Forecast Performance in Noncausal MAR(1,1) Processes

Andrew Hencic* and Joann Jasiak *

first version: June, 2017

revised: February 17, 2018

This paper examines the performance of nonlinear forecasts of noncausal processes from closed-form functional predictive density estimators. The processes considered have the mixed causal-noncausal MAR(1,1) dynamics and various non-Gaussian distributions with finite and infinite variance. The forecasts are assessed based on the forecast error behavior and the goodness of fit of the estimated predictive density.

*York University, Canada, *e-mail*: ahencic@yorku.ca.

*York University, Canada, *e-mail*: jasiakj@yorku.ca.

The second author gratefully acknowledges financial support of the Natural Sciences and Engineering Council of Canada (NSERC)

1 Introduction

This paper examines the performance of forecasts of noncausal processes obtained from closed-form functional estimators of predictive density. This forecasting method provides simulation-based prediction intervals at multiple finite horizons in one step. It has been proposed by Gourieroux and Jasiak (2016) as an alternative to the point prediction methods based on conditional expectations, which have no closed-form expressions and need to be approximated by simulations. These point forecasting methods have beeen introduced for univariate noncausal processes by Lanne, Luoto and Saikkonen (2012) and for constrained noncausal multivariate processes by Lanne,, Saikkonen (2013).

The objective of the paper is to distinguish the effects of noncausal persistence and the error distribution of the process on the forecast accuracy. We capture the persistence effect by varying the conditional mean parameters of predicted models. In addition, we assess the forecasts from processes with infinite error mean and variance distribution and compare them to forecasts from processes with finite moments. Our study includes noncausal processes with the Cauchy error distribution and t-student error distributions with degrees of freedom less than 4.

The forecast performance is evaluated in terms of its accuracy by studying the properties of forecast errors as well as the properties of the predictive density estimator.

The paper is organized as follows. Section 1 recalls the definition of the noncausal MAR(1,1) process, the filtering and simulation methods as well as the forecasts from closed-form predictive density [Gourieroux, Jasiak (2016)].

Section 2 investigates the effect of noncausal persistence and error distribution on the forecast error in one-step ahead predictions. First, we examine the forecast error properties in point predictions. Next, we study the size and coverage of the prediction interval based on selected quantiles of the predictive density.

Section 3 examines the goodness of fit of the predictive density estimators in one-step ahead predictions. We compare the performance of the Cauchy specific closed-form predictive density that has been derived for the Cauchy noncausal process with the closed-form general formula of predictive density that can be applied to noncausal processes with any error distribution. In the context of the general formula of predictive density, we study the effect of the number of simulations on forecast accuracy.

Section 4 studies the accuracy and of point predictions and the size of prediction interval in multiple step ahead predictions. We build the iso-density curves of the predictive density at horizon 2.

Section 5 concludes.

2 Noncausal MAR(1,1) Process

This section provides the definition of the MA(1,1) process, explains the approach for filtering and simulation and shows the closed-form formulas of predictive density used in forecasting.

2.1 Definition

The mixed noncausal autoregressive MAR(r,s) process is defined as:

$$\Phi(L)\Psi(L^{-1})y_t = \varepsilon_t,$$

where the errors are independent, identically distributed and such that $E(|\varepsilon_t|^\delta) < \infty$ for $\delta > 0$ and Φ and Ψ are two polynomials of degrees r and s, respectively, with roots strictly outside the unit circle and such that $\Phi(0) = \Psi(0) = 1$.

When $\psi_1 = \psi_2 - ... = \psi_s = 0$, equation (2.1) defines a purely causal autoregressive process. If $\phi_1 = \phi_2 = ... = \phi_r = 0$, the equation above defines a purely noncausal process. If both polynomials contain non-zero coefficients, then equation (2.1) describes a mixed causal-noncausal process. The mixed process contains both leads and lags of y_t. When the autoregressive orders are $r = s = 1$, the above equation defines the MAR(1,1) process:

$$(1 - \phi L)(1 - \psi L^{-1})y_t = \varepsilon_t \tag{2.1}$$

It follows from Lanne, Saikkonen (2011), and Lanne, Luoto, Saikkonen (2012), that process (y_t) has the following unobserved components:

$$u_t \equiv (1 - \phi L)y_t \leftrightarrow (1 - \psi L^{-1})u_t = \varepsilon_t, \tag{2.2}$$

and

$$v_t \equiv (1 - \psi L^{-1})y_t \leftrightarrow (1 - \phi L)v_t = \varepsilon_t, \tag{2.3}$$

which can be interpreted as the"causal" and "noncausal" components of process (y_t) *. Gourieroux, Jasiak (2016) show that i) u_t is ε-noncausal and y-causal and ii) v_t is ε-causal and y-noncausal.

