Final Report

Bayesian Vector Autoregression

Ward B. Eiling (9294163), June 19, 2024

Table of contents

1 Introduction **2**

2 The Bayesian Person-specific VAR(p) Model **2**
2.1 Introduction and Notation 2
2.2 Priors 3

3 Methods **4**
3.1 Bayesian Estimation 4
3.2 Bayesian Model Selection 5

4 Results **5**
4.1 Bayesian Estimation 5
4.2 Posterior Analysis 6
4.3 Bayesian Model Selection 6

5 Discussion **7**

6 References **8**

7 Appendix A **9**
7.1 Model 1: Uninformative Conjugate Prior 9
7.2 Model 2: Normal-Wishart Prior 12
7.3 Model 3: Independent Normal-Wishart Prior 15

8 Appendix B **18**

1 Introduction

The rising popularity of the network model of psychopathology, which captures the dynamic interplay between symptoms over time, has led to the widespread use of vector autoregressive (VAR) models. VAR models are popular due to their flexibility in choosing the number of lags and their relative simplicity. However, the rich parameterization of VAR models introduces risks such as overfitting, imprecise inference, and significant uncertainty in future projections (Karlsson, 2012). Overfitting occurs when the model mistakes sample-specific noise for the true underlying signals or patterns in the population (Yarkoni & Westfall, 2017).

In line with extensive coverage in econometrics over the past few decades, the field of psychopathology has recently reemphasized the risk of overfitting in VAR models (Bulteel, Tuerlinckx, et al., 2018; Bulteel, Mestdagh, et al., 2018; Lafit et al., 2022). Overfitting in traditional regression models (e.g., ordinary least squares regression) often arises from *over-parameterization*: having too many parameters relative to the sample size (Babyak, 2004). More specifically, for each added lag, the number of regression coefficients increases by the square of the number of endogenous variables[1]. This issue is particularly relevant for small time series, which are common in applied psychopathology research due to considerations of client burden, potential for attrition, and limited resources. Consequently, researchers often opt for VAR models with a single lag, regardless of whether this makes sense substantively. This practice is not without consequences: "While being powerful forecast devices that can fit the data well, VAR-models may require relatively large lag lengths p in order to match the time series properties of the data which, with the many parameters to estimate can cause poor forecasting performance" (Karlsson, 2012).

The main Frequentist argument for utilizing the Bayesian approach to VAR models is that it may mitigate the issue of overfitting and the 'curse of dimensionality' by imposing a structure through informative data-centric prior beliefs, thus enabling the use of larger models (Bańbura et al., 2010). For instance, the Minnesota prior introduced by Litterman (1979) shrinks the parameters towards a useful benchmark, promoting the use of higher lags, reducing parameter uncertainty, and improving forecasting accuracy (Karlsson, 2012). From a Bayesian perspective, it may also be said that this prior captures widely held beliefs about the long-run properties of the data, which are not evident in the typically short samples used for estimation. Moreover, Bayesian estimates are conditional on the data and prior beliefs, reflecting inherent uncertainty in the posterior distribution. This differs from Frequentist estimates, which are point estimates that do not capture this uncertainty.

2 The Bayesian Person-specific VAR(p) Model

2.1 Introduction and Notation

In the multivariate person-specific VAR(p) model, data from each individual are treated separately, with each variable regressed on all endogenous variables M (where $m = 1, \dots, M$), including itself, for every lag up to p. The model is represented as:

$$\mathbf{y}_t = \mathbf{C} + \sum_{i=1}^{p} \mathbf{A}_i \mathbf{y}_{t-i} + \boldsymbol{\epsilon}_t$$

[1]The number of estimated parameters in a VAR model (with intercept) is given by $M + M^2 \times p$, where M represents the total number of endogenous variables and p the number of lags.

Here, $\mathbf{C}$ is an $M \times 1$ vector of intercepts, $\mathbf{A}_i$ are $M \times M$ coefficient matrices for lagged values, and $\boldsymbol{\epsilon}_t$ is an $M \times 1$ vector of error terms at time t. The error terms are assumed to be independent and identically distributed with a normal distribution, mean zero, and covariance matrix $\boldsymbol{\Sigma}$, allowing for contemporaneous correlations between error terms of different variables ($\sigma_{ij} \neq 0$ for $i \neq j$).

For simplicity, the VAR model can be expressed in matrix form using the matric-variate Normal Distribution. Define $\mathbf{Y}$ as a $T \times M$ matrix of observations, and $\mathbf{x}_t = (1, \mathbf{y}'_{t-1}, \dots, \mathbf{y}'_{t-p})$ as the regressor vector for each time period t. The regressor matrix $\mathbf{X}$ collects these vectors for all time periods, and with $K = 1 + M \times p$, $\mathbf{X}$ is a $T \times K$ matrix. Stack all coefficient matrices into $\mathbf{A} = (\mathbf{C}, \mathbf{A}_1, \dots, \mathbf{A}_p)$. In vectorized form, this becomes $\boldsymbol{\alpha} = \text{vec}(\mathbf{A})$, a $KM \times 1$ vector of coefficients. The VAR model can then be written compactly in two equivalent forms

$$\mathbf{Y} = \mathbf{X}\mathbf{A} + \mathbf{E}$$

$$\mathbf{y} = (\mathbf{I}_M \otimes \mathbf{X})\alpha + \boldsymbol{\epsilon}$$

where $\mathbf{y}$ and $\boldsymbol{\epsilon}$ are the vectorized forms of $\mathbf{Y}$ and $\mathbf{E}$, $\mathbf{I}_M$ is the $M \times M$ identity matrix, and $\otimes$ denotes the Kronecker product. Given the assumptions about the error terms, the likelihood function of the VAR model is expressed in terms of $\boldsymbol{\alpha}$ and $\boldsymbol{\Sigma}$. The posterior distribution of $\boldsymbol{\alpha}$ given $\boldsymbol{\Sigma}$ and the data is normally distributed, and the inverse of $\boldsymbol{\Sigma}$ follows a Wishart distribution:

$$\boldsymbol{\alpha} \mid \boldsymbol{\Sigma}, \mathbf{y} \sim \mathcal{N}(\hat{\boldsymbol{\alpha}}, \boldsymbol{\Sigma} \otimes (\mathbf{X}'\mathbf{X})^{-1})$$

$$\boldsymbol{\Sigma}^{-1} \mid \mathbf{y} \sim \mathcal{W}(\mathbf{S}^{-1}, T - K - M - 1)$$

where $\hat{\boldsymbol{\alpha}} = \text{vec}(\hat{\mathbf{A}})$ with $\hat{\mathbf{A}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ is the ordinary least squares (OLS) estimate of $\mathbf{A}$, and $\mathbf{S} = (\mathbf{Y} - \mathbf{X}\hat{\mathbf{A}})'(\mathbf{Y} - \mathbf{X}\hat{\mathbf{A}})$ is the sum of squared errors (SSE).

2.2 Priors

In the context of Bayesian VAR models, various priors can be used to impose structure and facilitate estimation. To maintain simplicity, we considered the natural conjugate prior and the independent Normal-Wishart prior.

