ARL-TR-8741 ● July 2019

Bayesian Reduced-Rank Regression with Stan

by Benjamin T Files, Mac Strelioff, and Rasmus Bonnevie

NOTICES

Disclaimers

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.

ARL-TR-8741 ● JULY 2019

Bayesian Reduced-Rank Regression with Stan

by Benjamin T Files
Human Research and Engineering Directorate, CCDC Army Research Laboratory

Mac Strelioff
University of California, Irvine, Computational Cognitive Neuroscience

Rasmus Bonnevie
Technical University of Denmark, Cognitive Systems

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)* July 2019	**2. REPORT TYPE** Technical Report	**3. DATES COVERED (From - To)** 01 September 2017–30 August 2019

4. TITLE AND SUBTITLE Bayesian Reduced-Rank Regression with Stan	**5a. CONTRACT NUMBER** **5b. GRANT NUMBER** **5c. PROGRAM ELEMENT NUMBER**
6. AUTHOR(S) Benjamin T Files, Mac Strelioff,* and Rasmus Bonnevie	**5d. PROJECT NUMBER** **5e. TASK NUMBER** **5f. WORK UNIT NUMBER**
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) CCDC Army Research Laboratory ATTN: FCDD-RLH-FA 12025 Waterfront Drive Playa Vista, CA 90094	**8. PERFORMING ORGANIZATION REPORT NUMBER** ARL-TR-8741
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	**10. SPONSOR/MONITOR'S ACRONYM(S)** **11. SPONSOR/MONITOR'S REPORT NUMBER(S)**

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

ORCID ID(s): Benjamin Files, 0000-0002-1141-7886

*Currently at Insight Data Science

14. ABSTRACT

Reduced-rank regression enables characterizing the relationship between several predictors and outcome measures when their relationship can be accounted for with a relatively small number of latent dimensions. In contrast to full-rank multivariate regression, reduced-rank regression avoids estimating redundant regression coefficients and efficiently uncovers the underlying lower-dimensional latent variables that characterize the relationship between predictors and outcomes. Here, we report on an implementation of reduced-rank regression in a Bayesian framework using Markov Chain Monte Carlo, No-U-Turn Sampling as implemented in Stan, a popular open-source Bayesian inference engine. This implementation supports robust error modelling and calculation of posterior uncertainty intervals.

15. SUBJECT TERMS

rank reduction; multivariate regression; Bayes; Stan; simulation-based calibration

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** Benjamin T Files
a. REPORT Unclassified	**b. ABSTRACT** Unclassified	**c. THIS PAGE** Unclassified	UU	21	**19b. TELEPHONE NUMBER (Include area code)** (310) 577-6212

Contents

List of Figures

Acknowledgments

The authors wish to thank the Stan community for helpful advice and discussion. This work was supported by US Army Combat Capabilities Development Command Army Research Laboratory’s 6.1 Training Effectiveness project.

1. Introduction

Multivariate regression characterizes the relationship between multiple predictors and multiple outcome measures. Here, we adopt the standard formulation, $Y = X\beta + \epsilon$, with $Y \in \mathbb{R}^{n \times c}$, $X \in \mathbb{R}^{n \times p}$, coefficient matrix $\beta \in \mathbb{R}^{p \times c}$, and error term $\epsilon \in \mathbb{R}^{n \times c}$, with n observations, p covariates, and c outcome variables. We focus specifically on reduced-rank regression models, in which the coefficient matrix, β, is assumed to have rank $k < \min(p, c)$.

Identifying a lower-rank structure of the relationship between *X* and *Y* has at least two advantages. The first is interpretability: the simpler structure clarifies a small number of independent combinations of covariates that each has a relationship with a combination of outcomes. The second advantage is in data efficiency. All other things held equal, a model with fewer parameters can estimate those parameters with less uncertainty compared to a model with more parameters. So, with the same number of observations, uncertainty intervals will be smaller when the underlying low-rank structure of β is estimated.