The above unobserved component representation of process y_t can be used in filtering and forecasting.

2.2 Filtering and Simulation

Let us suppose that process y is observed over a period of length T and $(y_1, \ldots, y_T)$ denotes the observed sequence. The values of unobserved components u and v and errors ε can be computed from a set of observations $(y_1, \ldots, y_T)$ as follows.
(i) From equation (2.1) for $t = 2, \ldots, T-1$, we obtain the values $\varepsilon_2, \ldots, \varepsilon_{T-1}$ as functions of $(y_1, \ldots, y_T)$.

(ii) From equation (2.3) : $u_t = (1 - \phi L)y_t, t = 2, \ldots, T$, we obtain $u_2, \ldots, u_T$.

(iii) From equation (2.4) : $v_t = (1 - \psi L^{-1})y_t, t = 1, \ldots, T-1$, we obtain $v_1, \ldots, v_{T-1}$.

When an additional observation y_{T+1} becomes available, the set of unobserved components can be updated by computing ε_T, u_{T+1} and v_T.

The above formulas can be used to simulate the unobserved components as follows.

step 1: Simulate a long path of i.i.d. errors ε_t^s.

step 2: Use formulas (2.3)-(2.4) to simulate the paths of the ε-causal and ε-noncausal components :

$$u_t^s = \varepsilon_t^s + \psi u_{t+1}^s, \ t = 1, ..., 2T,$$

$$v_t^s = \varepsilon_t^s + \phi v_{t-1}^s, \ t = -T, ..., T,$$

*A process is causal with respect to a given filtration if its current value belongs to the associated information set.

starting from a far terminal condition (resp. far initial condition) $u^s_{2T} = u_0$, say (resp. $v^s_{-T} = v_0$).

2.3 Forecasting

The information set $(y_1, \ldots, y_T)$ is equivalent to information set $(v_1, \ldots, v_2, \varepsilon_2, \ldots, \varepsilon_{T-1}, u_T, \ldots, u_T)$, as shown in Gourierroux, Jasiak (2016).

Therefore, the information contained in $(y_1, \ldots, y_{T+H})$ is equivalent to the information in $(v_1, \varepsilon_2, \ldots, \varepsilon_{T+H-1}, u_{T+H}, \ldots, u_{T+H})$, and it is also equivalent to that in $(v_1, \varepsilon_2, \varepsilon_{T-1}, u_T, u_{T+H})$, because $(1 - \psi(L^{-1}))u_t = \varepsilon_t, t = T, \ldots, T + H - s$ by formula (2.3).

Thus, instead of predicting the future value of y, at horizon H, we can equivalently predict the future value of the ε-noncausal component u, by finding the the predictive density $\hat{\Pi}$ at horizon H for a noncausal process of order 1:

For $s = 1$, we get :

$$\begin{aligned} & \hat{\Pi}(u_{T+1}, \ldots, u_{T+H} | \hat{u}_T) \\ \equiv \quad & \frac{\hat{g}(\hat{u}_T - \hat{\psi} u_{T+1}) \hat{g}(u_{T+1} - \hat{\psi} u_{T+2}) \ldots \hat{g}(u_{T+H-1} - \hat{\psi} u_{T+H}) \sum_{t=1}^{T} \hat{g}(u_{T+H} - \hat{\psi} \hat{u}_t)}{\sum_{t=1}^{T} \hat{g}(\hat{u}_T - \hat{\psi} \hat{u}_t)}. \end{aligned} \tag{2.4}$$

The estimator of predictive density $\hat{\Pi}$ of the ε-noncausal component u given above can be used to generate the future values or future paths of the observable process y and its unobservable causal and noncausal components over a given horizon H.