The *Normal-Wishart prior* is the *natural conjugate prior* for normal multivariate regressions. Unlike the well-known Minnesota prior, the natural conjugate prior provides a complete Bayesian treatment by considering $\boldsymbol{\Sigma}$ as an unknown parameter. This prior is specified as: $\boldsymbol{\alpha} \mid \boldsymbol{\Sigma} \sim \mathcal{N}(\boldsymbol{\alpha}, \boldsymbol{\Sigma} \otimes \mathbf{V})$ and $\boldsymbol{\Sigma}^{-1} \sim \mathcal{W}(\mathbf{S}^{-1}, v)$, where $\boldsymbol{\alpha}$, $\mathbf{V}$, $\mathbf{S}$, and v are hyperparameters chosen by the researcher. Under this prior, the posterior distributions are: $\boldsymbol{\alpha} \mid \boldsymbol{\Sigma}, \mathbf{y} \sim \mathcal{N}(\boldsymbol{\alpha}, \boldsymbol{\Sigma} \otimes \mathbf{V})$ and $\boldsymbol{\Sigma}^{-1} \mid \mathbf{y} \sim \mathcal{W}(\mathbf{S}^{-1}, v)$, with updated parameters: $\mathbf{V} = [\mathbf{V}^{-1} + X'X]^{-1}$, $\boldsymbol{\alpha} = \text{vec}(\mathbf{A})$, $\mathbf{S} = \mathbf{S} + (\mathbf{Y} - \mathbf{X}\mathbf{A})'(\mathbf{Y} - \mathbf{X}\mathbf{A})$ and $v = T + v$. This allows for direct analytical solutions for posterior inferences and predictions. A special case of the conjugate prior is the *non-informative prior*, which sets $v = \mathbf{S} = \mathbf{V}^{-1} = c\mathbf{I}$ and $c \to 0$, yielding posterior results based on Frequentist OLS quantities without coefficient shrinkage.

The *independent Normal-Wishart prior* combines the flexibility of the well-known Minnesota prior with the Bayesian completeness of the natural conjugate prior. It assumes that the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\Sigma}$ are independently distributed as follows: $\boldsymbol{\alpha} \sim \mathcal{N}(\boldsymbol{\beta}, \mathbf{V}_\beta)$ and $\boldsymbol{\Sigma}^{-1} \sim \mathcal{W}(\mathbf{S}^{-1}, v)$. This prior's posterior distributions are then: $\boldsymbol{\alpha} \mid \boldsymbol{\Sigma}, \mathbf{y} \sim \mathcal{N}(\boldsymbol{\alpha}, \boldsymbol{\Sigma} \otimes \mathbf{V}_\beta)$ and $\boldsymbol{\Sigma}^{-1} \mid \mathbf{y} \sim \mathcal{W}(\mathbf{S}^{-1}, v)$, where $\boldsymbol{\beta}$, $\mathbf{V}_\beta$, $\mathbf{S}$, and v are prior hyperparameters. This approach is more flexible than the Normal-Wishart prior, allowing incorporation of additional explanatory variables and more complex prior covariance structures.

3 Methods

3.1 Bayesian Estimation

Extensive testing with simulated multivariate time series data ensured accurate estimates and achieved convergence, though not elaborated here for brevity. To illustrate the Bayesian VAR(1) model, we examine the short-term relationship between mood satisfaction and restlessness using 100 timepoints in a single patient from the dataset of Kossakowski et al. (2017), where each item was measured on a 5-point scale. Three prior specifications were considered for the BVAR model:

- **Model 1: The conjugate non-informative prior.** No hyperparameters are specified.
- **Model 2: The conjugate Normal-Wishart Prior.** Combines a normal prior for coefficients with an inverse-Wishart prior for the covariance matrix. Hyperparameters chosen are $\boldsymbol{\alpha} = 0_{KM\times 1}$, $\mathbf{V} = 10\mathbf{I}_K$, $v = M + 1$, and $\mathbf{S}^{-1} = \mathbf{I}_M$.
- **Model 3: The Independent Normal-Wishart Prior.** Assumes independence between coefficients and covariance matrix. Hyperparameters chosen are $\boldsymbol{\beta} = 0_{KM\times 1}$, $\mathbf{V}_\beta = 10I_{KM}$, $v = M + 1$, and $\mathbf{S}^{-1} = \mathbf{I}_M$.

The estimation involved Monte Carlo integration for models 1 and 2, where the posterior directly depends on OLS estimates, and Gibbs sampling for model 3. The function `Bayesian_Bivariate_VAR1` was developed to estimate these models, involving the following steps:

1. **Data Preparation**: Endogenous variables were lagged to create matrices Y and X, with an intercept added to X. The number of observations (T), endogenous variables (M), and coefficients (K) are defined. The OLS estimators for the coefficients and the error covariance matrix are specified as initial values (for models 1 and 2).
2. **Prior Specification**: Hyperparameters and prior distributions were specified based on the selected prior.
3. **Posterior Sampling**: Posterior draws of coefficients and covariance matrix were obtained for each chain and iteration, excluding burn-in samples. All models were ran with 2 separate markov chains. For models 1 and 2, 10,000 iterations were used without burn-in[2], while model 3 employed 15,000 iterations with the first 5,000 as burn-in. Sampling methods varied by prior:
    - **Model 1**: The posterior of $\mathbf{A}$ is sampled from a multivariate normal distribution given $\boldsymbol{\Sigma}$, the data (incl. OLS coefficients). The posterior of $\boldsymbol{\Sigma}$ is sampled from an inverse Wishart distribution given the data (incl. SSE).
    - **Model 2**: The posterior of $\mathbf{A}$ is sampled from a multivariate normal distribution given $\boldsymbol{\Sigma}$, the data (incl. OLS coefficients) and prior information. The posterior of $\boldsymbol{\Sigma}$ is sampled from an inverse Wishart distribution given $\mathbf{A}$, the data (incl. SSE) and prior information.
    - **Model 3**: The posterior of $\mathbf{A}$ is sampled from a multivariate normal distribution given $\boldsymbol{\Sigma}$, the data and prior information. The posterior of $\boldsymbol{\Sigma}$ is sampled from an inverse Wishart distribution given $\mathbf{A}$, the data (incl. SSE) and prior information.

[2] A burn-in period is not strictly necessary for the models two with the conjugate priors, as they are depend directly on and are initialized with OLS quantities.

4. **Posterior Analysis**: Summary statistics computed from posterior densities, including mean, standard deviation, and credible intervals, provided insights into parameter estimates.
5. **Convergence Diagnostics**: Trace plots, density plots, autocorrelation functions, Gelman-Rubin statistic, and Monte Carlo error were employed to assess convergence

3.2 Bayesian Model Selection

To compare the performance of the three different models, we used the Bayes Factor as a model selection criterion. The Bayes Factor quantifies the relative evidence for one model over another by taking the ratio of their marginal likelihoods, which represent the integral of the likelihood function over the parameter space.

There are various methods to estimate the marginal likelihood of a gibbs-sampler, including the Harmonic Mean Estimator (Newton & Raftery, 1994), Chib's method (Chib, 1995), and the closed-form solution of the marginal likelihood is available for the two conjugate priors, as a multivariate/matricvariate t-distribution (Karlsson, 2012; Koop & Korobilis, 2010; Murphy, 2007). Due to its simplicity, we opted for approximating the marginal likelihood using the Harmonic Mean Estimator, which is calculated for each chain[3]:

$$p(\mathbf{y} \mid \boldsymbol{\alpha}^{(n)}, \boldsymbol{\Sigma}^{(n)}) = \frac{1}{N} \sum_{n=1}^{N} p(\mathbf{y} \mid \boldsymbol{\alpha}^{(n)}, \boldsymbol{\Sigma}^{(n)})$$

where $p(y \mid \alpha^{(n)}, \Sigma^{(n)})$ denotes the likelihood of the data given the parameters at iteration n and N is the total number of iterations. After doing this for all three models, we computed the Bayes factor (Kass & Raftery, 1995) for each chain as the ratio of two marginal likelihoods:

$$BF_{ij} = \frac{p(\mathbf{y} \mid M_i)}{p(\mathbf{y} \mid M_j)} = \frac{m_i}{m_j}$$

where M_i and M_j are the two models being compared, and m_i and m_j are the marginal likelihoods of the respective models. A Bayes Factor greater than 1 indicates that the data is more likely under model i than model j.

