Economic forecasts with Bayesian autoregressive distributed lag model: choosing optimal prior in economic downturn

Bušs, Ginters

Department of Probability Theory and Mathematical Statistics, Riga Technical University, Central Statistical Bureau of Latvia

13 September 2009

Economic Forecasts with Bayesian Autoregressive Distributed Lag Model: Choosing Optimal Prior in Economic Downturn

Ginters Bušs*
Riga Technical University and
Central Statistical Bureau of Latvia

This version: October 28, 2009

Abstract

Bayesian inference requires an analyst to set priors. Setting the right prior is crucial for precise forecasts. By using an autoregressive distributed lag model, this paper analyzes how optimal Litterman prior changes when an economy is hit by a recession. The results show that a sharp economic slowdown changes the optimal prior in two directions. First, it changes the structure of the optimal weight prior by setting smaller weight on the lagged dependent variable compared to variables containing more recent information. Second, greater uncertainty brought by a rapid economic downturn requires more space for coefficient variation which is set by the overall tightness parameter. It is shown that the optimal overall tightness parameter may increase to such an extent that Bayesian ADL becomes equivalent to frequentist ADL.

Keywords: Forecasting, Bayesian inference, Bayesian autoregressive distributed lag model, optimal prior, Litterman prior, business cycle, mixed estimation

JEL code: C1, C2, C3, C5, C13, E1

*Central Statistical Bureau of Latvia, Lacplesa 1, Riga, LV-1301, Latvia; e-mail: ginters.buss@rtu.lv

1 Introduction

Bayesian inference requires an analyst to set a prior. Setting the right prior is crucial for precise forecasts. This paper analyzes how optimal Litterman prior changes when an economy is hit by a recession. By an 'optimal Litterman prior' in this paper we define Litterman hyperparameters that minimize the root mean squared error from one-period ahead forecasts.

Although the question about what hyperparameters to use has been addressed in a series of papers by, among others, Litterman and coauthors (Litterman (1979), Doan, Litterman and Sims (1984), Litterman (1986)) and LeSage and coauthors (LeSage and Magura (1991), LeSage and Pan (1995), LeSage and Krivelyova (1999)), the role of a business cycle on the optimal prior, to the best of our knowledge, has not been discussed. Thus, this paper analyzes how (if any) prior hyperparameters should be altered for the best one-period ahead forecasting performance when there is a switch in a phase of a business cycle. For this task, an autoregressive distributed lag model (ADL) is chosen. The prior is set up like in Litterman (1979). The model is solved by 'mixed estimation' set forth in Theil and Goldberger (1961). Latvia's gross domestic product (GDP) was found to be well suited for the analysis. The results show that a sharp economic slowdown changes the optimal prior in two directions.

First, a lagged dependent variable loses its dominance as the key explanatory variable and, instead, more current information contained in leading indicator-type variables is of greater importance to improve forecasts. This changes the structure of the optimal weight prior, setting smaller weight on the lagged dependent variable compared to variables containing more recent information.

Second, greater uncertainty brought by a swift economic downturn requires more space for coefficient variation, which is set by the overall tightness parameter. Particularly, the results show that, in economic downturn, the optimal overall tightness parameter may increase to such an extent that Bayesian ADL becomes equivalent to frequentist ADL, which may imply that a greater uncertainty in an economy requires more skills from an analyst to set the right prior such that, during great economic uncertainty, one may become more comfortable using frequentist rather than Bayesian inference.

The paper is organized as follows. Section 2 describes the model and its estimation procedure. Section 3 presents the results from a case study. Finally, Section 4 concludes.

2 Methodology

2.1 The Model

Consider an autoregressive distributed lag model (ADL) of order (p, q):

$$y_t = \sum_{m=1}^{p} \beta_m y_{t-m} + \sum_{n=0}^{q} \gamma_n' x_{t-n} + \xi' z_t + \epsilon_t \tag{1}$$

where y_t is the dependent variable, x_t is a $d \times 1$ vector of key explanatory variables $x = [x_1 \ x_2 \ \dots \ x_d]$, z_t is (a vector of) other explanatory variable(s) potentially containing a constant, a dummy variable for an outlying effect, etc., and $\epsilon_t \sim N(0, \sigma^2)$. The Bayesian prior is set to