To uncover the low-rank structure of the regression coefficient, it can be expressed as $\beta = BA^T$, with $B \in \mathbb{R}^{p \times k}$, and $A \in \mathbb{R}^{c \times k}$. The matrices *A* and *B* are not uniquely identifiable. Column permutations, sign reversals, and changes in scale can all produce identical β. Which pair of *A* and *B* to prefer differentiates a number of approaches to fitting the general reduced-rank regression model. Principal components regression, canonical correlation analysis, partial least squares regression, and reduced-rank regression are all algorithms to fit a reduced-rank regression model, and are related via the generalized eigenvalue problem (Borga et al. 1997).

Although these methods exist within the frequentist framework, past work has developed Bayesian versions as well. Bayesian reduced-rank regression methods have a number of advantages relative to their frequentist counterparts. Bayesian methods permit the construction of posterior uncertainty intervals around coefficients, can provide evidence in favor of null relationships, can generate posterior predictive distributions, and can represent detailed prior knowledge of parameter values via flexible prior distributions.

The present approach is similar to the first Bayesian treatment of reduced-rank regression (Geweke 1996); whereas here, we restrict *B* to be lower trapezoidal, the Geweke approach is to force a subset of *Y* to depend only on *A* or *B* by replacing the first k columns of *A* or *B* with an identity matrix (i.e., $B^T = [I_k | B^{*T}]$ or $A^T = [I_k | A^{*T}]$), a choice that can affect the model's suitability to a particular problem. Geweke introduces a Gibbs sampler to draw from the posteriors, but it is conditional on reference prior distributions, which might not be desirable (Gelman et al. 2017). Vidaurre and colleagues introduced a Bayesian sparse partial least squares that used variational Bayes (Vidaurre et al. 2013). Other recent work introduces a Bayesian method for sparse reduced-rank regression (Goh et al. 2017). It estimates the posterior distribution for the elements of β, which is sparse, without identifying posteriors for A or B. Although their method produces an analogue to B and A on each iteration, some post-hoc column, sign, and scale transformation would need to be applied to obtain a coherent posterior on B and A.

Here, we introduce a fully Bayesian modelling approach to reduced-rank regression that uses weakly informative priors, supports robust error modeling, and produces well-defined posterior distributions for analogues to B and A. It is implemented as a program for Stan (Carpenter et al. 2017), a popular Bayesian inference engine that, given a Bayesian model and some data, will automatically generate draws from the posterior distribution. To accomplish this, it uses a variant of Hamiltonian Monte Carlo sampling called the No-U-Turn Sampler. Stan's modeling language also facilitates simulation by generating simulated data from a known set of parameters drawn from their respective priors. In the following sections, we describe our modeling approach in detail and present results from simulation studies to confirm that the model is well-specified. We conclude with some recommendations for future work.

2. Methods, Assumptions, Procedures

2.1 Bayesian Reduced-Rank Regression Model

A straightforward approach to Bayesian reduced-rank regression would be to specify some priors for A and B, specify some error term, and let the sampler take it from there. This approach is problematic because of the sign, column, and scale ambiguities described previously. The same ambiguities occur in latent variable models, and here we borrow the solution from latent variable modeling (Geweke and Zhou 1996; Leung and Drton 2016). We first resolved the scale ambiguity by ensuring the priors on B and A strictly decrease from a mode at zero so that, all other things being equal, the coefficients with smaller values are preferred.

To resolve the sign and column ambiguities, we applied the LQ factorization to B, such that $BA^T = LQA^T$, where L is lower-trapezoidal with positive values on the diagonal and Q is unitary. This structure resolves the remaining ambiguities, as the lower-trapezoidal restriction precludes column permutation, and the positive diagonal restriction precludes sign changes. For convenience, we define $\hat{A}^T = QA^T$. Because Q is unitary, the distribution of $\hat{A}$ is the same as the distribution of

A subject to the criterion that the elements of A have a distribution that is symmetric around zero and that elements in a given row of A all share the same distribution.