The approach consists of four steps outlined below. The future values of y are computed from the future values of u that are drawn in $\hat{\Pi}$ by applying a sampling importance resampling (SIR) method [see Rubin (1988), Geldfand, Smith (1992)], or alternatively a Metropolis-Hasting algorithm. More specifically, the procedure is as follows:

Step 1 : Use data $(y_1, \ldots, y_T)$ to compute the filtered values of in-sample unobserved components u: $\hat{\varepsilon}_2, \ldots, \hat{\varepsilon}_{T-1}$, $\hat{v}_1, \ldots, \hat{v}_{T-1}$, $\hat{u}_2, \ldots, \hat{u}_T$.

Step 2 : Compute the approximate predictive density $\hat{\Pi}$.
Step 3: Use the SIR method to simulate future u's: $u^s_{T+1}, \ldots, u^s_{T+H}$.
Step 4: Use the recursive formulas (2.2)-(2.4) to compute the future values $y^s_{T+1}, \ldots, y^s_{T+H}$, $\hat{\varepsilon}_{T-s+1}, \ldots, \hat{\varepsilon}_{T+H-s}$, $\hat{v}_{T-s+1}, \ldots, \hat{v}_{T+H-s}$.

In special cases, the predictive density of the ε-noncausal component u admits a closed form. In particular, when the error ε follows a Cauchy distribution, the predictive density is:

$$l(u_{T+1}|u_T) = \frac{1}{\pi}\frac{1}{1+(u_T-\psi u_{T+1})^2}\frac{1+(1-|\psi|)^2u_T^2}{1+(1-|\psi|)^2u_{T+1}^2}, \tag{2.5}$$

$$\begin{aligned} l(u_{T+1}, u_{T+2}|u_T) &= \frac{1}{\pi^2}\frac{1}{1+(u_T-\psi u_{T+1})^2}\frac{1}{1+(u_{T+1}-\psi u_{T+2})^2} \\ &\quad \times\frac{1+(1-|\psi|)^2u_T^2}{1+(1-|\psi|)^2u_{T+2}^2}. \end{aligned}$$

3 One-step-ahead Forecasts

This Section illustrates the effect of noncausal persistence and error distribution on the forecast error in one-step ahead predictions. First, we examine the forecast error properties in point predictions. Next, we study the size and coverage of the prediction interval based on selected quantiles of the predictive density.

3.1 Point Predictions

In order to study the effect of noncausal persistence, we simulate samples of MAR(1,1) processes with constant causal coefficient $\phi = 0.3$ and the noncausal coefficient that takes the following values $\psi = 0.0, 0.3, 0.5, 0.9$.

To observe the effect of error distribution, we consider MAR(1,1) processes with the following error distributions:

1. Cauchy(0,1) with infinite marginal mean and variance and finite conditional moments.

2. t-student with 3 degrees of freedom with finite variance and infinite moments of higher orders

3. t-student with 4 degrees of freedom with finite skewness and infinite kurtosis

The simulated MAR(1,1) processes with Cauchy distributed errors are plotted in Figure 1 for sample size $T = 200$ observations. We observe that the simulated series with Cauchy errors display large spikes in their trajectories. These spikes that can be interpreted as bubbles, grow at a faster or slower rate depending on the value of noncausal coeffficient ψ.

[Insert Figure 1: Simulated MAR(1,1) Cauchy processes]

The forecasts of the processes are obtained from formula (2.4) to make the forecast errors comparable across different error distribution and for 1000 replications.

Table 1 shows the values of the median, interquartile range, max and min and range of the forecast errors of the MAR(1,1) processes with Cauchy, t-student-3 and t-student-4 error distributions and noncausal coefficients $\psi = 0.0, 0.3, 0.5, 0.9$.

Table 1. Forecast Errors of MAR(1,1)

Cauchy Distribution				
ψ	0.0	0.3	0.5	0.9
Median	0.039	-0.017	-0.026	0.055
Interquartile Range	2.155	2.820	3.598	2.745
Max	448.83	583.99	12,486.21	424,638.96
Min	-2,917.35	2,917.24	-3,414.91	-6,130.85
Range	3,366.22	3,501.232	15,901.116	430,769.815
t-Distribution 3 d.f.				
ψ	0.0	0.3	0.5	0.9
Median	0.045	-0.024	-0.000	-0.031
Interquartile Range	1.479	1.501	1.727	1.611
Max	4.320	4.461	5.099	6.340
Min	-6.128	-6.574	-6.548	-6.514
Range	10.449	11.035	11.648	12.855
t-Distributio 4 d.f.				
ψ	0.0	0.3	0.5	0.9
Median	0.050	0.006	-0.037	-0.016
Interquartile Range	1.427	1.464	1.637	1.477
Max	9.586	8.778	8.110	6.339
Min	-5.831	-6.335	-7.233	-7.310
Range	15.417	15.113	15.343	13.649

