Supplementary Information: Adding Regularized Horseshoes to the Dynamics of Latent Variable Models

Garret Binding*and Piotr Koc†

1 Horseshoe and Spike-and-Slab Priors

Our discussion in the main text focused on the regularized horsesehoe (Piironen and Vehtari, 2017) as an extension of the original horseshoe prior (Carvalho et al., 2010; Polson and Scott, 2011). Throughout the literature, the parallels between the (regularized) horseshoe and spike-and-slab priors are noted, as they predate the horseshoe as sparsity-inducing priors (Mitchell and Beauchamp, 1988; George and McCulloch, 1993). In the original specifications as a mixture of distributions, draws from a spike-and-slab prior are either drawn from a spike (e.g., a normal distribution with a very small standard deviation) or from a slab (e.g., a Cauchy distribution). If the standard deviation of the normal distribution is set to 0 (a δ spike), then the discrete spike-and-slab prior combines a binary indicator with a draw from the slab (Piironen and Vehtari, 2017).

However, others have suggested modeling the binary indicator as a continuous probability instead (Ishwaran and Rao, 2005; Brandt et al., 2023). Then, following the parameterization employed in the main text, a spike-and-slab prior can be formally represented as:

$$\begin{aligned} &\gamma_i \sim N(0, \pi^2 \lambda_i^2), \\ \pi \sim Beta(0.5, 0.5), \quad &\lambda_i \sim StudentT^+(1, 0, 1) \quad \forall i = 1, \ldots, I, \end{aligned} \tag{1}$$

or as:

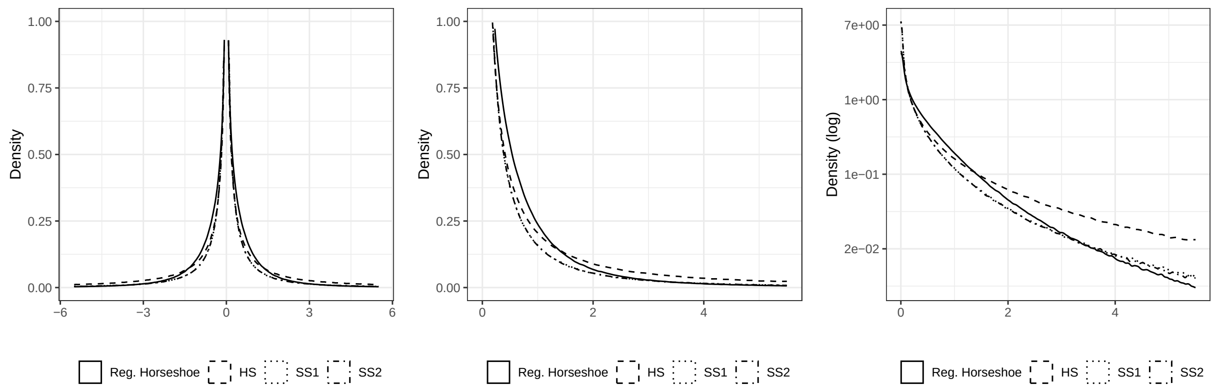
$$\begin{aligned} &\gamma_i \sim N(0, \pi_i^2 \lambda^2), \\ \pi_i \sim Beta(0.5, 0.5) \quad \forall i = 1, \ldots, I, \quad &\lambda \sim StudentT^+(1, 0, 1). \end{aligned} \tag{2}$$

In the first specification (SS1), π is a hyperparameter that determines the share of non-zero parameters overall combined with a draw from a Cauchy-distribution for each i, while this pattern is reverted in the second specification (SS2): an overall scaling parameter λ governs the size of deviations, while π_i is estimated for each i.

In Figure 1, we show the empirical density functions for the horseshoe (HS), the regularized horseshoe (RHS), the first specification of the spike-and-slab prior (SS1), and the second specification of the spike-and-slab (SS2). Overall, all four distributions exhibit very similar density functions with marked spikes at 0. There are some difference in the tails: the original horseshoe assigns more probability to the tails than either of the other three distributions, while the distributions can also be distinguished based on how markedly they descend from the spikes at 0. The regularized horseshoe exhibits the most gradual descent of the four.

*University of Basel and University of Zurich. Email: garret.binding@unibas.ch

†GESIS – Leibniz Institute for the Social Sciences and Institute of Philosophy and Sociology of the Polish Academy of Sciences.



(a) All draws.

(b) Draws > 0.

(c) Draws > 0, log(y).