4 Results

4.1 Bayesian Estimation

To assess whether the sampled data of the three models converged to a stable constant distribution, we employed (1) history/trace plots, (2) autocorrelation plots, (3) the Gelman-Rubin statistic and (4) Monte Carlo error. For both the initial assessment with simulated data and the empirical data, the traceplots, which show the evolution of the sampled values over the iterations, displayed stable intermingling of the chains, representing a 'fat hairy catterpillar' (see Appendix A; Figure 1, 4, 7). Autocorrelation plots demonstrated rapid decay to low values: models 1 and 2 exhibited correlations near 0 at from lag 1 onwards, while model 3 showed autocorrelations below 0.15 at lag 1 and to 0 at lag 2 (see Appendix A; Figure 3, 6, 9), indicating efficient mixing

[3]Computing the Marginal likelihoods and Bayes Factors separately for each chain allows us to detect potential instability of the Harmonic Mean Estimator.

of chains. The Gelman-Rubin statistic for all parameters in all models was exactly 1.000 (see Table 2, 3, and 4), indicating minimal variance between chains and suggesting good convergence. The Monte Carlo error, quantifying uncertainty in posterior estimates, was less than 5% of the sample standard deviation for each parameter in all models (see Table 2, 3, and 4), further supporting robust convergence.

Note that although we have not found any evidence that the markov chains failed to convergence, we have not and cannot prove that convergence occurred using the discussed diagnostic checks. Nevertheless, the strong convergence observed in models 1 and 2 aligns with expectations due to their closed-form posterior solutions.

4.2 Posterior Analysis

The posterior means of the coefficients and covariance matrix $\mathbf{\Sigma}$ for each model are summarized in Appendix B (Tables 2, 3, and 4). Upon closer examination, models 1 and 2 exhibit similar estimates for the coefficients in matrix $\mathbf{A}$, while model 3 shows notably smaller estimates, particularly for the intercepts. The elements of the covariance matrix $\mathbf{\Sigma}$ show minimal variation across all models, albeit slightly differing in model 3. Posterior standard deviations of the coefficients and covariance matrix are comparable across models, except for the intercepts, which are substantially smaller in model 3. The credible intervals also reflect these differences, aligning with the smaller means and standard deviations observed in model 3.

Notably, the intercepts exhibit the highest uncertainty, evident from their larger standard deviations and wider credible intervals. The minimal variation in estimates of $\mathbf{\Sigma}$ across all models suggest that the choice of prior has minimal impact on its estimation, indicating robustness to prior specification. In contrast, the varying estimates of matrix $\mathbf{A}$ across models underscore the significant influence of prior choice, particularly the prior independence present and the lack of incorporation of the OLS coefficients in sampling procedure in model 3.

Given that the psychological measures were based on a 5-point Likert scale (ranging from 1 to 5), it is concerning that intercept values below 1 suggest an equilibrium level below the lowest possible score when past influences are absent. Moreover, credible intervals indicating 95% confidence that intercepts fall between 0.25 and 1.5 raise questions about the suggested confidence in and plausibility of these estimates. Models 1 and 2 provide more plausible intervals, suggesting that we may be 95% confident that the baseline level of restlessness falls between 0.6 and 2.2, and that the baseline level of mood satisfaction falls between 1.8 and 4.1.

4.3 Bayesian Model Selection

Table 1 presents the Bayes Factors computed for each model comparison. The Bayes Factor comparing model 1 to model 2 indicates that after observing the data, model 1 is approximately 0.9 times more likely than model 2, suggesting a very slight preference for model 2, albeit not strongly supported. In contrast, the Bayes Factors comparing model 1 to model 3 indicates that after observing the data, model 1 is approximately 18,000 to 20,000 times more likely than model 3, suggesting an extreme preference for model 1 over model 3. Comparatively, the Bayes Factors for model 2 versus model 3 indicate that model 2 is approximately 20,000 to 23,000 times more likely than model 3, indicating an extreme preference for model 2 over model 3. These results show that models with conjugate priors are strongly favored over the model with the independent Normal-Wishart prior, with a preference for the Normal-Wishart prior model over the non-informative prior model.

Table 1: Bayes Factors for each chain

	BF_12	BF_13	BF_23
chain 1	0.872	18104.98	20753.76
chain 2	0.876	19825.90	22630.30

5 Discussion

The current investigation highlighted that the absence of evidence for non-convergence does not necessarily confirm that the Markov chains have indeed converged. Diagnostic checks such as history/trace plots, autocorrelation plots, the Gelman-Rubin statistic, and Monte Carlo error suggested that the Markov chains had reached a stable distribution. While this aligns with our expectations pertaining the Monte-Carlo processes of models 1 and 2, the convergence results may be misleading for model 3.

Posterior analysis revealed consistent estimates for the coefficients in matrix $\mathbf{A}$ across the first two models. However, Model 3 produced substantially different and implausible values, suggesting potential bias. This discrepancy may indicate a lack of convergence, potentially caused by poor initialization, leading the markov chain to a local rather than a global mode. Although Models 2 and 3 used identical hyperparameters and relied on the sum of squared errors (SSE) estimates, the sampling procedure in Model 3 did not depend on the OLS coefficients. This might explain why the $\mathbf{\Sigma}$ estimates were similar across all models, but the $\mathbf{A}$ estimates were not.

Bayes Factors indicated an extreme preference for models with conjugate priors over the model with the independent Normal-Wishart prior. There was a slight preference for the model with the Normal-Wishart prior over the non-informative prior. The significant support for the conjugate models further implies potential issues with the posterior estimates of Model 3. Additionally, the lack of fine-tuning of hyperparameters in Model 3 may have exacerbated this issue, while Model 2, relying on OLS coefficients, appeared more robust. However, due to time constraints, hyperparameters for Models 2 and 3 were not optimized to achieve intended shrinkage effects, evident in relatively weak model support and similar uncertainties (e.g., credible intervals).

Due to time constraints, we used the Harmonic Mean Estimator to compute marginal likelihoods. Ideally, for models 1 and 2 with conjugate priors, direct computation using the multivariate t-distribution would have provided closed-form solutions. Model 3 would have benefited from the method proposed by Chib (1995), known for producing more stable estimates. These choices could have influenced the Bayes Factors and overall result validity, potentially contributing to the extreme values observed. However, despite these methodological differences, consistent conclusions were drawn across different chains, suggesting reliable results from our chosen approach.

Similarly, due to time constraints, multiple lags were not implemented despite their potential benefits mentioned in the introduction, which could enhance Bayesian VAR model shrinkage. According to (Karlsson, 2012), multiple lags are crucial for capturing data's time series properties accurately. Consequently, the simplicity of models used in this empirical example might have resulted in limited validity and robustness.

6 References

Babyak, M. A. (2004). What you see may not be what you get: A brief, nontechnical introduction to overfitting in regression-type models. *Psychosomatic Medicine*, *66*(3), 411–421. https://doi.org/10.1097/01.psy.0000127692.23278.a9

Bańbura, M., Giannone, D., & Reichlin, L. (2010). Large bayesian vector auto regressions. *Journal of Applied Econometrics*, *25*(1), 71–92. https://www.jstor.org.proxy.library.uu.nl/stable/25608797

Bulteel, K., Mestdagh, M., Tuerlinckx, F., & Ceulemans, E. (2018). VAR(1) based models do not always outpredict AR(1) models in typical psychological applications. *Psychological Methods*, *23*(4), 740–756. https://doi.org/10.1037/met0000178

Bulteel, K., Tuerlinckx, F., Brose, A., & Ceulemans, E. (2018). Improved insight into and prediction of network dynamics by combining VAR and dimension reduction. *Multivariate Behavioral Research*, *53*(6), 853–875. https://doi.org/10.1080/00273171.2018.1516540

Chib, S. (1995). Marginal likelihood from the gibbs output. *Journal of the American Statistical Association*, *90*(432), 1313–1321. https://doi.org/10.2307/2291521

Karlsson, S. (2012). *Forecasting with bayesian vector autoregressions*. https://econpapers.repec.org/paper/hhsoruesi/2012_5f012.htm

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795. https://doi.org/10.1080/01621459.1995.10476572

Koop, G., & Korobilis, D. (2010). Bayesian multivariate time series methods for empirical macroeconomics. *Foundations and Trends® in Econometrics*, *3*(4), 267–358. https://doi.org/10.1561/0800000013

Kossakowski, J. J., Groot, P. C., Haslbeck, J. M. B., Borsboom, D., & Wichers, M. (2017). Data from 'critical slowing down as a personalized early warning signal for depression'. *Journal of Open Psychology Data*, *5*, 1–3. https://doi.org/10.5334/jopd.29

Lafit, G., Meers, K., & Ceulemans, E. (2022). A systematic study into the factors that affect the predictive accuracy of multilevel VAR(1) models. *Psychometrika*, *87*(2), 432–476. https://doi.org/10.1007/S11336-021-09803-Z/FIGURES/12

Litterman, R. B. (1979). *Techniques of forecasting using vector autoregressions*. https://doi.org/10.21034/wp.115

Murphy, K. (2007). *Conjugate bayesian analysis of the gaussian distribution*.