$$\begin{aligned}\beta_m &\sim N(\mathbb{1}_{\{1\}}(m), \sigma_m^2) \\ \gamma_{in} &\sim N(0, \sigma_{in}^2)\end{aligned} \tag{2}$$

where $\mathbb{1}_{\{\}}()$ is an indicator function, $m = 1, 2, \ldots, p$, $i = 1, 2, \ldots, d$, and $n = 0, 1, \ldots, q$. The specification of the standard deviation of the prior is *à la* Doan, Litterman and Sims (1984):

$$\begin{aligned}\sigma_m &= \theta k m^{-\phi} \\ \sigma_{in} &= \theta l (1+n)^{-\phi} \left(\frac{\hat{\sigma}_{u,i}}{\hat{\sigma}_{u,y}}\right)\end{aligned} \tag{3}$$

where $\hat{\sigma}_{u,y}$ and $\hat{\sigma}_{u,i}$ are the standard errors from a univariate autoregression involving y and x_i, respectively, so that $\hat{\sigma}_{u,i}/\hat{\sigma}_{u,y}$ is a scaling factor that adjusts for varying magnitudes of the involved variables. The parameter θ is referred as the overall tightness. The terms $m^{-\phi}$ and $(1+n)^{-\phi}$ are referred as lag decay functions for y and x_i, respectively, with $\phi \geq 0$ reflecting a shrinkage of the standard deviation with increasing lag length. The parameters k and l specify the relative tightness of the prior for variables y and x_i, respectively. Note that, for simplicity, we set l the same for all x_i.

2.2 Estimation

The model (1) to (3) can be estimated using the 'mixed estimation' method set forth in Theil and Goldberger (1961). For ease of exposition, drop z_t from (1) and rewrite it as

$$y = X\beta + \epsilon \tag{4}$$

where y is the $T \times 1$ vector of observations on the dependent variable, X the $T \times (p + (q+1)d)$ matrix of observations on the explanatory variables with rank $p + (q+1)d$, β the $(p + (q+1)d) \times 1$ vector of coefficients, and ϵ the $T \times 1$ vector of disturbances such that

$$E\epsilon = 0, \qquad \Sigma := E(\epsilon\epsilon') = \sigma^2 I_{T \times T}. \tag{5}$$

The Bayesian prior is included in

$$r = R\beta + \nu, \tag{6}$$

where r is a $(p+(q+1)d) \times 1$ vector $[1\ 0\ 0\ \ldots\ 0]'$, R is a $(p+(q+1)d) \times (p+(q+1)d)$ identity matrix, and ν is a $(p + (q+1)d) \times 1$ vector of disturbances such that

$$E\nu = 0 \tag{7}$$

and $E(\nu\nu')$ is a $(p+(q+1)d) \times (p+(q+1)d)$ diagonal matrix with diagonal elements being the variances specified in (3),

$$\Omega := E(\nu\nu') = \begin{bmatrix} \sigma_1^2 & 0 & & & \cdots & & & & 0 \\ 0 & \sigma_2^2 & & & & & & & \\ 0 & 0 & \ddots & & & & & & \\ & & & \sigma_p^2 & & & & & \\ \vdots & & & & \sigma_{10}^2 & & & & \vdots \\ & & & & & \sigma_{11}^2 & & & \\ & & & & & & \ddots & & \\ & & & & & & & \sigma_{d,q-1}^2 & 0 \\ 0 & & & & \cdots & & & 0 & \sigma_{dq}^2 \end{bmatrix} \tag{8}$$

The sample and the independent extraneous information may be combined by writing

$$\begin{bmatrix} y \\ r \end{bmatrix} = \begin{bmatrix} X \\ R \end{bmatrix} \beta + \begin{bmatrix} u \\ \nu \end{bmatrix}; \quad E \begin{bmatrix} u \\ \nu \end{bmatrix} = 0; \quad E\left(\begin{bmatrix} u \\ \nu \end{bmatrix} \begin{bmatrix} u' & \nu' \end{bmatrix} \right) = \begin{bmatrix} \Sigma & 0 \\ 0 & \Omega \end{bmatrix}. \tag{9}$$