We first examine the forecast errors for the Cauchy process. The medians of forecast errors are close to 0 in all processes. The interquartile range depends on the magnitude of the noncausal component, increasing slightly as ψ increases from 0.0 to 0.9. We observe a strong increase in the tails, as depicted by the range and the size of the tails. The range of the errors increases from just over 3,300, to over 430,000, as ψ is increased. Most of this is due to the growth of the right tail of the distribution. The maximum grows from 448 ($\psi = 0.0$) to over 424,000 ($\psi = 0.9$). The left tail, also greatly increases, albeit to a lesser extent, as the minimum values declines from -2,917.35, to -6,130.85.

Next, we examine the forecast errors for the process generated by the t-distribution with 3 degrees of freedom. Again, the median errors are close to 0 with an interquartile range that appears to increase as ψ increases. The most apparent difference is the reduction in the tails of the distribution relative to those of the errors from the Cauchy

generated process. The range of the errors for the Cauchy process with $\psi = 0.0$ is over 3,300 whereas that of the t-distributed process with $\psi = 0.0$ is just under 15. Importantly, as in the earlier case, as the magnitude of the noncausal coefficient increases, so to does the range of the forecast errors.

Processes generated by a t-distribution with 4 degrees of freedom also exhibit very small median forecast errors. They appear to decrease slightly as the values of ψ are increased. As in the case of the other two processes, the interquartile range does increase as ψ increases, however the tails of the distribution do not grow like in the case of the Cauchy process. The range of the errors stays relatively constant (roughly 15) across all values of ψ and, in fact, slightly decrease as ψ increases.

The densities of the forecast errors are plotted in Figures 2-4.

[Insert Figures 2-4 Forecast Error Densities]

Figure 2 shows the density of the forecast errors of the Cauchy process at horizon 1. First, the large tails of the forecast errors are evident from the densities regardless of the magnitude of the noncausal coefficient. We noted that the largest increases in the range occur in the right tail of the distribution, but from the density plots the extent of the shift from a left skewed to a right skewed distribution.

The densities of the student-t with 3 degrees of freedom are plotted in Figure 3. Again, the tails of the distribution grow rapidly as the noncausal parameter increases from 0.0 to 0.9. This is expected as the student t distribution with 3 d.f. has no finite variance.

Figure 4 plots the densities of the forecast errors of the process with errors generated by the student t with 4 degrees of freedom. As expected from examining the median, interquartile range and overall range in Table 1, the distributions do not exhibit the heavy tails associated with the other two processes. In fact, there is little discernable difference between the process with noncausal parameter $\psi = 0.0$ and $\psi = 0.9$. The only apparent difference between the four specifications is the more pronounced peak of the density at the median in the case of the noncausal parameter being equal to 0.0 versus when it is equal to 0.9. We observe that the medians of forecast errors are close to 0 in all processes. The interquartile range of forecast errors depends on the magnitude of the noncausal component, increasing slightly as ψ increases from 0.0 to 0.9. We observe a strong increase in the tails, as depicted by the range and size off tails. Left, right describe.

what do you see for student 3 and 4? ADd in the 3 next panels.

3.2 Interval Predictions

Let us now examine interval predictions obtained for fixed values of persistence coefficients $\phi = 0.3, \psi = 0.5$. We perform 1000 replications of the MAR(1,1) process with errors that follow the Cauchy and t-student distributions with 3 and 4 degrees of freedom.

Table 2 below shows the prediction intervals at 70 % provided by the 15th and 85th quantiles of the predictive distribution (2.4). Note that 5% of computed prediction intervals for the Cauchy process are of infinite length. This can be explained by computational difficulty in fitting the tails of the predictive distribution for some extreme values of the conditioning value y_T, which is addressed in the next section. In contrast 0.2% and 0.7% of Student-3 and Student-4 simulations gave infinite length prediction intervals.