Figure 1: Empirical Density Functions

2 More on Simulation Study

In this section, we provide additional results of the simulation study: (i) we add the original horseshoe prior and the two spike-and-slab priors to the comparisons (parameterized as discussed in Section 1 in the SI; S&S1 is the first, S&S2 the second parameterization), (ii) vary the extent of stasis from one year to a next (i.e., ideal points remain constant), and (iii) provide additional measures. We focus on measures of correlation between mean estimated and trues values (both Pearson's ρ and Kendall's τ), the root mean squared errors of and standard errors surrounding mean estimates, expected log pointwise density (using either LOO-CV or WAIC), and cross-validated accuracy (based on an additional left-out item) and duration of model estimation (in seconds). While the extent of stasis is held constant at 0 in the main text, we add two additional conditions of increasing stasis to the simulations presented in the SI. Therefore, the results discussed here are based on $3 \times 225 = 675$ datasets rather than the 225 focused on in the main text.

The raw results across all conditions are shown in Figure 2. For most quantities (apart from the duration), the differences between models are small enough and the range of results across all simulations large enough that it is near impossible to distinguish differences by eye. Therefore, the differences from the RHS are shown in Figure 3 in line with the presentation in the main text. In addition, Figure 4, Figure 5, and Figure 6 show the same set of quantities across the main factors varied in the simulation setup: the probability of θ being drawn from $N(0, 1)$ (instead of from a random walk), the size of σ (how large is a deviation from one time-period to the next), and the share of stasis (how often is $\theta_{t,i} = \theta_{t-1,i}$).

Of the three distributions discussed in the main text (RHS, standard normal, and Student's t), the RHS generally outperforms the other two across all metrics. The difference between the models is larger when the probability of a shock is larger (Figure 4) and when σ is smaller (Figure 5, while the extent of stasis doesn't matter much (Figure 6).

The sole exception to this pattern is the duration of model estimation. Here, both the standard normal and the Student's t-distribution are faster than the RHS. Arguably, this quantity is the least important one in our comparison. While the other metrics pertain to the actual estimates that would subsequently be used in applications, this one predominantly matters for researchers' time budgets.

Of the three distributions only discussed in the SI (S&S1, S&S2, and the horseshoe), only the original horseshoe and the first parametrization of the spike-and-slab prior perform (nearly) as well as the RHS. Note that both tend to outperform both the standard normal and the Student's t-distribution. At the same time, both the original horseshoe and the S&S1 take longer than the regularized horseshoe and run into convergence issues more frequently: while the regularized horseshoe runs into convergence issues (as evidenced by $\hat{R}$; Gelman and Rubin, 1992) in around 0.13% of observations of $\theta_{t,i}$, the original horseshoe has difficulties in 1.41% and the S&S1 in 2.06% of cases. For these reasons, we limit our discussion of these prior distributions to the SI.

3 More on Pemstein et al. (2010)

Pemstein et al. (2010) estimate static democracy scores across 197 countries and 63 years (a total of 9137 country-year observations) based on ten democracy indicators (see Table 1). Using the same data, RKF estimate a dynamic model as outlined in the main text. They compare the results when the prior of the random walks is the normal or on the t-distribution. Because the observed indicators are measured on an ordered scale, Pemstein et al. (2010) and RKF use an ordered logit specification to link observations to the latent trait. RKF highlight how the use of a dynamic (rather than a static) model decreases uncertainty in the resulting estimates, and how the use of a t-distribution (rather than a standard normal distribution) allows the model to better capture large changes in the latent trait over time. We replicate their re-analysis of Pemstein et al. (2010) and add the RHS prior to the comparison.

Table 1: Democracy indicators: number of observations and categories

Indicator	**Number of observations**	**Number of categories**	**Source**
Arat	3873	7	Arat (1991)
BLM	275	3	Bowman et al. (2005)
Bollen	510	10	Bollen (2001)
Freedom House	6438	13	Freedom House (2007)
Hadenius	129	8	Hadenius (1992)
PACL	9067	2	Przeworski et al. (2000)
PRC	6002	4	Gasiorowski (1996)
Polity	8050	21	Marshall et al. (2006)
Polyarchy	353	11	Coppedge and Reinicke (1991)
Vanhanen	8965	8	Vanhanen (2003)

Figure 7 shows the predictive accuracy of each specification across each indicator as well as across all indicators. The largest relative improvement of the RHS relative to the other two distributions is visible for the Polity indicator. This translates into an overall higher accuracy, as the Polity indicator is one of the more frequently observed indicators in this dataset. Otherwise, the RHS generally performs as well as the Student's t distribution.

Figure 8 shows different quantities concerning the convergence of the three models in the application for the estimated countries' democracy scores (θ). $\hat{R}$ is low (Gelman and Rubin, 1992) and the effective sample size (ESS) is high, indicating a good level of efficiency in the sampling process.

