{-1}(\beta_t - \beta_0)$ for $t = 1$, ..., n.

2. Let α be a $2K$-by-1 vector stacking β_0 and v. Conditional on β_t^*, σ_t^2 and the hyperparameters θ_0, draw α from a linear regression with the posterior $N(b_\alpha, B_\alpha)$, where $B_\alpha^{-1} = B_0^{-1} + \sum_{t=1}^n \frac{1}{\sigma_t^2}\tilde{x}_t\tilde{x}_t'$, $B_\alpha^{-1}b_\alpha = \sum_{t=1}^n \frac{1}{\sigma_t^2}\tilde{x}_t y_t$, B_0 is a $2K$-by-$2K$ diagonal matrix with the diagonal elements $\tau_0\tau_1$, ..., $\tau_0\tau_K$, τ_v, ..., τ_v, and $\tilde{x}_t = [x_t \;\; x_t \odot \beta_t^*]'$.

3. Compute back $\beta_t = \beta_t^* \odot v + \beta_0$ for $t = 1$, ..., n.

The resulting β_0, v and β_t from this ASIS step are used as their final draw in an MCMC sweep.

The second ASIS boosting is applied to the block of the latent variable z^* and the parameters μ and a and is based on the AR process for the original latent variable $z = \{z_t\}_{t=1}^n$:

$$\begin{aligned} z_t &= (\mathbf{1}_K - \rho) \odot \mu + \rho \odot z_{t-1} + \xi_t, \quad \xi_t \sim N(0, \mathrm{diag}(a^2)), \\ z_1 &\sim N(\mu, (I_K - \mathrm{diag}(\rho^2))^{-1}\mathrm{diag}(a^2)) \end{aligned} \tag{E2}$$

where $\mathbf{1}_K$ denotes a K-by-1 vector of ones. The steps are as follows:

1. In each MCMC sweep after running the Gibbs sampler of Section 3.1 and the first ASIS boosting, compute $z_t = \mu + a \odot z_t^*$ for $t = 1, ..., n$.

2. Draw μ from its posterior $p(\mu|z, \rho, a, \psi)$ based on Equation (E2) and the prior $\mu \sim N(0, \mathrm{diag}(\psi))$. The resulting posterior is a normal distribution $N(B_\mu^{-1} b_\mu, B_\mu^{-1})$ where the j^{th} entry of b_μ is $\frac{1}{a_j^2}\left((1-\rho_j^2)z_{j,1} + (1-\rho_j)\sum_{t=1}^{n-1}(z_{j,t+1} - \rho_j z_{j,t})\right)$ and B_μ is a diagonal matrix with the j^{th} diagonal entry $\frac{1}{\psi_j} + \frac{1}{a_j^2}\left(1 - \rho_j^2 + (n-1)(1-\rho_j)^2\right)$ for $j = 1, ..., K$.

3. Keep the sign of a.

4. Draw a^2 from its posterior $p(a^2|z, \rho, \mu)$ based on Equation (E2) and the prior $a_j^2 \sim G(0.5, 2\tau_a)$ for $j = 1, ..., K$. It can be derived that the posterior for each a_j^2 is a generalized inverse Gaussian distribution $GIG(\frac{1-n}{2}, \frac{1}{\tau_a}, \sum_{t=1}^n \xi_{j,t}^2)$ where $\xi_{j,1} = \sqrt{1-\rho_j^2}(z_{j,1} - \mu_1)$ and $\xi_{j,t} = z_{j,t} - (1-\rho_j)\mu_j - \rho_j z_{j,t-1}$ for $t = 2, ..., n$ and $j = 1, ..., K$. Update a as the square root of a^2 times the sign from the previous step.

5. Compute back $z_t^* = \mathrm{diag}(a)^{-1}(z_t - \mu)$ for $t = 1, ..., n$.

The resulting μ, a and z_t^* from this ASIS step are used as their final draw in an MCMC sweep.

F Alternative TVP Models with Dynamic Shrinkage

Two alternative TVP models with dynamic shrinkage features are compared to the proposed LMI model in the applications of this paper.

The RMI model is a restricted version of the MI model of Giordani and Kohn (2008) which is specified in Equation (2). Instead of the unrestricted 2^K scenarios in a model with K regressors, the RMI model allows only 3 scenarios for the mixture indicator $d_t = [d_{1,t} \ldots d_{K,t}]'$ at each time t: all $d_{j,t}$ equal zero (i.e. no parameter change at time t), all $d_{j,t}$ equal one (i.e. all parameters change at time t), and $d_{1,t} = 1$ and $d_{j,t} = 0$ for $j > 1$ (i.e. only the intercept changes at time t while other parameters remain constant). Denote the 3 scenarios as s_1, s_2 and s_3. The dynamics of d_t is Markovian following $p(d_t = s_j | d_{t-1} = s_j) = p_r$ and $p(d_t = s_i | d_{t-1} = s_j) = \frac{1-p_r}{2}$ for $i \neq j$. The prior for p_r is Beta$(50, 0.5)$ to favor persistent dynamics of d_t. Priors for other model parameters are the same as those in the LMI model.

