Handout for "Priors for the Long Run"

Junyuan Zou

April 22, 2015

Vector Autoregressions with a flat prior always suffer from the problem of in-sample overfitting and poor out-of-sample forecasting power. The problems come from the fact that the initial observations are taken as non-random. Therefore, the distance between the initial value and the steady state (or trend if the series is non-stationary) implied by the model doesn't enter the conditional likelihood function as a penalization. The fitted model will generate complex deterministic trend moving from the initial values to the steady state and explain an implausible proportion of the low frequency variation of the data. To see the severity of this problem, let's look at a 7 variable VAR with log-real GDP, log-real consumption, log-real investment, log-real wages, log hours worked, inflation and a short-term interest rate with different priors. Figure 1 shows the deterministic trend implied by models with different priors and the actual realizations.

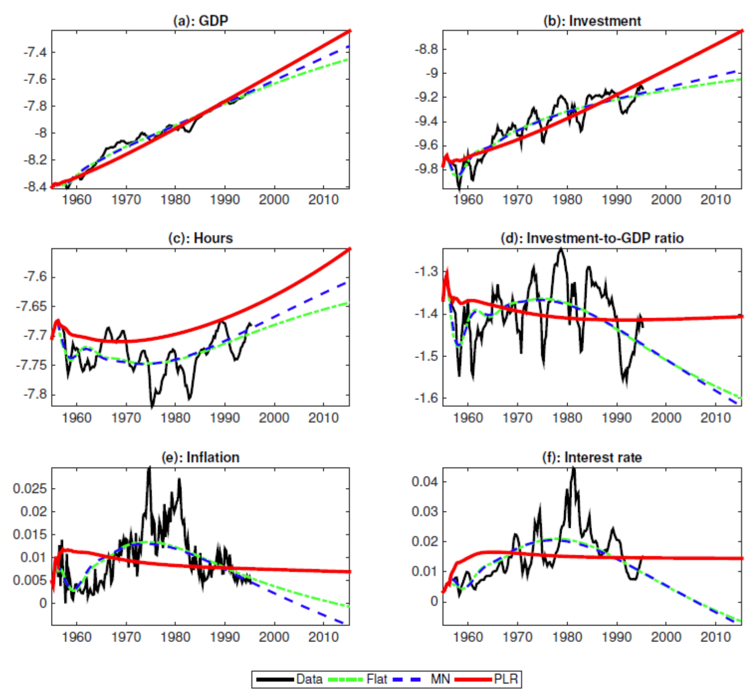

Figure 1: Deterministic Trend Implied by a 7-variable VAR

To solve the above problem, this paper proposes a class of priors which disciplines the long run behavior of the estimated model. The prior for the long run (PLR) incorporates the intuition of cointegration based on sharp predictions of economic theory. It decomposes the variable space to a-priori cointegrating space and an orthogonal space. The prior probability will be low on the VAR coefficients that imply large predictive power of initial conditions on the long run dynamics, especially for those variables that are non-stationary. Different from the literature, PLR doesn't involve pretesting of cointegration rank (as in Horvath and Watson, 1995), or using Bayesian model selection techniques (as in Vilani, 2005). It can be easily implemented using dummy variables.

To see how the prior is elicited, consider the following error correction representation of a VAR.

$$\Delta y_t = c + \Lambda \tilde{y}_{t-1} + \Gamma_1 \Delta y_{t-1} \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + \varepsilon_t,$$

where $\tilde{y}_{t-1} = H y_{t-1}$ is a vector of linear combinations of y_{t-1}. If $\tilde{y}_{it} = H_{i\cdot} y_t$ is small, it's likely to be a stationary linear combination of y_t, due to the error correction mechanism, it could have a relatively large effect on the long run dynamics. On the other hand, if $\tilde{y}_{it}$ is large, it indicates that the linear combination is likely to be non-stationary, therefore is not likely have a big impact on the long run behavior of the series. The PLR shrinks less the coefficients $\Lambda_{\cdot i}$ for those combinations that are close to zero to reflect the intuition of error correction mechanism. More specifically, the PLR for Γ conditional on a specific choice of matrix H is given by

$$\Gamma_{\cdot i} | H_i, \Sigma \sim N\left(0, \frac{\phi_i^2}{(H_{i\cdot} \bar{y}_0)^2} \Sigma\right), \quad i = 1, \ldots, n.$$

Different columns of Γ are assumed to be independent. ϕ_i are hyper-parameters controlling the overall tightness of the prior.