Newton, M. A., & Raftery, A. E. (1994). Approximate bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society. Series B (Methodological)*, *56*(1), 3–48. https://www.jstor.org/stable/2346025

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, *12*(6), 1100–1122. https://doi.org/10.1177/1745691617693393

7 Appendix A

7.1 Model 1: Uninformative Conjugate Prior

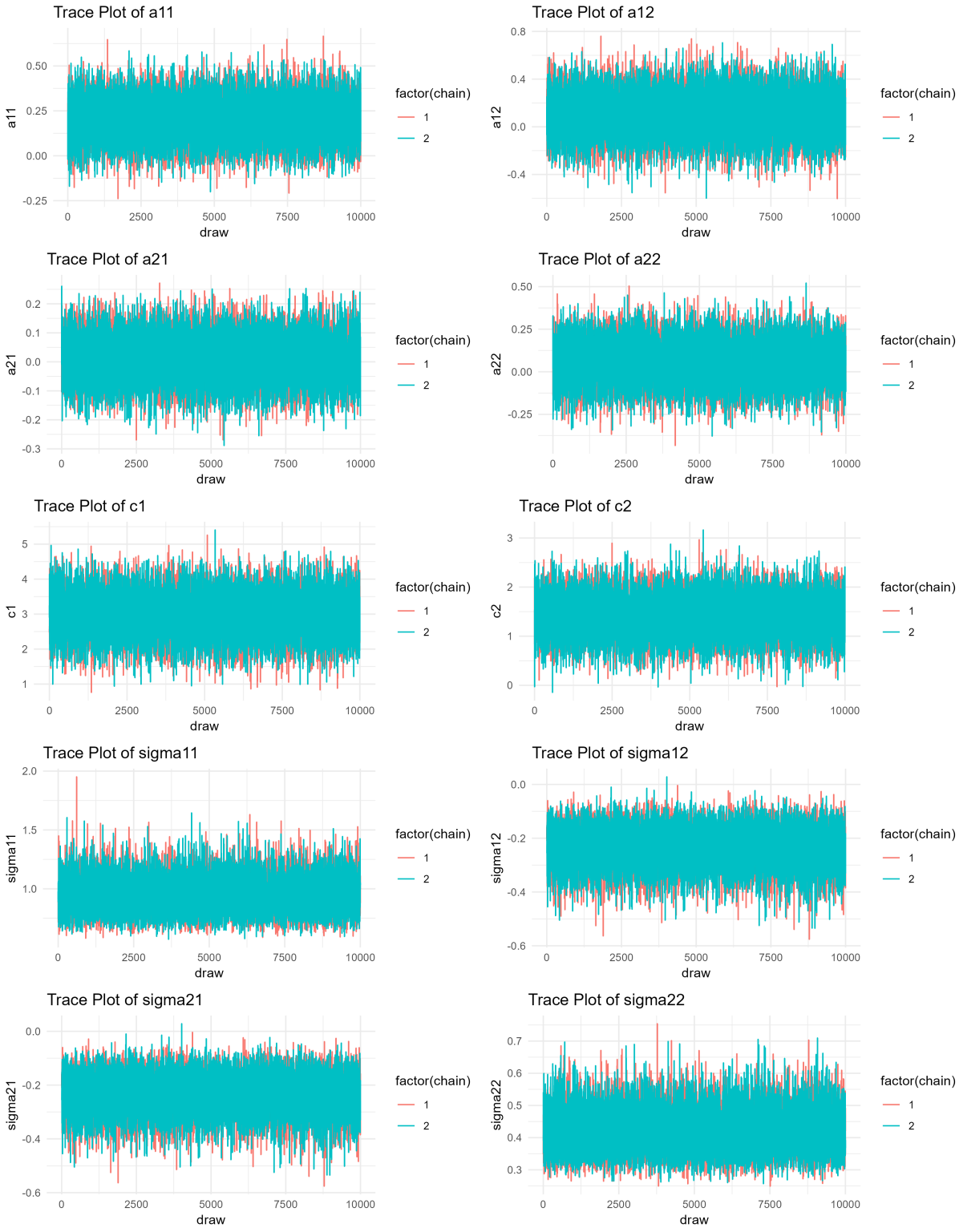

Figure 1: Trace plots of the coefficients and covariance matrix for the VAR(1) model with the uninformative conjugate prior

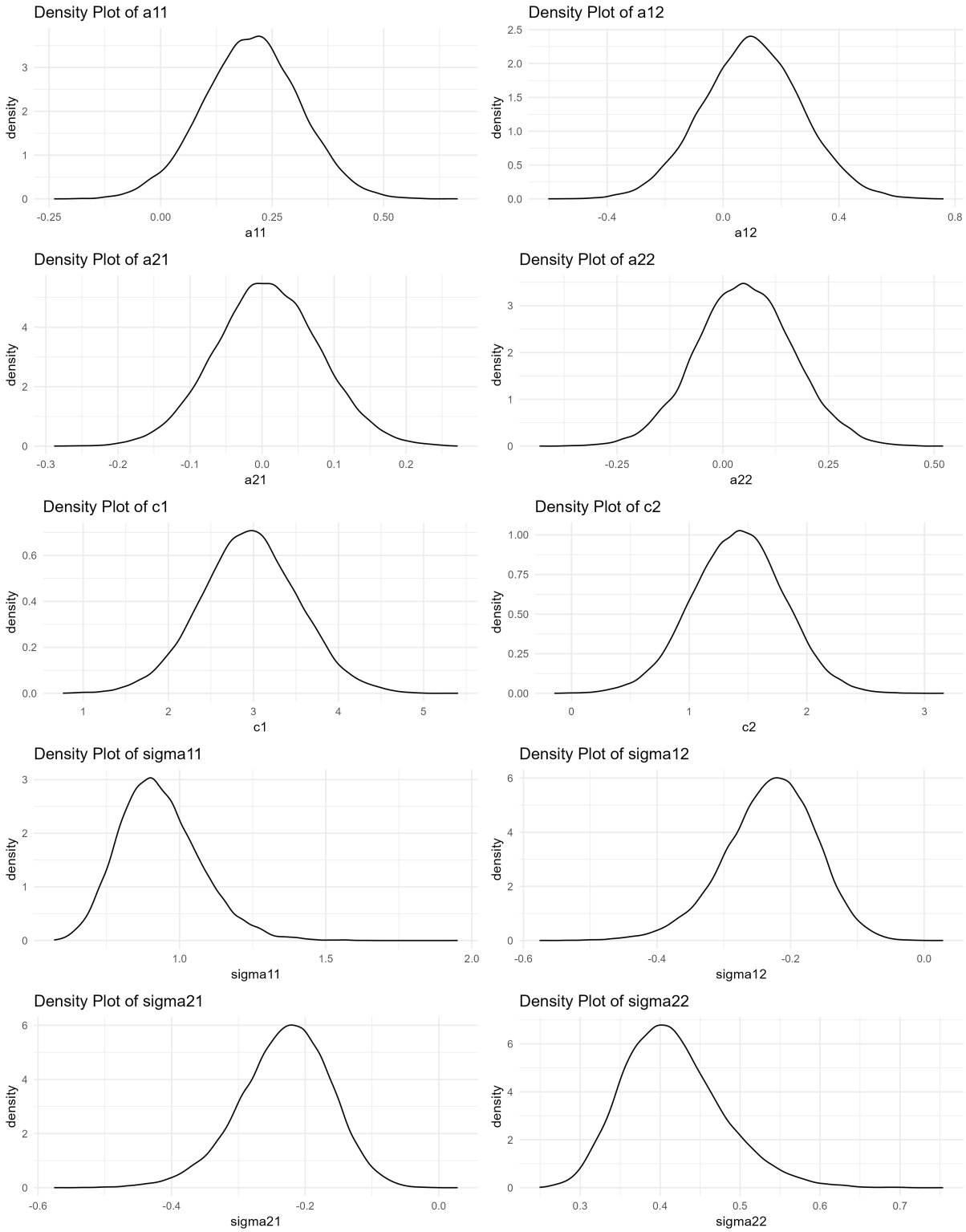

Figure 2: Density plots of the coefficients and covariance matrix for the VAR(1) model with the uninformative conjugate prior