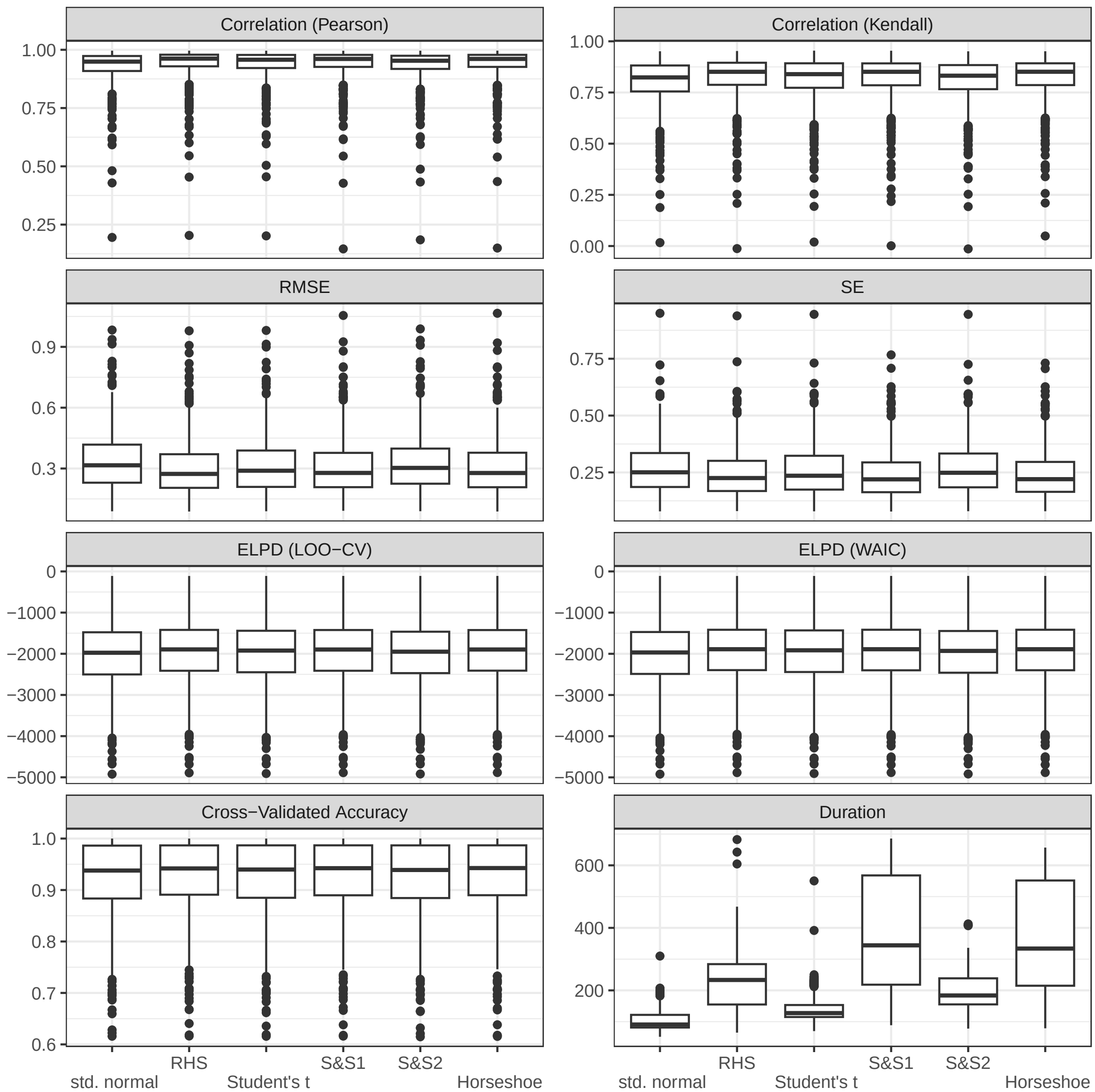


Figure 2: Simulation Results (raw values)