Here, we use a spherical normal prior for the elements of $\hat{A}$ with mean zero and standard deviation one. This is appropriate when the columns of Y have been z-transformed. When such a transformation is undesirable, the rows of $\hat{A}$ could be assigned to be normal with some other standard deviation, i.e., $\hat{A}_{i,j} \sim \mathcal{N}(0, \sigma_i)$, read as the element in the ith row and the jth column is normally distributed with mean zero and standard deviation σ_i.

The prior for L when B is spherical normal with mean zero and standard deviation σ_B is described in the context of factor analysis by Leung and Drton (2016). For off-diagonal elements, the distribution in L is the same as that in B. The ith diagonal of L is distributed proportional to

$$B_{i,i}{}^{k-i} e^{-B_{i,i}{}^2/2\sigma_B}, B_{i,i} > 0.$$

Although this distribution is not implemented natively in Stan, Stan makes it easy to implement any distribution for which the log of the probability density function can be written.

The Leung and Drton prior on L applies to B with identically distributed elements. This restriction can be conveniently relaxed to situations where the rows of B are identically distributed, but the columns are not, i.e., $B_{i,j} \sim \mathcal{N}(0, \lambda_i)$, by reformulating $B = \Lambda\hat{B}$ where $\hat{B}$ is spherical standard normal and $diag(\Lambda) = \{\lambda_1, \lambda_2, \dots \lambda_p\}$. In the present implementation, the priors on the diagonal elements of Λ are standard half-normal, $\lambda_i \sim \mathcal{N}_+(0,1)$.

Finally, the elements of the error term are modeled as drawn independently from a generalized Student's T distribution with ν degrees of freedom, mean 0 and standard deviation σ. The hyper parameters ν and σ are given fixed prior distributions, with $\nu \sim Gamma(2, \frac{1}{10})$ and $\sigma \sim \mathcal{N}_+(0,1)$. Although these hyper-priors might be inappropriate in some situations, they are generally recommended as reasonable defaults (Stan Development Team 2017). The Stan program implementing this model is in Listing 1.

Listing 1: Stan program for reduced-rank regression

```
functions {
  real ld_diag_lpdf(real x, int i, int k, real c0) {
    return (k-i)*log(x) - square(x)/(2*c0);
  }
}
data {
  int n;
  int p;
  int c;
  int k;
  matrix[n, p] X;
  matrix[n, c] Y;
}
transformed data {
  int ntrap = p*k - k*(k-1)/2 - k;
}
parameters {
  real<lower=0> sigma;
  real<lower=0> nu;
  vector<lower=0>[k] diags;
  vector[ntrap] lowtrap;
  matrix[k, c] Ahat;
  vector<lower=0>[p] lambda;
}
transformed parameters {
  matrix[p,k] L;
  { //hide integer
    int idx;
    idx=0;
    L = rep_matrix(0, p, k);
    for (col in 1:k) {
      L[col, col] = diags[col];
      for (r in (col+1):p) {
        idx+=1;
        L[r, col] = lowtrap[idx];
      }
    }
  } //close hider
}
model {
  matrix[n, c] mu;
  lowtrap ~ normal(0,1);
  for (i in 1:k)
    diags[i] ~ ld_diag(i, k, 1);
  to_vector(Ahat) ~ normal(0,1);
  lambda ~ normal(0, 1);
  nu ~ gamma(2, 0.1);
  sigma ~ normal(0,1);
  mu = diag_post_multiply(X, lambda)*L*Ahat;

  to_vector(Y) ~ student_t(nu, to_vector(mu), sigma);
}
```