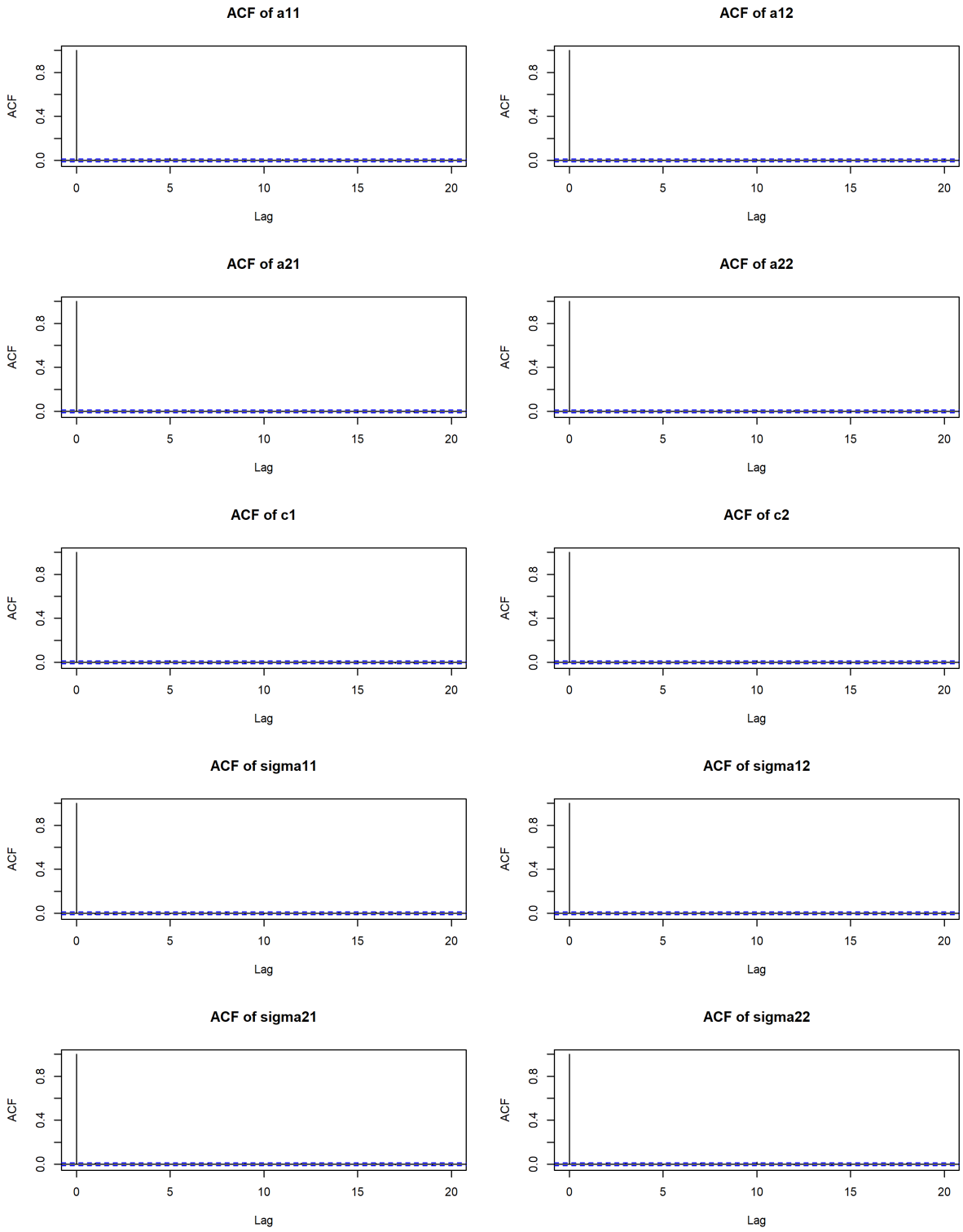

Figure 3: Auto-correlation plots of the coefficients and covariance matrix for the VAR(1) model with the uninformative conjugate prior

7.2 Model 2: Normal-Wishart Prior

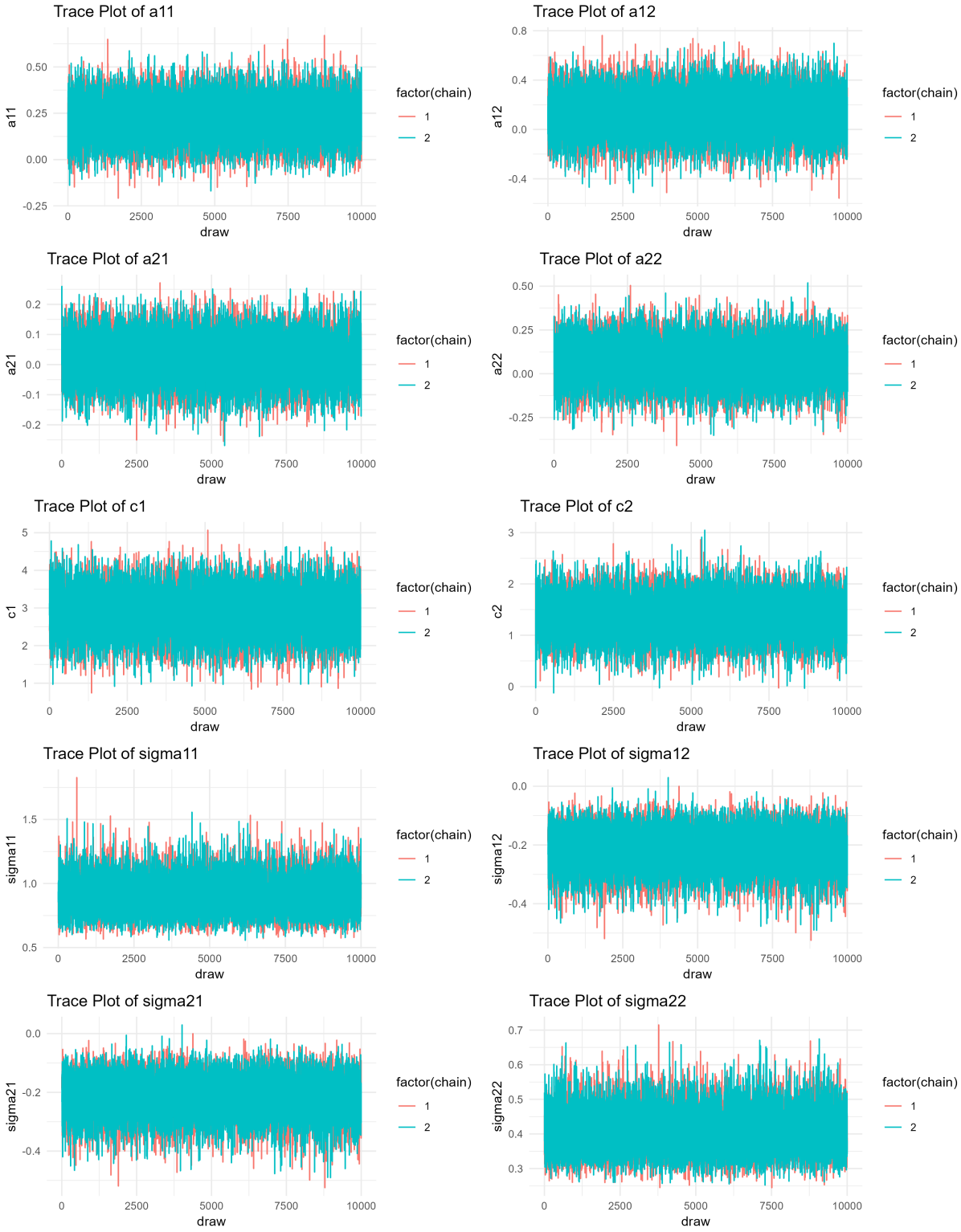

Figure 4: Trace plots of the coefficients and covariance matrix for the VAR(1) model with the Normal-Wishart prior