An application of generalized least squares (GLS) procedure leads to estimating β as

$$\hat{\beta} = \left(\begin{bmatrix} X' & R' \end{bmatrix} \begin{bmatrix} \Sigma & 0 \\ 0 & \Omega \end{bmatrix}^{-1} \begin{bmatrix} X \\ R \end{bmatrix} \right)^{-1} \begin{bmatrix} X' & R' \end{bmatrix} \begin{bmatrix} \Sigma & 0 \\ 0 & \Omega \end{bmatrix}^{-1} \begin{bmatrix} y \\ r \end{bmatrix} \tag{10}$$

or

$$\hat{\beta} = \left[X'\Sigma^{-1}X + R'\Omega^{-1}R \right]^{-1} \left[X'\Sigma^{-1}y + R'\Omega^{-1}r \right]. \tag{11}$$

Normalizing R:

$$\tilde{R} := \begin{bmatrix} \frac{\sigma}{\sigma_1} & 0 & & & \cdots & & & & 0 \\ 0 & \frac{\sigma}{\sigma_2} & & & & & & & \\ & & \ddots & & & & & & \\ & & & \frac{\sigma}{\sigma_p} & & & & & \\ \vdots & & & & \frac{\sigma}{\sigma_{10}} & & & & \vdots \\ & & & & & \frac{\sigma}{\sigma_{11}} & & & \\ & & & & & & \ddots & & \\ & & & & & & & \frac{\sigma}{\sigma_{d,q-1}} & 0 \\ 0 & & & & \cdots & & & 0 & \frac{\sigma}{\sigma_{dq}} \end{bmatrix}$$

and r:

$$\tilde{r} := \left[\frac{\sigma}{\sigma_1} \; 0 \; 0 \; \ldots \; 0 \right]$$

gives $E(\nu\nu') = \sigma^2 I$, and the GLS estimator in (11) reduces to an ordinary least squares estimator:

$$\hat{\beta} = \left[X'X + \tilde{R}'\tilde{R} \right]^{-1} \left[X'y + \tilde{R}'\tilde{r} \right]. \tag{12}$$

Model	RMSE	RMSE1sthalf	RMSE2ndhalf
SARMA(01)(01)	0.0328737	$\boxed{0.0160291}$	0.0436398
AR(1)	0.0275043	0.0194567	0.0336810
AR(2)	0.0263058	0.0203990	0.0311106
FADL(1,0)(D)	0.0277540	0.2011203	0.0330832
FADL(2,0)(D)	0.0289995	0.0272706	0.0306310
FADL(2,1)(D)	0.0253833	0.0196827	0.0300202
FADL(2,1)(E)	0.0257016	0.0216257	0.0292142
FADL(2,1)(D+E)	0.0247125	0.0220415	0.0271218
FADL(3,2)(D)	0.0260984	0.0216730	0.0298754
FADL(3,2)(E)	0.0257382	0.0217008	0.0292230
FADL(3,2)(D+E)	0.0253316	0.0251711	$\boxed{0.0254912}$
BADL(2,1)(D+E)(.95,.1,.8,0)	0.0239113	0.0196482	0.0275217
BADL(2,1)(D+E)(.05,1,2,0)	0.0264237	0.0258526	0.0269828
BADL(3,2)(D+E)(1,.35,.2,0)	$\boxed{0.0223288}$	0.0171109	0.0265400
BADL(3,2)(D+E)(.8,.25,.2,0)	0.0225414	0.0166686	0.0271732

Table 1: A brief comparison of SARMA, AR, FADL and BADL. The two latter models are specified by their orders, (p, q), key exogenous variables, e.g. (D+E), and the Bayesian ADL with a single key exogenous variable is specified by its prior, (k,l,θ,ϕ), where prior weight $w := [k\ l]$. The least RMSE in each sample space is framed.

3 Results

The dependent variable of the model (1) is Latvia's quarterly GDP series from 1995Q1 till 2009Q1. The key explanatory variables x are two quarterly series, the output in manufacturing industry (according to Nace revision 1.1 subsequently called D) and output in electricity, gas and water supply industry (E). All three series are chained priced as of year 2000 and twice regularly and once seasonally differenced. The second regular difference is performed for better forecasting performance during the latter part of the GDP series due to a sharp economic downturn (see Buss, 2009 for a discussion). Series D and E are published before the GDP flash estimate is released, thus we can potentially use these series to forecast GDP before its other components are known. The model may contain a constant and other explanatory variables, all contained in z in (1). All calculations are performed in Scilab with the aid of its econometrics toolbox Grocer.