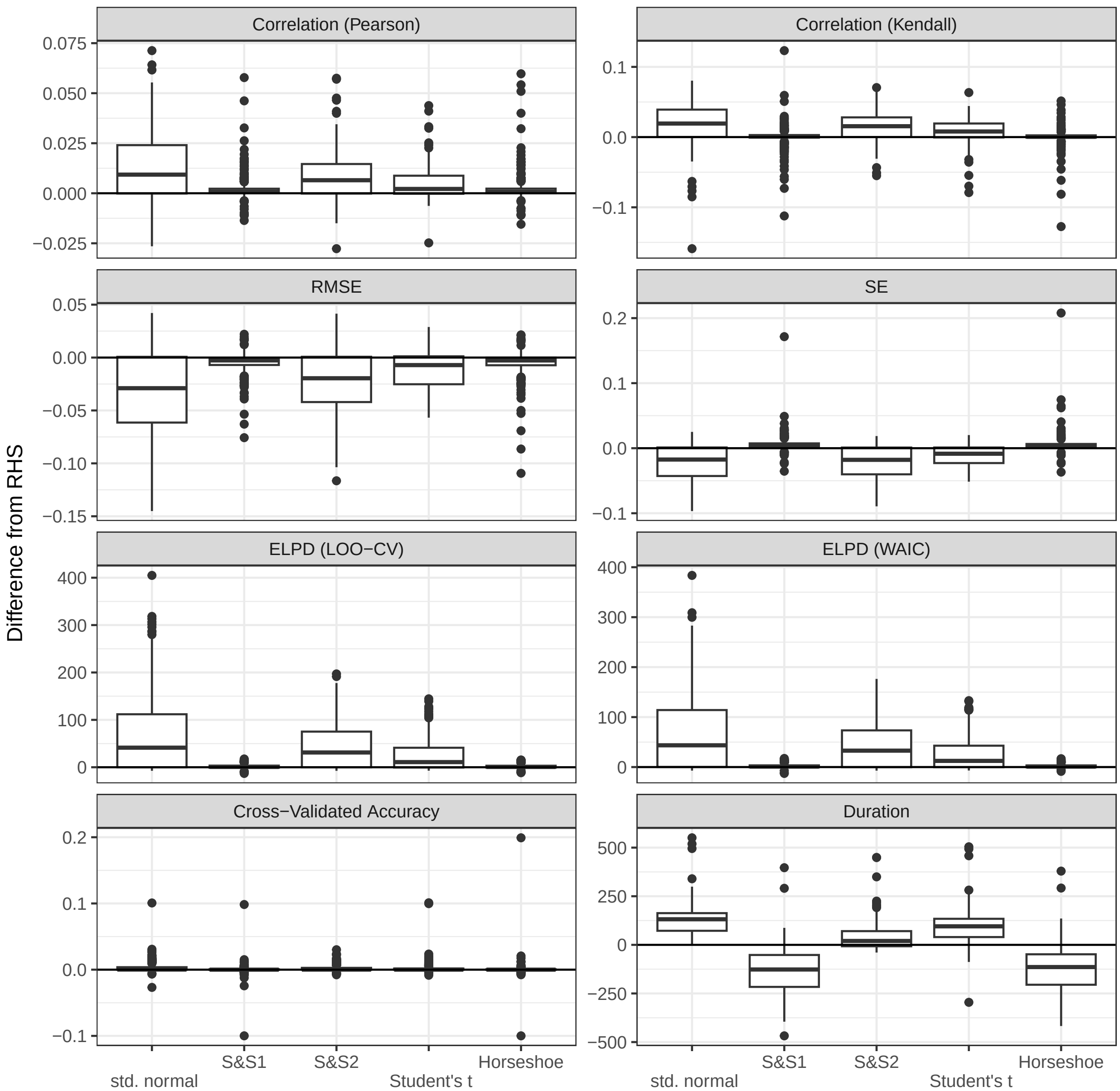


Figure 3: Simulation Results

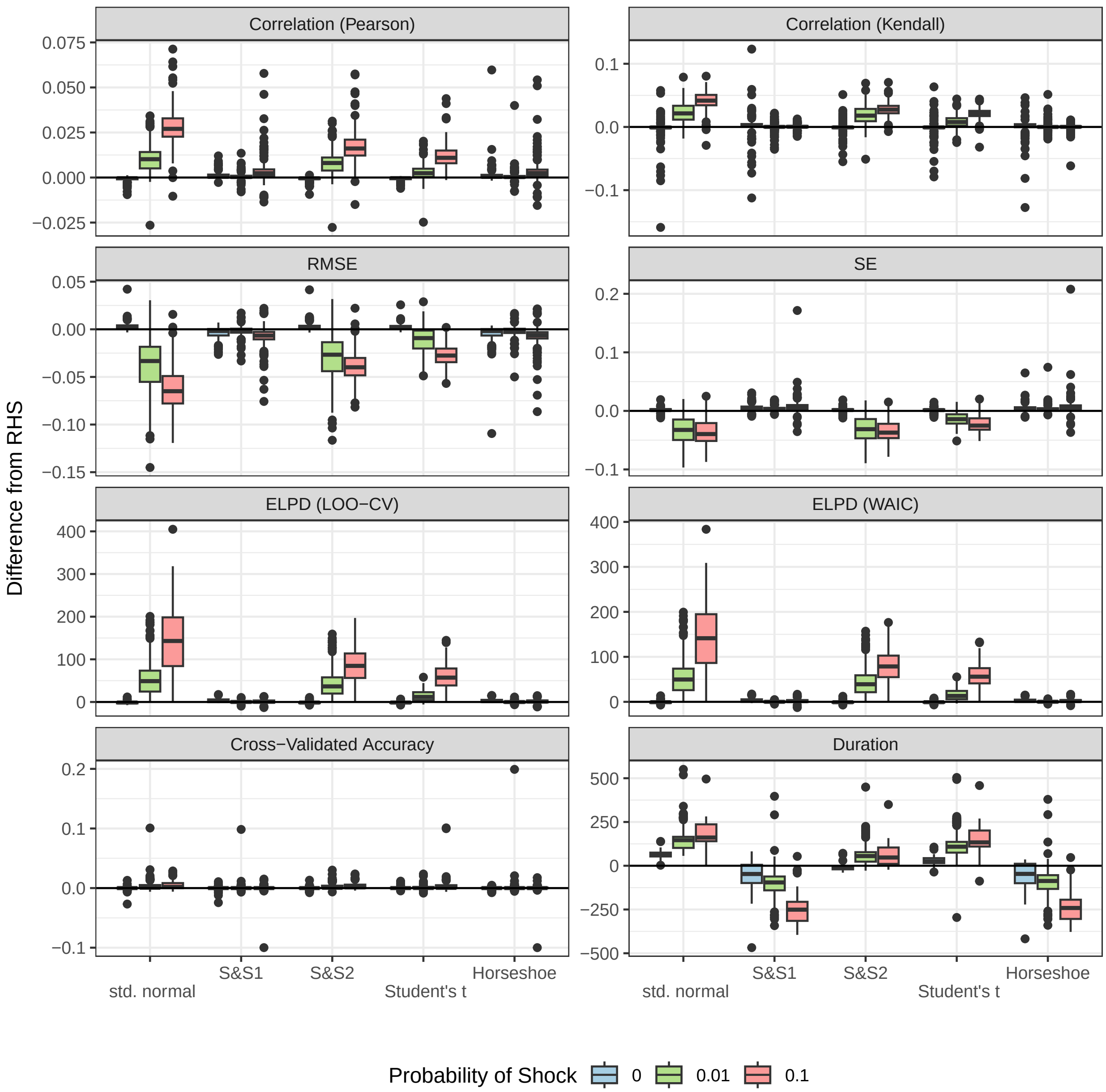