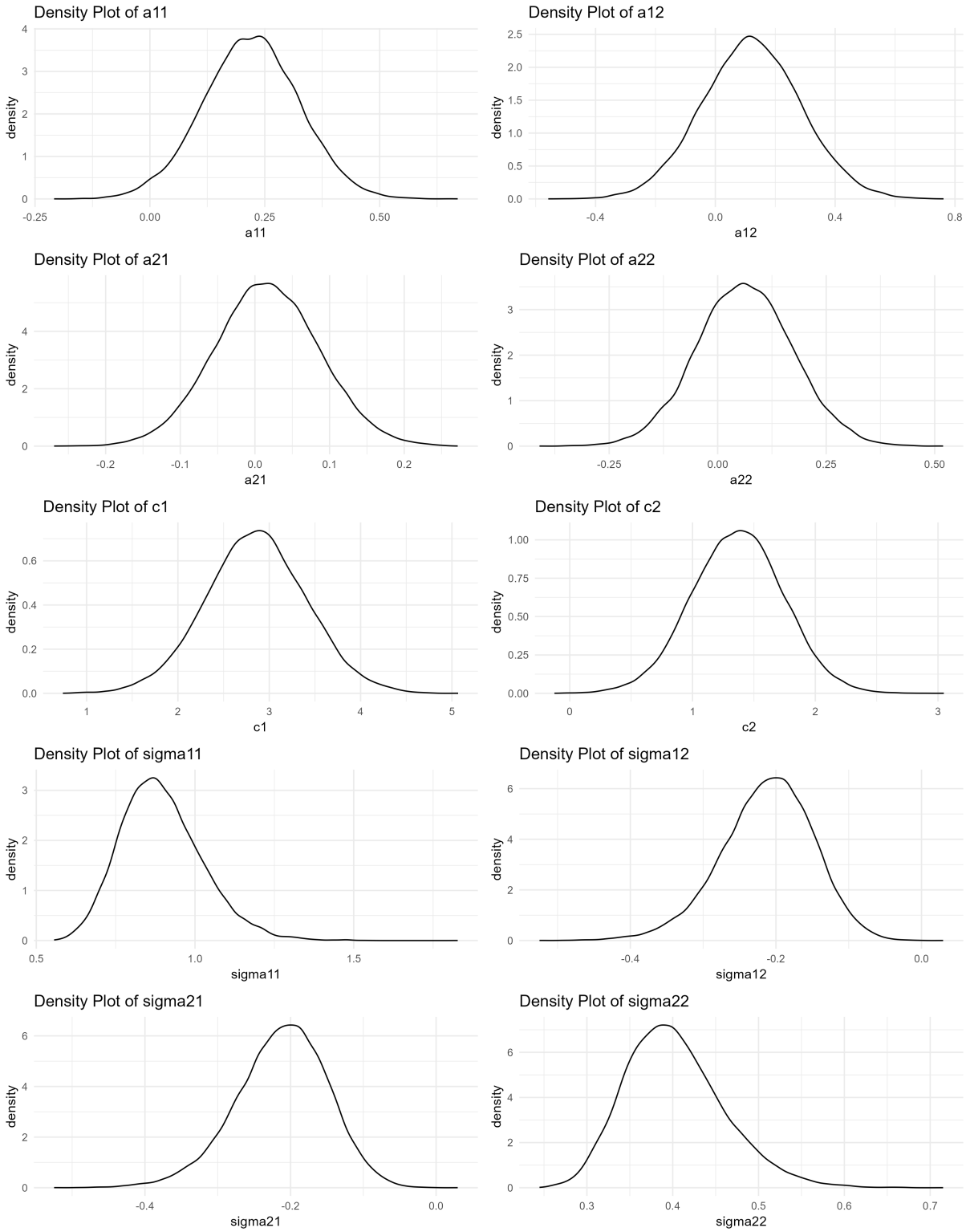

Figure 5: Density plots of the coefficients and covariance matrix for the VAR(1) model with the Normal-Wishart prior

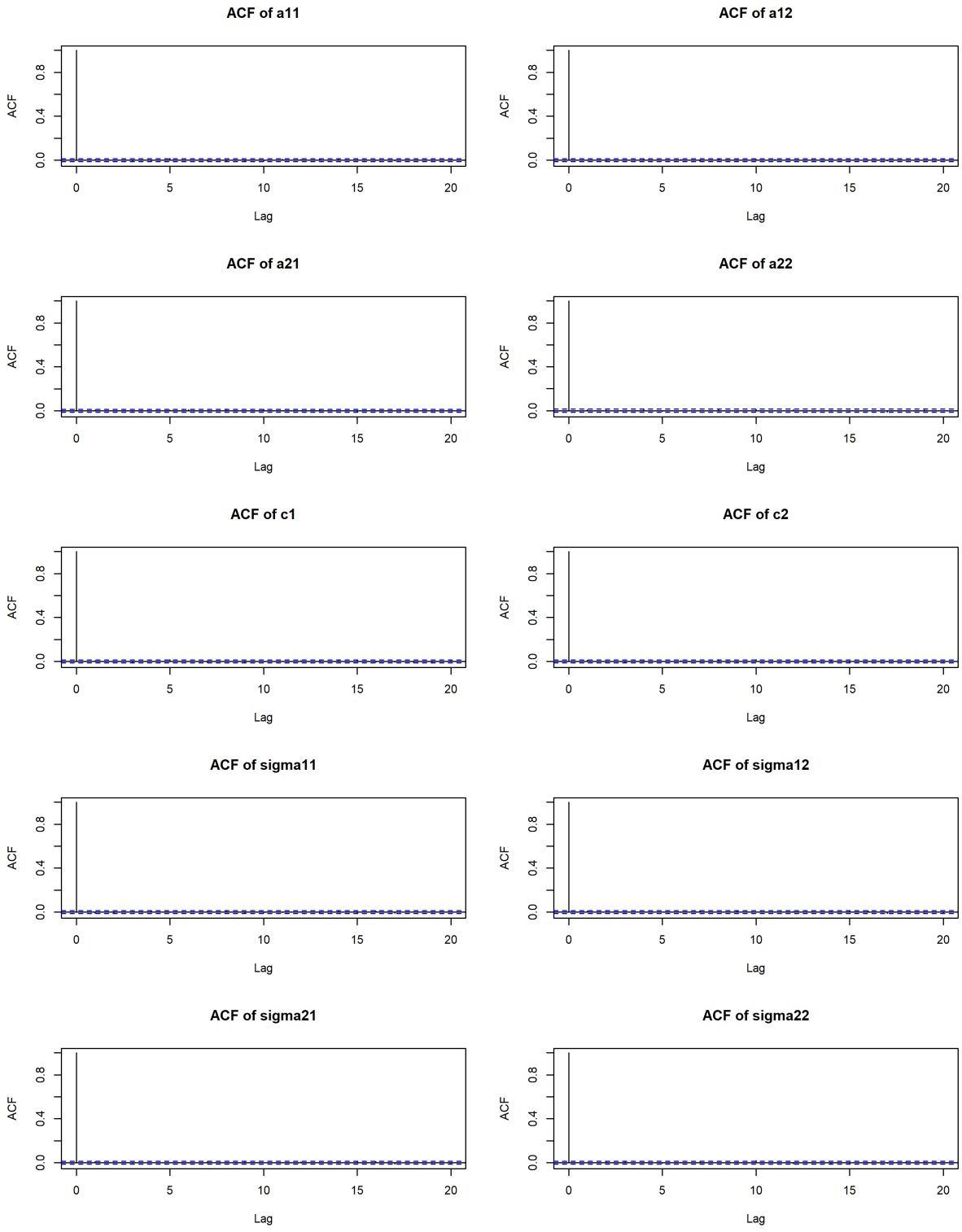

Figure 6: Auto-correlation plots of the coefficients and covariance matrix for the VAR(1) model with the Normal-Wishart prior