2.2 Model Evaluation

To validate the model implemented in Listing 1, we used simulation-based calibration (SBC; Talts et al. 2018). SBC is an update to a previous method (Cook et al. 2006); both follow similar logic based on the relationship between parameters drawn from the priors of a model and the posterior distributions for those parameters conditioned on data simulated by the generative model. The goal of SBC is to ensure the model is well-specified, meaning that the model is a correct and error-free representation of a specific data-generating process. The steps of SBC are as follows: First, draw a set of model parameters from the model's priors, $\theta^* \sim p(\theta)$. Next, simulate some data according to the generative model, $y^* \sim p(y|\theta^*)$. Now generate S independent draws from the posterior distribution, $\{\theta_1, \theta_2, \dots \theta_S\} \sim p(\theta|y^*)$. Finally, for each element of θ^*, θ_i^*, compute the rank statistic $R_i = \sum_{s=1}^{S} \mathbb{I}(\theta_{s,i} < \theta_i^*)$, where $\mathbb{I}(x)$ is the indicator function. Under a well-specified model, the rank statistic R will have a discrete uniform distribution over the integers $[0, S]$, inclusive (Talts et al. 2018, Theorem 1). To assess the uniformity of R_i, the SBC procedure is repeated many times, and a histogram of the obtained values of R_i is constructed. Under uniformity, after N repetitions of the SBC procedure, the count of occurrences of a given value of R_i will follow a binomial distribution, $Binomial(N, \frac{1}{S+1})$.

The above holds when independent draws from the posterior are available. The Stan program for our reduced-rank regression method produces correlated draws from the posterior, so the draws were thinned to obtain approximately independent draws. The goal for each simulation was to obtain 1024 nearly independent draws. Sixteen independent chains were run with 8,192 warmup and 8,192 post-warmup draws, and the draws were thinned by a factor of 128 (i.e., every 128th draw was saved and the rest were discarded), for a total of 1024 post-warmup draws. If, after this procedure, the estimated effective number of draws for any variable was less than 910, that simulation was not used. The first sample of retained simulations was discarded, setting S to 1023, so that $N \times \frac{1}{S+1}$ was a whole number.

This entire process of drawing parameter values from their respective priors, simulating data, and then fitting the model to the simulated data was repeated until $N = 1024$ to obtain an expected count per bin of 1. After pooling 64 adjacent bins, the expected count per pooled bin is 128. Uniformity of R was assessed graphically with a histogram and imposed lines representing the 1st and 99th percentiles of a binomial distribution with N = 1024 and p = 64/1024. These methods will be insensitive to small deviations from uniformity but are sufficient to see large deviations that would be expected in a miss-specified model.

Stan can operate in simulation mode to generate simulated data sets. Listing 2 shows the Stan program to simulate data based on our reduced-rank regression model. Sampling from the distribution specified in Leung and Drton (2016) for L is not implemented in Stan, so the approach instead was to draw from the implied priors on B and $A,$ decompose B into L and Q, and compose $\hat{A}^T$ from A and Q. This approach therefore relies on the assumption that the Leung and Drton prior is correct and well-implemented.

Listing 2: Stan program for simulating data from the reduced-rank regression model

```
data {
  int n;
  int p;
  int c;
  int k;
}
transformed data {
  int ntrap = p*k - k*(k-1)/2 - k;
}
parameters {
}
transformed parameters {
}
model {
}
generated quantities {
  // parameters
  real<lower=0> sigma;
  real<lower=0> nu;
  vector<lower=0>[p] lambda;
  matrix[p,k] B;
  matrix[c,k] A;
  // transformed parameters
  matrix[k, c] Ahat;
  matrix[p,k] L;
  matrix[p,p] B_fat;
  matrix[p,p] qt_fat;
  matrix[p,p] lt_fat;
  // model values
  matrix[n, p] X;
  matrix[n,c] Y;
  matrix[n,c] mu;
  // parameter samples
  sigma = fabs(normal_rng(0,1));
  nu = gamma_rng(2, 0.1);
  for (i in 1:p)
    lambda[i] = fabs(normal_rng(0, 1));
  for (j in 1:k){
    for (i in 1:p)
      B[i,j] = normal_rng(0,1);
    for (i in 1:c)
      A[i,j] = normal_rng(0,1);
  }
  // transformations
  B_fat = append_col(B, rep_matrix(rep_vector(0,p), p-k));
  qt_fat = qr_Q(B_fat');
  lt_fat = qr_R(B_fat');
  L = lt_fat[1:k,]';
  Ahat = qt_fat[1:k,1:k]' * A';
  // model code
  for (i in 1:n) {
    for (j in 1:p) {
      X[i,j] = normal_rng(0,1);
    }
  }
  mu = diag_post_multiply(X, lambda)*L*Ahat;
  for (i in 1:n) {
    for (j in 1:c) {
      Y[i,j] = student_t_rng(nu, mu[i,j], sigma);
    }
  }
}
```

When fitting this model to data (simulated or real), it is important to provide initial parameter values to the sampler that are reasonable and similar (but not identical) for each independent chain. R code for generating reasonable initialization values is shown in Listing 3. These initial estimates are then jittered and provided as initial values to Stan.