Figure 4: Simulation Results by Shock Probability

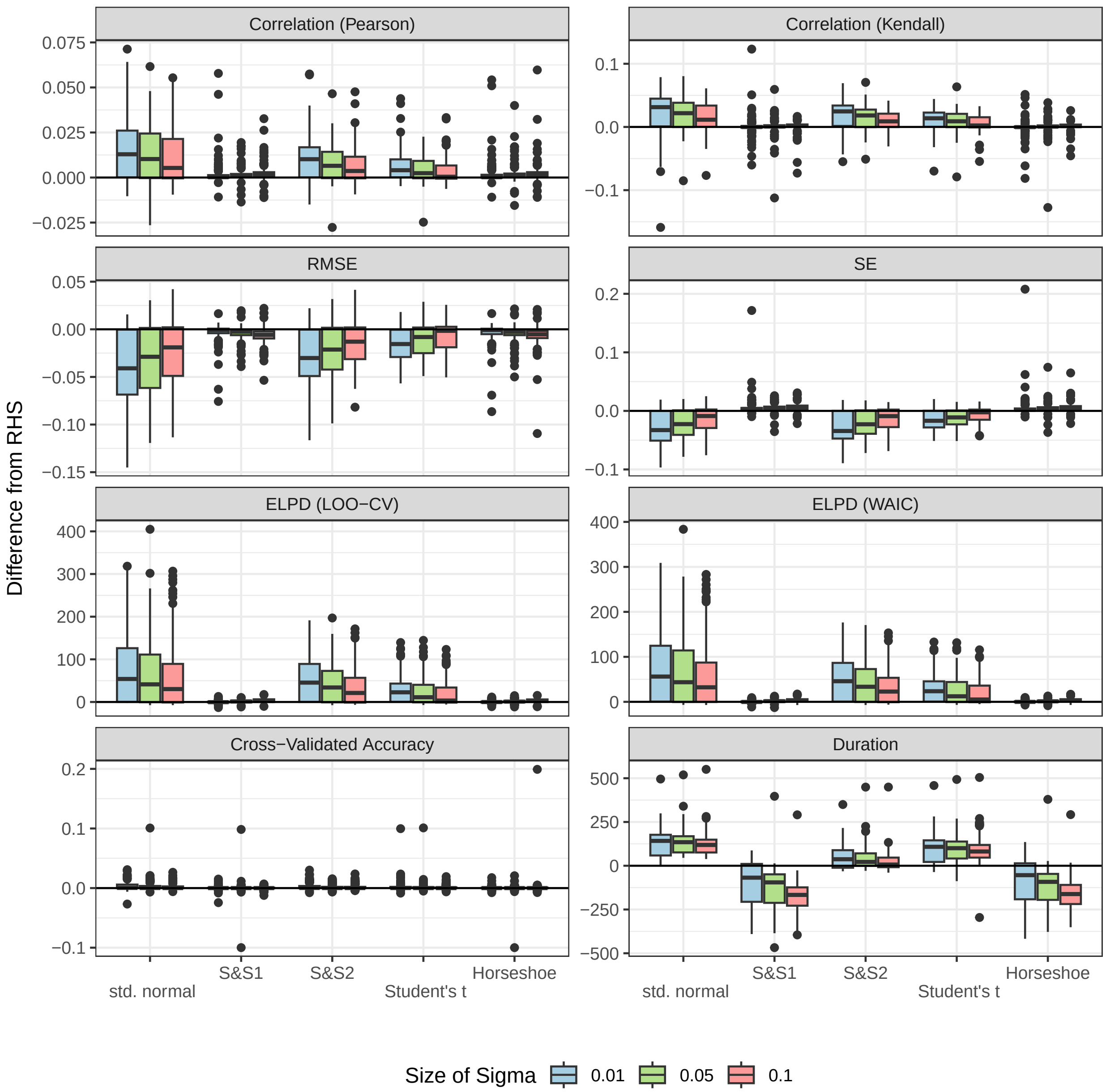