3.1 Warm-up

To start, Table 1 shows root mean squared forecast errors (RMSE) for the whole sample, the first half of the sample (RMSE1sthalf) and the second half of the sample (RMSE2ndhalf) from one-period ahead pseudo real-time forecasts beginning at sample size 17 from simple benchmark seasonal autoregressive moving average model (SARMA), autoregressive models (AR), and frequentist and Bayesian autoregressive distributed lag models (FADL and BADL, respectively) of order (p, q) with explanatory variable in parenthesis. Notation (D+E) means the variables are summed to result in a single explanatory variable. The Bayesian counterpart of ADL requires to specify the hyperparameters for (3), called Litterman prior consisting of four parameters, k, l, θ, and ϕ, with $w := [k\ l]$ for one-dimensional x. The forecasts are called *pseudo* real-time because they are made on the revised values of left-hand-side and right-hand-side variables in (1); although the revisions for the specific variables used in this analysis tend to be relatively small, they might underestimate RMSE. Nonetheless, this does not harm for our purpose.

The sample is split in halves because the first half contains a smooth growth whereas the second half contains rapid economic downturn (see the GDP series in Figure 1), so we can analyze how the forecasting performance of the models changes with the business cycle and, especially, how Bayesian prior has to be altered for the best forecasting performance.

Figure 1 about here

The least RMSE in each column is framed. It can be seen that Bayesian ADL models compare well with other models. It can also be seen that the BADL(3,2) models give the most precise one-period ahead forecasts for the whole sample as well as for the first half of the sample among all the ADL models considered, but they are outperformed by FADL for the second half of the model. This observation suggests that the optimal Bayesian prior might be different for the first half of the model (smooth positive growth) compared to the second half of the sample when there is a rapid economic downturn. We check this hypothesis further by employing grid search for the optimal prior.

3.2 Search for optimal priors

First, the grid search is performed for BADL(2,1)(D+E). The weight vector $[k\ k]$ is 2-dimensional, one element, k, for the dependent variable and one, l, for a single explanatory variable x, both ranging from .05 to 1 with step size .05. The overall tightness, θ, is set to range from .6 to 2.5 with step .1, and the lag decay, ϕ, from 0 to 1 with step .2. So, the grid size is $20 \times 20 \times 20 \times 6$ containing overall 48000 prior combinations for each one-period ahead forecast with sample size ranging from 17 to 51. The minimum RMSE for the whole sample is attained at the coordinate [19 2 3 1] with the corresponding values $[k\ l\ \theta\ \phi]$=[.95 .1 .8 0] with a boundary value at $\phi = 0$. The boundary for ϕ can not be decreased further since negative values would presume lags of a higher order be more informative which is counterintuitive. Figures 2(a) and 2(b) show the inverse of the RMSE as a function of the prior for the whole sample.

Figure 2 about here

Figure 2(a) shows the inverse of the RMSE as a function of the weight vector (the x and y axes represent k and l, respectively) given the rest of parameters, θ and ϕ, at their RMSE-minimizing values. It can be seen that the values of k have the major impact on the RMSE with acceptable range about (.4,1), otherwise the RMSE increases substantially. On the contrary, values of l have less influence on the RMSE given k, nonetheless, a peak is evident at $l = .1$ for all acceptable values of k.

Similarly, Figure 2(b) shows the inverse of the RMSE as a function of θ and ϕ (representing x and y axes, respectively) given the RMSE-minimizing weight vector. It can be seen that the values of both θ and ϕ have a nontrivial impact on RMSE at its optimum with the maximizing values .8 and 0, respectively. The maximizing value of $\phi = 0$ might be due to the small number of lags, which is one for each RHS variable in this model.

Now, calculating the minimum RMSE for the second half of the sample, the optimum value is attained at the coordinate [1 20 15 1] with the corresponding values [k l θ ϕ]=[.05 1 2 0] with three boundary values for k, l and ϕ. It can already be seen that the optimal prior weight is different compared to the full sample. Figures 2(c) and 2(d) show the inverse of the RMSE as a function of the prior for the second half of the sample. Figure 2(c) looks almost like the inverse of Figure 2(a). Now, the RMSE is increasing with k, with an optimum at the lowest k considered; other values of k would significantly increase the RMSE at all levels of l, the latter being also critical for optimal RMSE with acceptable range about (.3,1), otherwise the forecast error increases substantially. This observation is in line with our hypothesis that, during sharp decline in the economy, explanatory variables containing most recent information are more important than the lagged dependent variable.