Table 2. Prediction intervals of MAR(1,1)

Cauchy Distribution			
mean LB	mean UB	mean length	freq in
-3.523	4.532	8.061	0.64
t-Distribution 3 d.f.			
-1.468	1.393	2.882	0.73
t-Distribution 4 d.f.			
-1.347	1.447	2.794	0.71

The accuracy of interval predictions can be assessed from the number of times the true value y_{T+1} is inside the prediction interval, which is given in column 4 of Table 2. The length of the prediction interval decreases across the rows of Table 2.

3.3 Predictive density

Let us consider the four trajectories of MAR(1,1) processes with Cauchy errors displayed in Figure 1 and denote them by $y_t^1, y_t^2, y_t^3, y_t^4$. To illustrate the predictive densities of $y_{T+1}^i, i = 1, ..., 4$ we compute them for each of these processes given their respective values at y_T^i. The values of the processes at $T = 200$ and their predictions are as follows:

Process 1: $Y_T^1 = 1.7706691$, $\hat{y}_{T+1}^1 = 0.70232739$

Process 2: $Y_T^2 = 25.037418$, $\hat{y}_{T+1}^2 = 3.0556950$

Process 3: Y_T^3 =70.185559, $\hat{y}_{T+1}^3$= 11.430161

Process 4: Y_T^4 = 253.94725, $\hat{y}_{T+1}^4$= 74.468086

The predictive densities are plotted in Figure 5.

[Insert Figure 5]

Figure 5 shows the difficulty in fitting the tails of the predictive density for high values of the conditioning variable and in the presence of high persistence.

4 Two-step Ahead Forecast

4.1 Iso-density curves

5 Conclusion

References

[1] Jasiak, J. and C. Gourieroux (2016) "Filtering, Prediction and Simulation Methods for Noncausal Processes", Journal of Time Series Analysis, Vol 37, 405-430

[2] Lanne, M., Luoma, A. and J. Luoto (2012): " Bayesian Model selection and Forecasting in Noncausal Autoregressive Models", Journal of Applied Econometrics, 27, 812-830.

[3] Lanne, M., Luoto, J. and P. Saikkonen (2012) : "Optimal Forecasting of Noncausal Autoregressive Time Series", International Journal of Forecasting, 28, 623-631.

[4] Lanne, M. and P. Saikkonen (2011) : "Noncausal Autoregressions for Economic Time Series", Journal of Time Series Econometrics, De Gruyter, 3, 1-32.