Figure 5: Simulation Results by σ

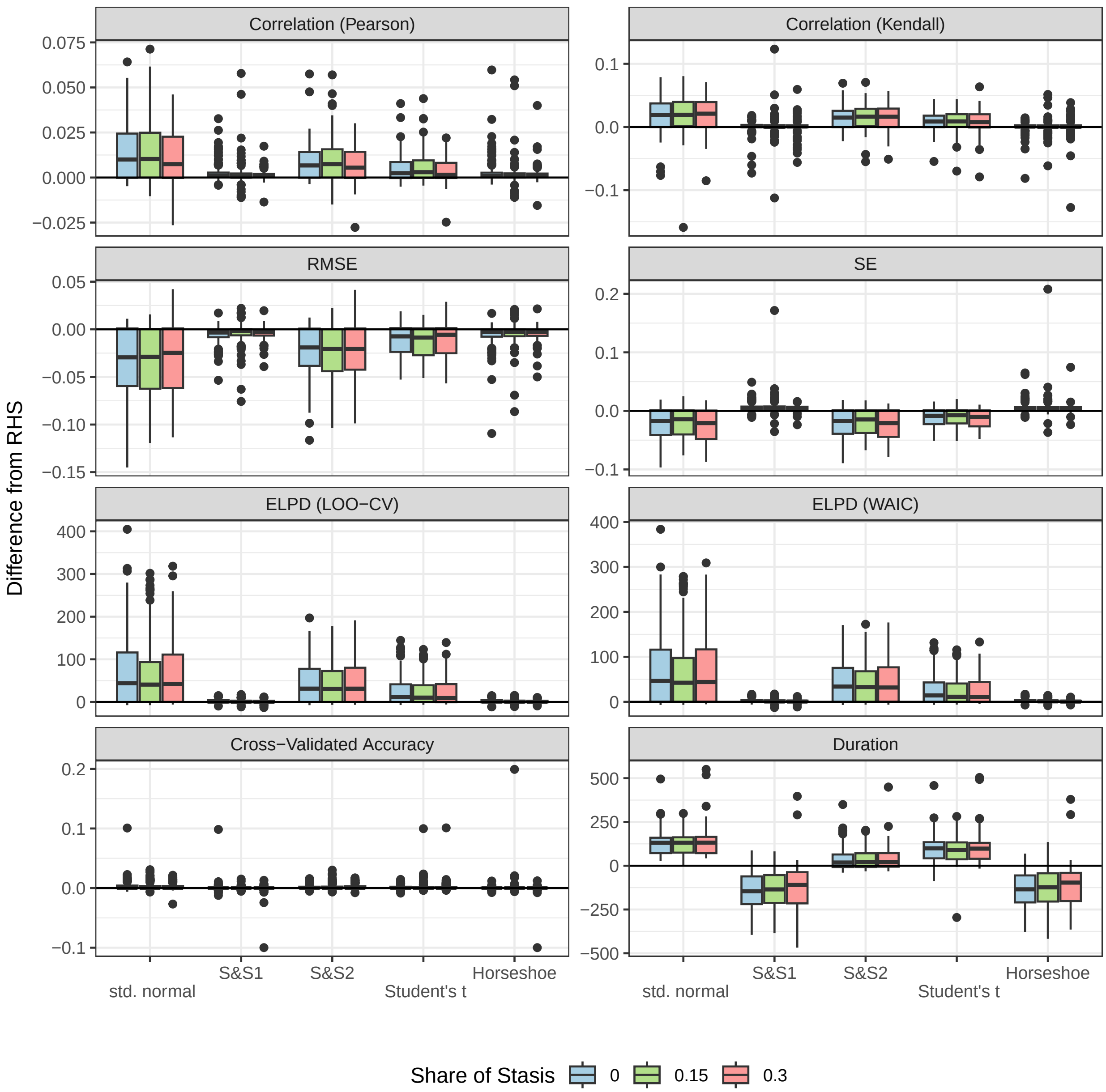


Figure 6: Simulation Results by Stasis