Figure 2(d) shows that, for the second half of the sample, the optimal tightness parameter is higher compared to the full sample, with acceptable values in about (1,2.5), otherwise the forecast error increases substantially. This observation is as expected since the model coefficients should be given more flexibility during a rapid change in an economy. For acceptable θ, the values of lag decay parameter, ϕ, is of less importance. The forecasting performance of BADL(2,1)(D+E) for the first half of the sample is not impressive and thus not presented here.

Having explored BADL(2,1)(D+E), we now check the results for BADL(3,2)(D+E) whose forecasting performance for all sample spaces considered, as it can be seen in Table 1, is promising. The grid space is formed by k and l being from .05 to 1 with step .05, θ from .1 to 1 with step .1, and ϕ from 0 to 1 with step .1. The coordinate for the least RMSE for full sample is [20 7 2 1] with the prior values [k l θ ϕ]=[1 .35 .2 0], showing some resemblance with the results for BADL(2,1)(D+E). The inverse RMSE for full sample around the optimal prior values is shown in Figures 3(a) and 3(b). The behavior of the inverse RMSE around its optimal value is similar to that of BADL(2,1)(D+E).

Figure 3 about here

We can see from Table 1 about the model's BADL(3,2)(D+E) comparatively competitive forecasting performance for the first half of the sample. Figures 3(c) and 3(d) show the inverse RMSE around its optimum as a function of prior parameters for the first half of the sample. We see that the results are similar to the results from a full sample with optimal $k = .8$, $l = .25$, $\theta = .2$ and $\phi = 0$. It can also be seen that l has more influence on the RMSE compared to the full sample, with lowest RMSE concentrating on the lowest part of l space.

Regarding the results for the second half of the sample, the coordinate of the optimal value is [20 20 10 1], with all values being at a boundary and suggesting a greater θ (i.e., more flexibility for coefficient values). An extensive search for the optimal θ resulted to its value around 10^5 with RMSE being the same as for FADL(3,2)(D+E) at least up to and including the 7^{th} digit after a comma, shown in Table 1. The latter result might suggest that during a sharp decline in an economy one might wish to set the overall tightness parameter, θ, so loose that one is more comfortable to use frequentist version of ADL.

4 Conclusions

Bayesian inference requires an analyst to set priors. Setting the right prior is crucial for precise forecasts. This paper analyzes how optimal prior changes with business cycle, specifically, when an economy is hit by a recession. Latvia's GDP is well suited for this analysis. The results show that when economy is growing, the optimal overall tightness parameter is less than one, and the optimal weight vector sets a higher weight on a lagged dependent variable compared to other explanatory variables. However, a swift economic downturn changes the optimal prior considerably in two directions.

First, a lagged dependent variable loses its dominance as the key explanatory variable and, instead, more current information contained in leading indicator-type variables is of greater importance to improve forecasts. This changes the structure of the weight prior, setting smaller weight on the lagged dependent variable compared to variables containing more recent information.

Second, greater uncertainty brought by a rapid economic downturn requires more space for coefficient variation, which is set by the overall tightness parameter. Particularly, the results show that, in economic downturn, the optimal overall tightness parameter may increase to such an extent that Bayesian ADL becomes equivalent to frequentist ADL, which may imply that a greater uncertainty in an economy requires more skills from an analyst to set the right prior such that, during great economic uncertainty, one may become more comfortable using frequentist rather than Bayesian inference.

References

[1] Buss, G. (2009), "Comparing Forecasts of Latvia's GDP Using Simple Seasonal ARIMA Models and Direct Versus Indirect Approach", MPRA Paper 16832, University Library of Munich, Germany

[2] Doan, T., R. B. Litterman and C. A. Sims (1984), "Forecasting and Conditional Projection Using Realistic Prior Distributions", *Econometric Reviews*, 3, 1-100

[3] LeSage, J. P. (1999), *Applied Econometrics using MATLAB*

[4] LeSage, J. P., A. Krivelyova (1999), "A Spatial Prior for Bayesian Vector Autoregressive Models", *Journal of Regional Science*, 39, 297-317