The DHS model of Kowal et al. (2019) follows:

$$
\begin{aligned}
&y_t = x_t'\beta_t + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_t^2), \\
&\Delta\beta_{j,t} \sim N(0, \tau_{v,0}\tau_{v,j}\phi_{j,t}), \quad \beta_{j,0} \sim N(0, v_{j,0}^2), \\
&\log(\phi_{j,t}) = \rho_j \log(\phi_{j,t-1}) + \xi_{j,t}, \quad \xi_{j,t} \sim Z(0.5, 0.5, 0, 1), \quad \phi_{j,0} = 1 \qquad \text{(F1)}
\end{aligned}
$$

where $\sqrt{\tau_{v,0}} \sim C^+(0, \frac{1}{\sqrt{nK}})$ and $\sqrt{\tau_{v,j}} \sim C^+(0, 1)$ for $j = 1, \ldots, K$. The logarithm of the local variance $\phi_{j,t}$ follows an autoregressive process with a Z-distributed innovation $\xi_{j,t}$ which is obtained as the logarithm of an inverted-beta distributed random variable. The Z distribution can be sampled as a scale mixture of normal distributions. See Kowal et al. (2019) for the motivation and details of the DHS model. In the applications, the prior for ρ_j is $N(0.95, 1)I\{-1 < \rho_j < 1\}$. The priors of other model parameters are the same as those in the LMI model.

G Equity Premium Data

Table G1: List of Predictors for Equity Premium

Name	Description
Dividend price ratio	Log dividends minus log price
Dividend payout ratio	Log dividends minus log earnings
Stock variance	Sum of squared daily returns on the S&P500 index
Book-to-market ratio	Ratio of book to market value for the Dow Jones Industrial Average index
Net equity expansion	Ratio of 12-month moving sums of net issues by NYSE listed stocks divided by the total end-of-year market capitalization of NYSE stocks
Treasury bill rate	Quarterly change of 3-month secondary market Treasury bill rate
Long term yield	Quarterly change of long-term government bond yield from Ibbotson's *Stocks, Bonds, Bills and Inflation Yearbook*
Term spread	Long term yield minus treasury bill rate
Default yield spread	Difference between BAA and AAA-rated corporate bond yields
Default return spread	Difference between long-term corporate and government bond returns
Inflation rate	Consumer price index (all urban consumers)
Investment-to-capital ratio	Ratio of aggregate (private non-residential fixed) investment to aggregate capital

Note: The data is publicly available from Amit Goyal's website https://sites.google.com/view/agoyal145/?redirpath=/. Detailed descriptions of the variables can be found in Welch and Goyal (2008).

H Inflation Rate Data

Table H1: List of Predictors for Inflation Rate

Name	Description
GDP	Log change of real GDP
Investment	Log change of real gross private domestic investment
Expenditure	Log change of real government consumption expenditures and gross investment
Imports	Log change of imports of goods and services
Potential GDP	Log change of real potential GDP
Employee	Log change of total non-farm employees
Unemployment	Change of unemployment rate
Wage	Log change of average hourly earnings of production and non-supervisory employees
House start	Log change of new privately-owned housing unites started
House supply	Change of the ratio of houses for sale to houses sold
Public debt	Change of the ratio of public debt to GDP
Consumer debt	Log change of consumer credit to households and non-profit organizations
Mortgage	Log change of one-to-four-family residential mortgages
Energy price	Log change of consumer price index for energy in U.S. city average
Producer price	Log change of producer price index for all commodities
Short rate	Change of 3-month Treasury bill rate
Term spread	Difference between 10-year Treasury constant maturity rate and 3-month Treasury bill rate
S&P500	Log change of average daily closing price
M1	Log change of M1 money stock
M2	Log change of M2 money stock

Note: Data on the S&P500 index is from Robert Shiller's website http://www.econ.yale.edu/ shiller/data.htm. Data on all other variables are from the FRED database of the U.S. Federal Reserve Bank of St. Louis.

References

Fruhwirth-Schnatter, S. and H. Wagner (2010). Stochastic model specification search for gaussian and partially non-gaussian state space models. *Journal of Econometrics 154*, 85–100.

Gerlach, R., C. Carter, and R. Kohn (2000). Efficient bayesian inference for dynamic mixture models. *Journal of the American Statistical Association 95*, 819–828.

Giordani, P. and R. Kohn (2008). Efficient bayesian inference for multiple change-point and mixture innovation models. *Journal of Business and Economic Statistics 26*, 66–77.

Kastner, G. and S. Fruhwirth-Schnatter (2014). Ancillarity-sufficiency interweaving strategy (asis) for boosting mcmc estimation of stochastic volatility models. *Computational Statistics and Data Analysis 76*, 408–423.

Kowal, D., D. Matteson, and D. Ruppert (2019). Dynamic shrinkage processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 81*, 781–804.

Makalic, E. and D. Schmidt (2016). A simple sampler for the horseshoe estimator. *IEEE Signal Processing Letters 23*(1), 179–182.

Omori, Y., S. Chib, N. Shephard, and J. Nakajima (2007). Stochastic volatility with leverage: Fast and efficient likelihood inference. *Journal of Econometrics 140*, 425–449.

Welch, I. and A. Goyal (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies 21*(4), 1455–1508.

Yu, Y. and X. Meng (2011). To center or not to center: That is not the question - an ancillarity-sufficiency interweaving strategy (asis) for boosting mcmc efficiency. *Journal of Computational and Graphical Statistics 20*(3), 531–570.