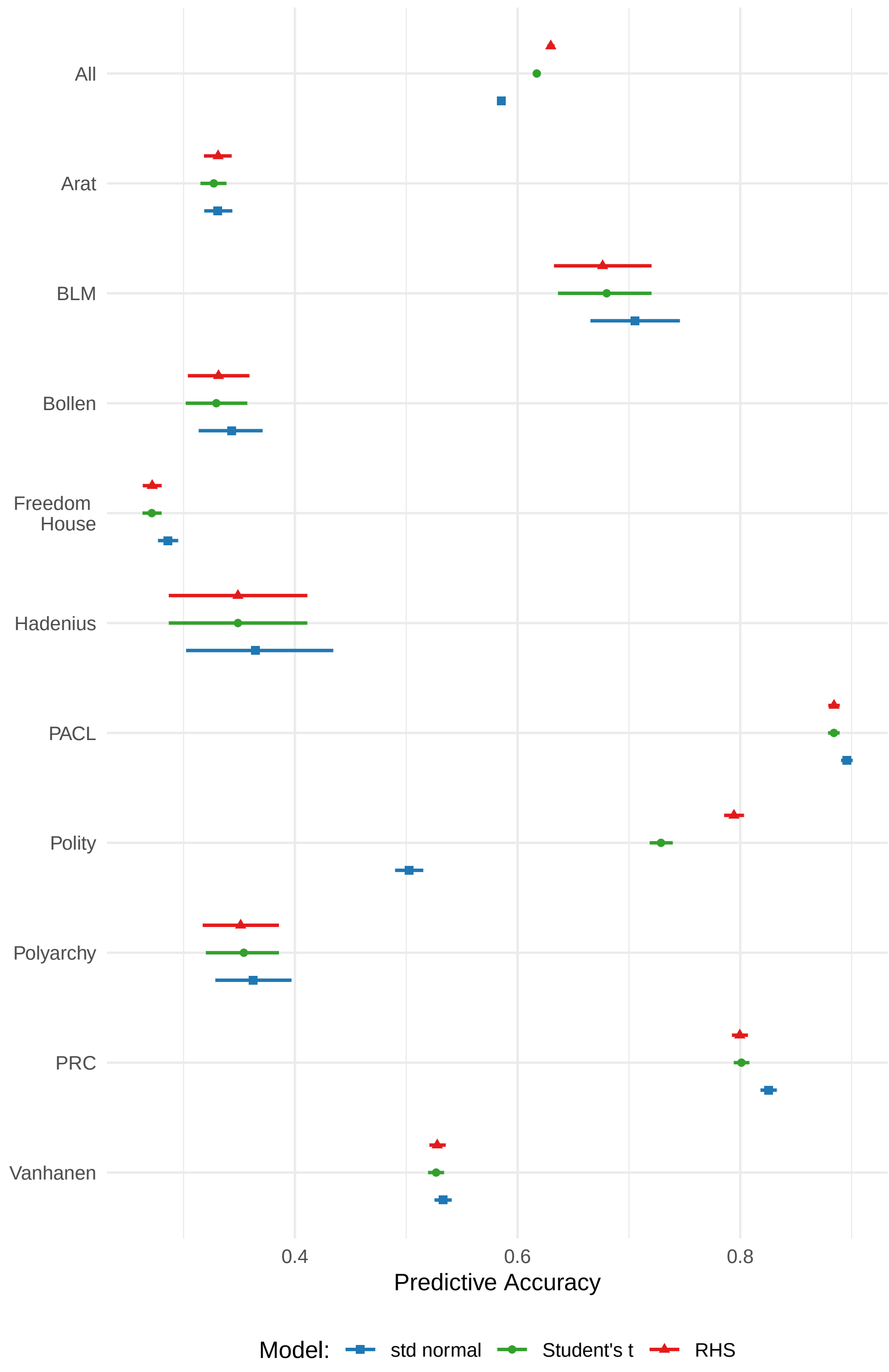
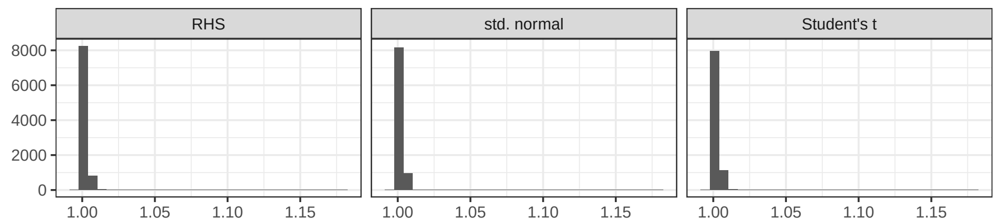
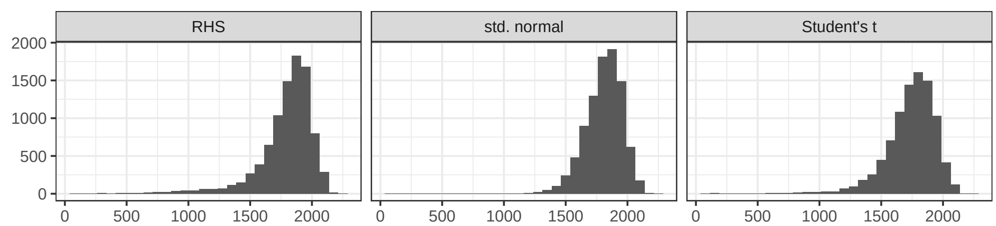