[5] LeSage, J. P., M. Magura (1991), "Using Interindustry Input-Output Relations as a Bayesian prior in Employment Forecasting Models", *International Journal of Forecasting*, 7, 231-238

[6] LeSage, J. P., Z. Pan (1995), "Using Spatial Contiguity as Bayesian Prior Information in Regional Forecasting Models", *International Regional Science Review*, 18, 33-53

[7] Litterman, R. B. (1979), "Techniques of Forecasting Using ector Autoregressions", Working Paper 115, Federal Reserve Bank of Minneapolis

[8] Litterman, R. B. (1986), "Forecasting with Bayesian Vector Autoregressions - Five Years of Experience", *Journal of Business & Economic Statistics*, 4, 25-38

[9] Theil, H. and A. S. Goldberger (1961), "On Pure and Mixed Statistical Estimation in Economics", *International Economic Review*, 2, 65-78

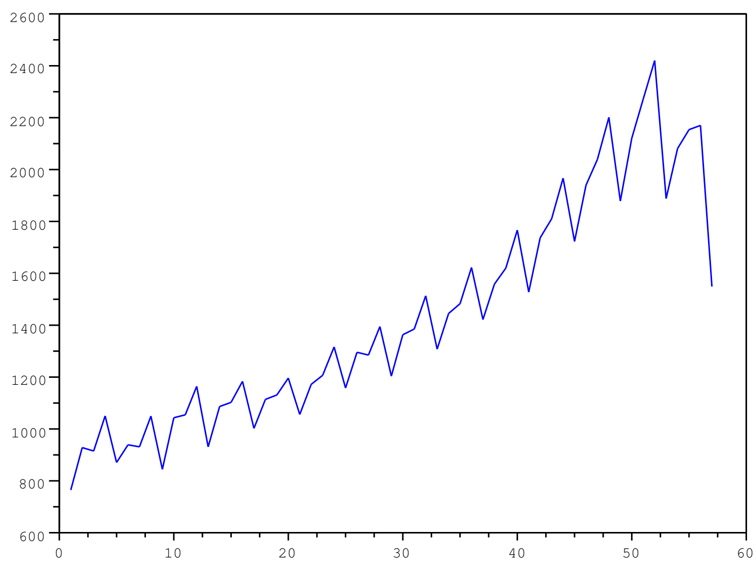
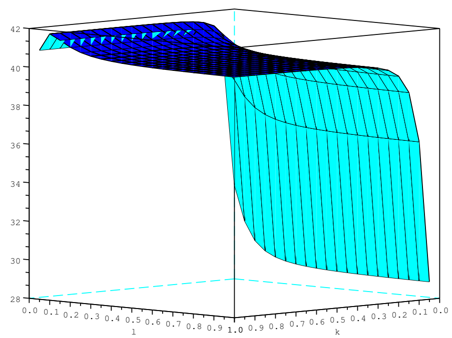
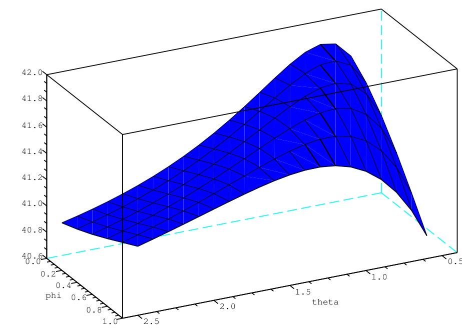


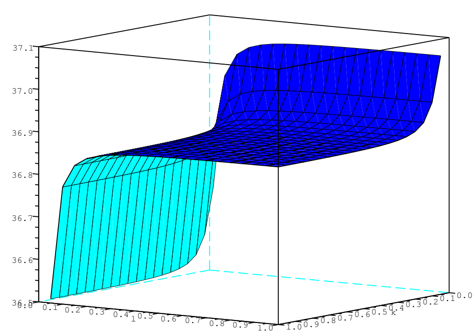
Figure 1: Latvia's seasonally unadjusted GDP series from 1995Q1 till 2009Q1. Horizontal axis represents time.