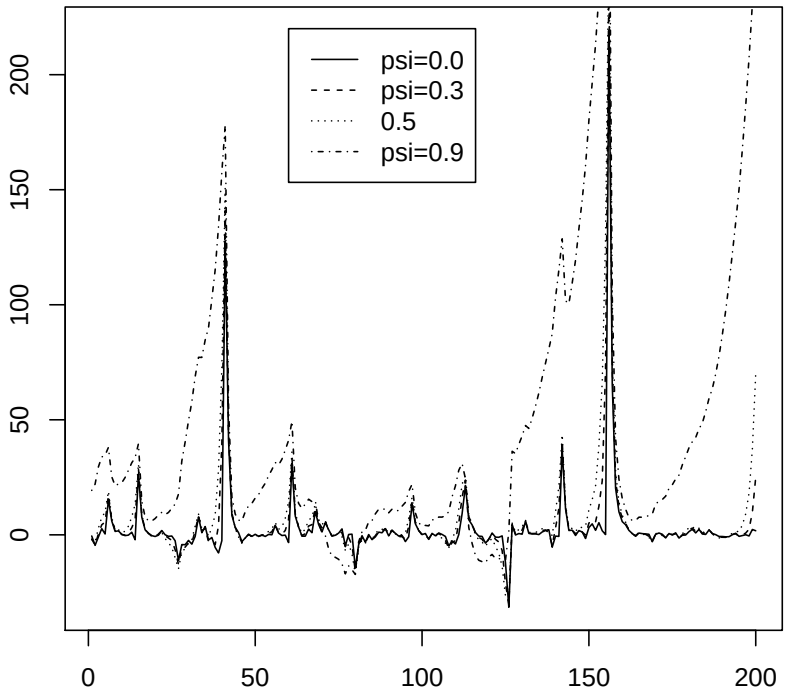

Figure 1: Simulated MAR(1,1) Cauchy processes

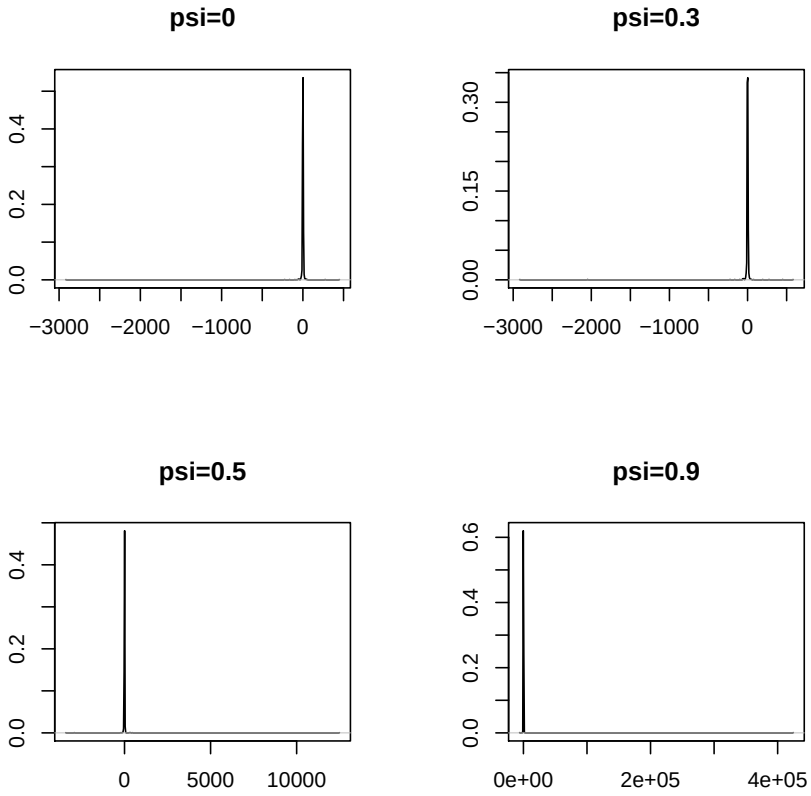

Figure 2: Forecast Error Density, Cauchy

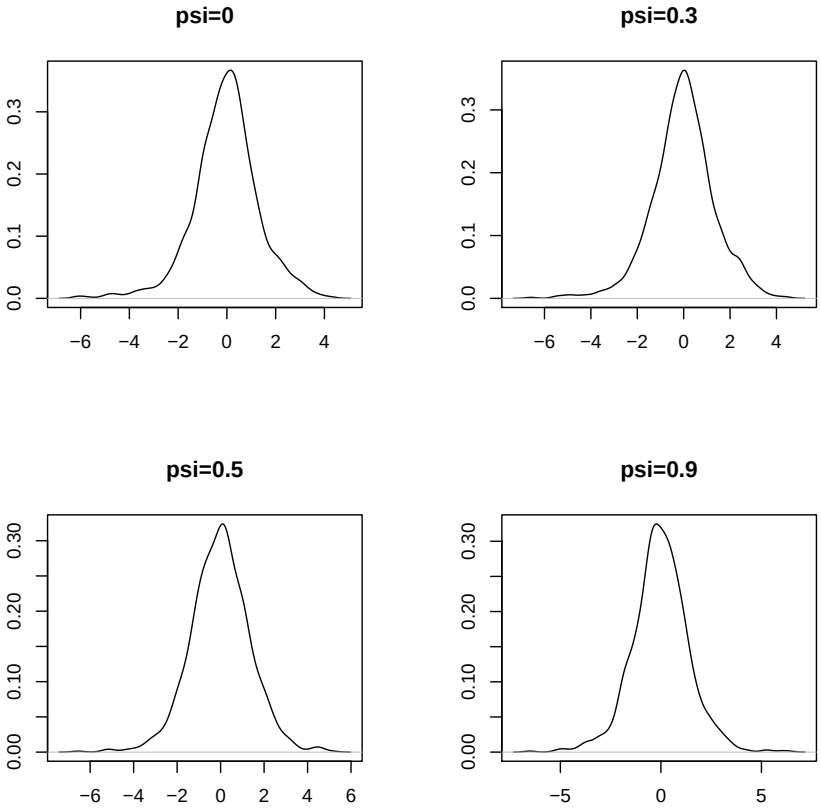