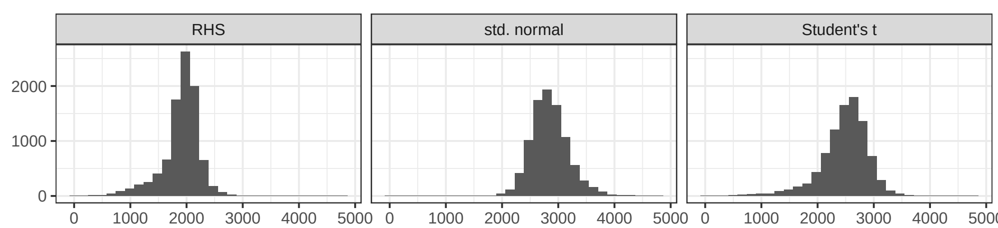
Figure 7: Predictive Accuracy by Indicator (median and 95% ci)



(a) $\hat{R}$.



(b) Effective sample size (tail).



(c) Effective sample size (bulk).

Figure 8: Convergence diagnostics

References

Arat, Z. F. (1991). *Democracy and Human Rights in Developing Countries*. Boulder, CO: Lynne Rienner Publishers.

Bollen, K. A. (2001). Cross-national indicators of liberal democracy, 1950-1990.

Bowman, K., F. Lehoucq, and J. Mahoney (2005). Measuring political democracy: Case expertise, data adequacy, and central america. *Comparative Political Studies 38*(8), 939–970.

Brandt, H., S. M. Chen, and D. J. Bauer (2023). Bayesian penalty methods for evaluating measurement invariance in moderated nonlinear factor analysis. *Psychological Methods*. Advance online publication.

Carvalho, C. M., N. G. Polson, and J. G. Scott (2010). The horseshoe estimator for sparse signals. *Biometrika 97*(2), 465–480.

Coppedge, M. and W. H. Reinicke (1991). Measuring polyarchy. In A. Inkeles (Ed.), *On Measuring Democracy: Its Consequences and Concomitants*, pp. 47–68. New Brunswick, NJ: Transaction Publishers.

Freedom House (2007). Freedom in the world. `http://www.freedomhouse.org`.

Gasiorowski, M. J. (1996). An overview of the political regime change data set. *Comparative Political Studies 29*(4), 469–483.

Gelman, A. and D. B. Rubin (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science 7*(4), 457 – 472.

George, E. I. and R. E. McCulloch (1993). Variable selection via gibbs sampling. *Journal of the American Statistical Association 88*(423), 881–889.

Hadenius, A. (1992). *Democracy and Development*. Cambridge: Cambridge University Press.

Ishwaran, H. and J. S. Rao (2005). Spike and slab variable selection: Frequentist and bayesian strategies. *The Annals of Statistics 33*(2), 730–773.

Marshall, M. G., K. Jaggers, and T. R. Gurr (2006). Polity iv: Political regime characteristics and transitions. `http://www.cidcm.umd.edu/polity/`.

Mitchell, T. J. and J. J. Beauchamp (1988). Bayesian variable selection in linear regression. *Journal of the american statistical association 83*(404), 1023–1032.

Pemstein, D., S. A. Meserve, and J. Melton (2010). Democratic compromise: A latent variable analysis of ten measures of regime type. *Political Analysis 18*(4).

Piironen, J. and A. Vehtari (2017). Sparsity information and regularization in the horseshoe and other shrinkage priors. *Electronic Journal of Statistics 11*(2), 5018 – 5051.

Polson, N. G. and J. G. Scott (2011). Shrink globally, act locally: sparse Bayesian regularization and prediction. In *Bayesian statistics 9*, pp. 501–538. Oxford Univ. Press, Oxford. With discussions by Bertrand Clark, C. Severinski, Merlise A. Clyde, Robert L. Wolpert, Jim e. Griffin, Philiip J. Brown, Chris Hans, Luis R. Pericchi, Christian P. Robert and Julyan Arbel.

Przeworski, A., M. E. Alvarez, J. A. Cheibub, and F. Limongi (2000). *Democracy and Development: Political Regimes and Material Well-Being in the World, 1950-1990*. Cambridge: Cambridge University Press.

Vanhanen, T. (2003). *Democratization: A Comparative Analysis of 172 Countries*. New York: Routledge.