7.3 Model 3: Independent Normal-Wishart Prior

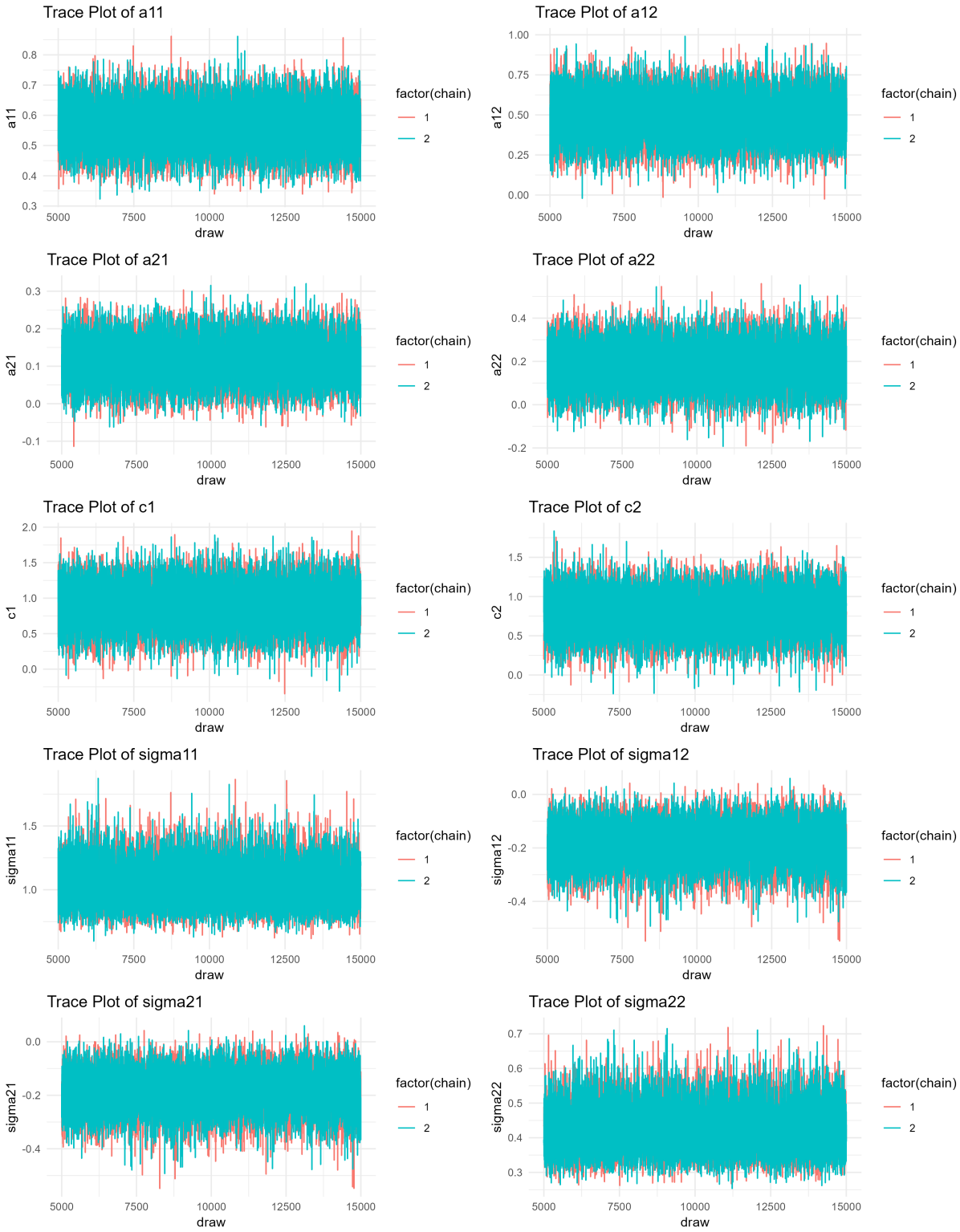

Figure 7: Trace plots of the coefficients and covariance matrix for the VAR(1) model with the Independent Normal-Wishart prior

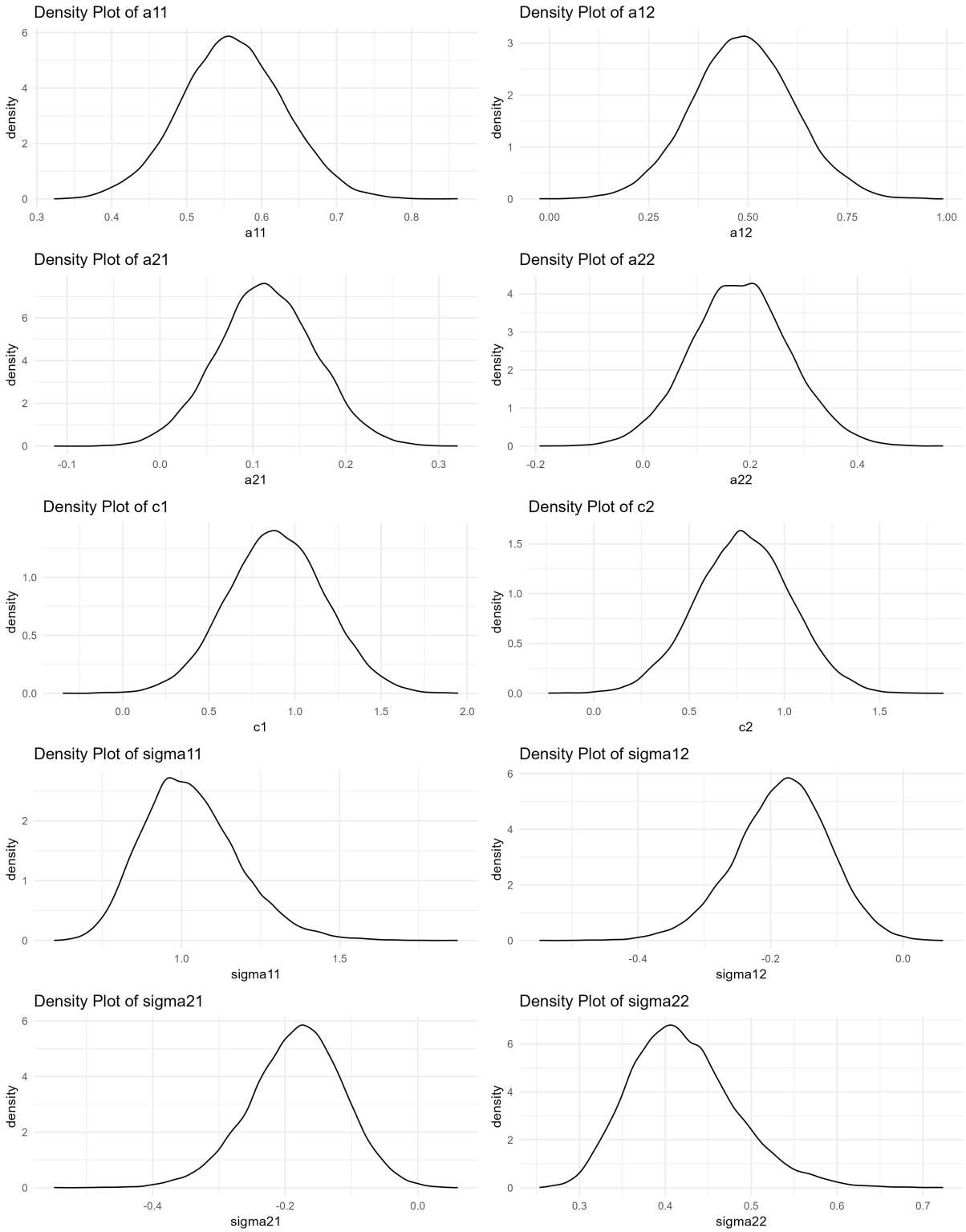

Figure 8: Density plots of the coefficients and covariance matrix for the VAR(1) model with the Independent Normal-Wishart prior