Figure 3: Forecast Error Density, Student-3

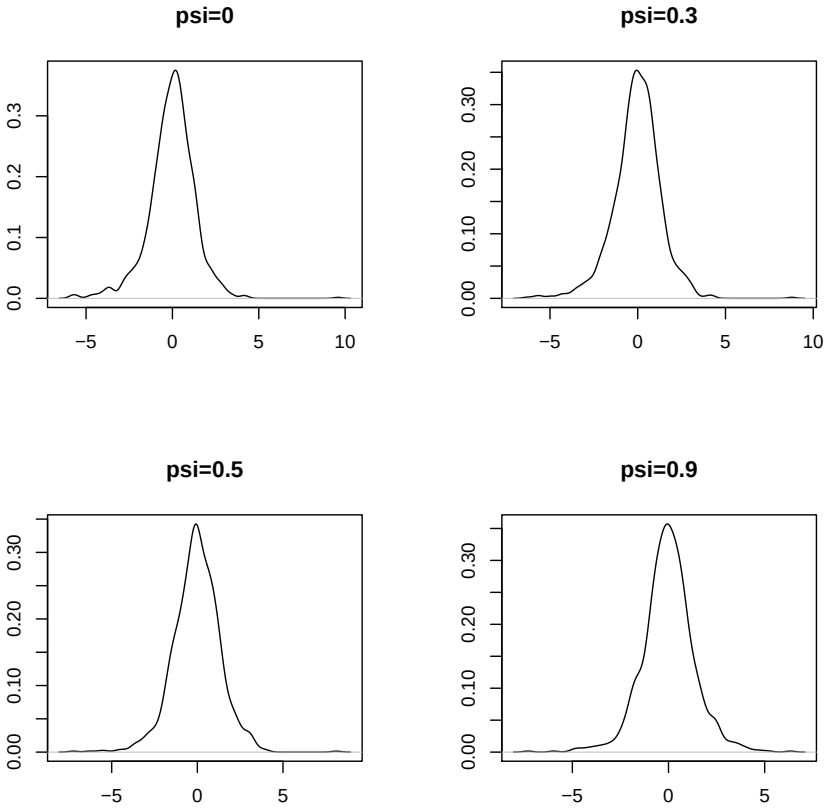

Figure 4: Forecast Error Density, Student-4

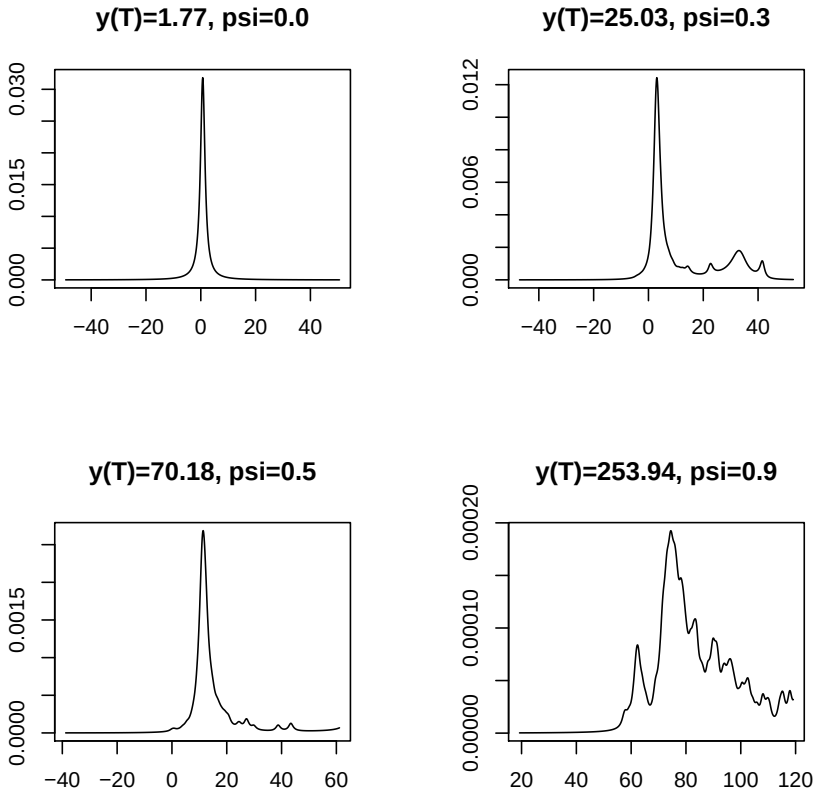

Figure 5: Predictive Density