This paper applies the proposed PLR to the estimation of a small scale and a medium scale VAR and compare its forecasting performance with other priors proposed by the literature, including a flat prior, a Minnesota prior (Litterman, 1979), a Minnesota prior and a sum-of-coefficients prior (Sims and Zha, 1998), and a naive prior with infinitely tight Minnesota prior. The mean squared forecast error are depicted in Figure 2. As we can see from the graph, expect for the hours worked, the PLR improves over the Minnesota prior for all variables and horizons. It also improves over the Sims and Zha prior uniformly except for the case of real wage. The biggest improvement lies in the nominal variables. Compared to the naive priors, the PLR is also generally good.

There are also some potential drawbacks of the PLR proposed in this paper. One of the concern is the way to set up the prior depends on the economic theory, even if the model is reduced-formed. Technically speaking, the prior relies heavily on the choice of cointegration matrix H. This increase of degree of freedom is hard to discipline and makes data mining possible.

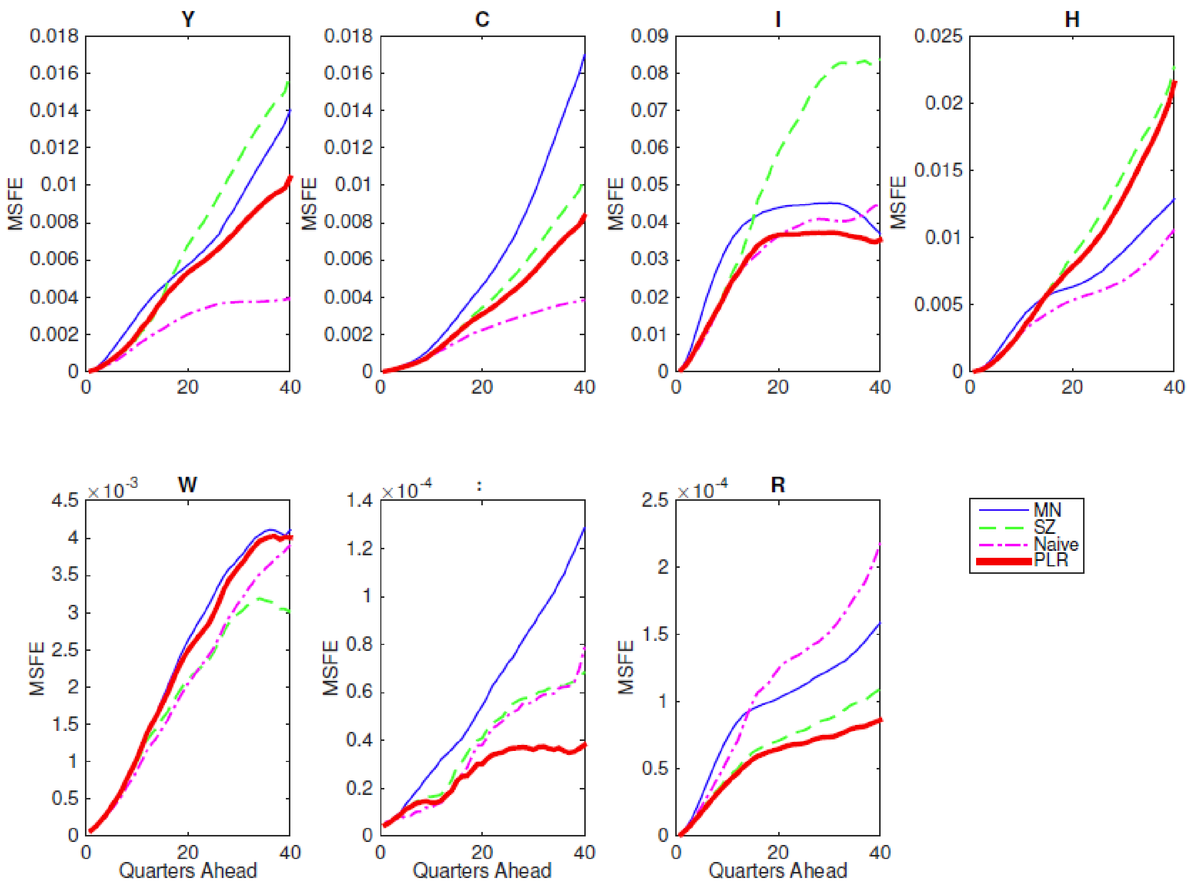

Figure 2: Mean squared forecast error in models with seven variables