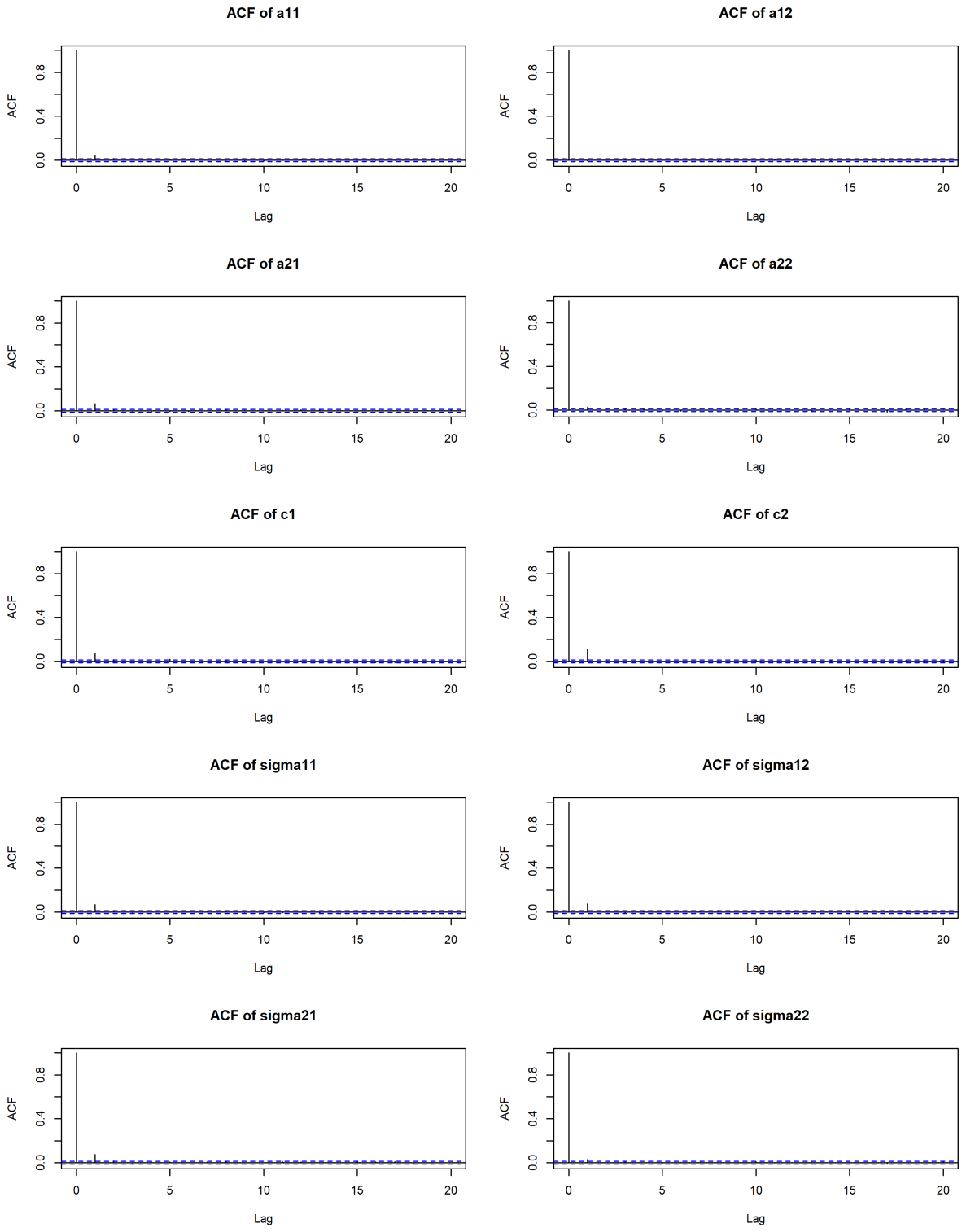

Figure 9: Auto-correlation plots of the coefficients and covariance matrix for the VAR(1) model with the Independent Normal-Wishart prior

8 Appendix B

Table 2: Posterior draws of the coefficients for model 1

	Mean	SD	2.5%	50%	97.5%	Monte Carlo error	R_hat
a11	0.204	0.108	-0.012	0.205	0.415	1e-05	1
a12	0.100	0.171	-0.235	0.100	0.438	1e-05	1
a21	0.008	0.072	-0.133	0.008	0.151	0e+00	1
a22	0.052	0.114	-0.169	0.051	0.281	1e-05	1
c1	2.956	0.575	1.827	2.954	4.098	3e-05	1
c2	1.416	0.387	0.650	1.419	2.167	2e-05	1
sigma11	0.933	0.139	0.698	0.920	1.240	1e-05	1
sigma12	-0.231	0.068	-0.377	-0.227	-0.109	0e+00	1
sigma21	-0.231	0.068	-0.377	-0.227	-0.109	0e+00	1
sigma22	0.416	0.062	0.312	0.410	0.553	0e+00	1

Table 3: Posterior draws of the coefficients for model 2

	Mean	SD	2.5%	50%	97.5%	Monte Carlo error	R_hat
a11	0.221	0.104	0.012	0.221	0.424	1e-05	1
a12	0.121	0.166	-0.202	0.120	0.448	1e-05	1
a21	0.016	0.070	-0.121	0.016	0.155	0e+00	1
a22	0.062	0.111	-0.153	0.062	0.282	1e-05	1
c1	2.856	0.552	1.766	2.855	3.952	3e-05	1
c2	1.367	0.373	0.629	1.369	2.093	2e-05	1
sigma11	0.895	0.129	0.675	0.883	1.180	1e-05	1
sigma12	-0.213	0.063	-0.348	-0.209	-0.098	0e+00	1
sigma21	-0.213	0.063	-0.348	-0.209	-0.098	0e+00	1
sigma22	0.403	0.058	0.305	0.397	0.531	0e+00	1

Table 4: Posterior draws of the coefficients for model 3

	Mean	SD	2.5%	50%	97.5%	Monte Carlo error	R_hat
a11	0.561	0.069	0.426	0.560	0.697	0e+00	1
a12	0.487	0.128	0.236	0.486	0.741	1e-05	1
a21	0.115	0.053	0.012	0.114	0.219	0e+00	1
a22	0.181	0.092	0.001	0.180	0.363	0e+00	1
c1	0.893	0.283	0.340	0.891	1.449	1e-05	1
c2	0.786	0.249	0.295	0.786	1.269	1e-05	1
sigma11	1.029	0.152	0.772	1.016	1.364	1e-05	1
sigma12	-0.183	0.071	-0.331	-0.179	-0.052	0e+00	1
sigma21	-0.183	0.071	-0.331	-0.179	-0.052	0e+00	1
sigma22	0.421	0.062	0.317	0.415	0.560	0e+00	1