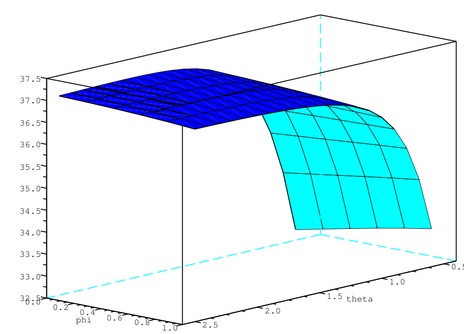
(a) Full sample. Optimal $k = .95$ and optimal $l = .1$.



(b) Full sample. Optimal $\theta = .9$ and optimal $\phi = 0$.

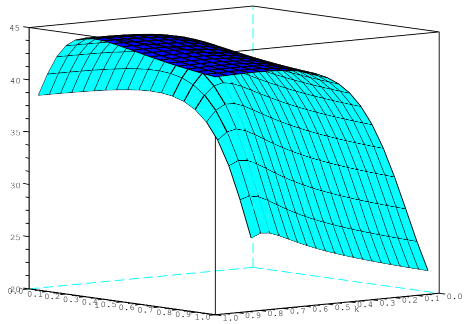


(c) Second half of the sample. Optimal $k = .05$ and optimal $l = 1$.

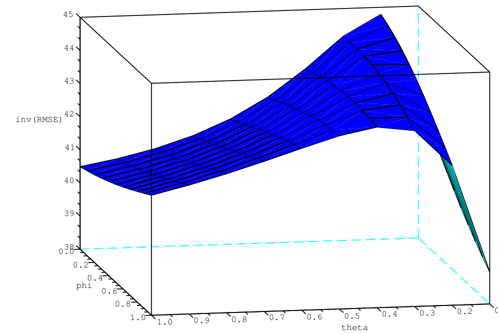


(d) Second half of the sample. Optimal $\theta = 2$ and optimal $\phi = 0$.

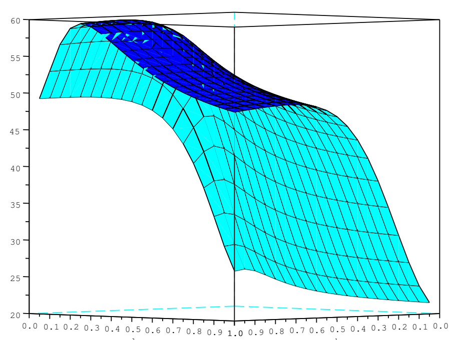
Figure 2: Results from grid search for optimal prior for BADL(2,1)(D+E). Figures 2(a) and 2(b) represent a full sample, whereas Figures 2(c) and 2(d) represent the second half of the sample. The figures on the left (2(a) and 2(c)) show $RMSE^{-1}$ (z axis) as a function of a weight vector (k,l) (x and y axis, respectively) at the RMSE-minimizing θ and ϕ. The figures on the right (2(b) and 2(d)) show $RMSE^{-1}$ (z axis) as a function of θ and ϕ (x and y axis, respectively) at the RMSE-minimizing weight vector.



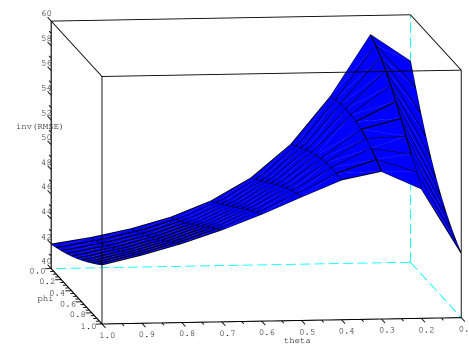
(a) Full sample. Optimal $k = 1$ and optimal $l = .35$.



(b) Full sample. Optimal $\theta = .2$ and optimal $\phi = 0$.



(c) First half of the sample. Optimal $k = .8$ and optimal $l = .25$.



(d) First half of the sample. Optimal $\theta = .2$ and optimal $\phi = 0$.

Figure 3: Results from grid search for optimal prior for BADL(3,2)(D+E). Figures 3(a) and 3(b) represent a full sample, whereas Figures 3(c) and 3(d) represent the first half of the sample. The figures on the left (3(a) and 3(b)) show $RMSE^{-1}$ (z axis) as a function of a weight vector (k,l) (x and y axis, respectively) at the RMSE-minimizing θ and ϕ. The figures on the right (3(c) and 3(d)) show $RMSE^{-1}$ (z axis) as a function of θ and ϕ (x and y axis, respectively) at the RMSE-minimizing weight vector.