Listing 3: R code for initializing analyses

```
Dbeta = ginv(X)%*%Y # ML solution
Dhat=apply(Dbeta,MARGIN=1,FUN=sd) # extract SD
beta=Dbeta/Dhat

# Decompose Beta into L and Ahat
tmp=qr(t(beta))
iL=t(qr.R(tmp))
iL=iL[,1:settings$k]
iQ=t(qr.Q(tmp))
iQ=iQ[1:settings$k,]
iL = iL/sqrt(settings$k*settings$c)
iAhat = iQ*sqrt(settings$k*settings$c)
# flip iB so diags are positive
sgn=t(sign(iL[diag_idx]))
iL = iL%*%diag(c(sgn))
iAhat = diag(c(sgn))%*%iAhat
```

This code snippet uses only values known a priori to estimate good starting points for model fitting. In particular, it uses the predictors *X*, outcomes *Y*, requested number of latent dimensions, *k*, and number of outcomes, *c*. These estimates were then jittered so the different chains would not start in identical locations.

3. Results and Discussion

The simulation procedure was repeated 6,400 times, but only 1,258 simulations met our convergence and independence criteria. Of the 1,258 that met the criteria, 1,024 were randomly selected for analysis.

Histograms show the rank frequency for each element of Λ (Fig. 1), L (Fig. 2), $\hat{A}^T$ (Fig. 3), and error parameters ν and σ (Fig. 4). The histograms are generally consistent with a uniform distribution, as illustrated with a quantile-quantile plot of ranking bin counts versus the quantiles of the binomial distribution (Fig. 5). Over all the estimated parameters, 2 bin counts (of 592 bins) are below the first percentile, a proportion of 0.0034. Seven counts are above the 99th percentile, a proportion of 0.012. These summaries are consistent with a well-specified model.

Despite fairly extreme thinning and long warmup periods, many of our simulations did not meet our criteria. Those that did meet our criteria suggest the model is well-specified. However, that a large proportion of simulations did not meet our independence criteria suggests that even longer warmup periods may be necessary in practice. When independent draws are required, then even more extreme thinning might be necessary. For purposes other than SBC, such as fitting the model to nonsimulation data, independent draws are typically not required. The results of this calibration show that when the chains do converge and samples are approximately independent, the estimates do not show any substantial bias that would indicate a miss-specified model.

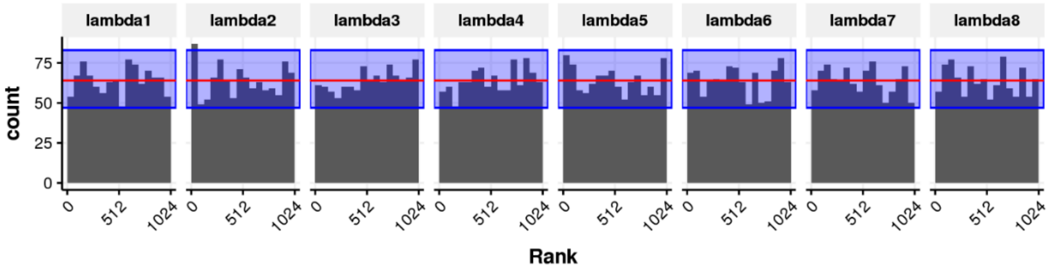

Fig. 1 Histograms of the rank of the simulated values of the eight elements of Λ relative to the draws from their respective posterior distributions in 1,024 simulations. The bin width is 64. The middle horizontal line indicates the expected bin count, with the box showing the 1st and 99th percentiles of a binomial distribution.

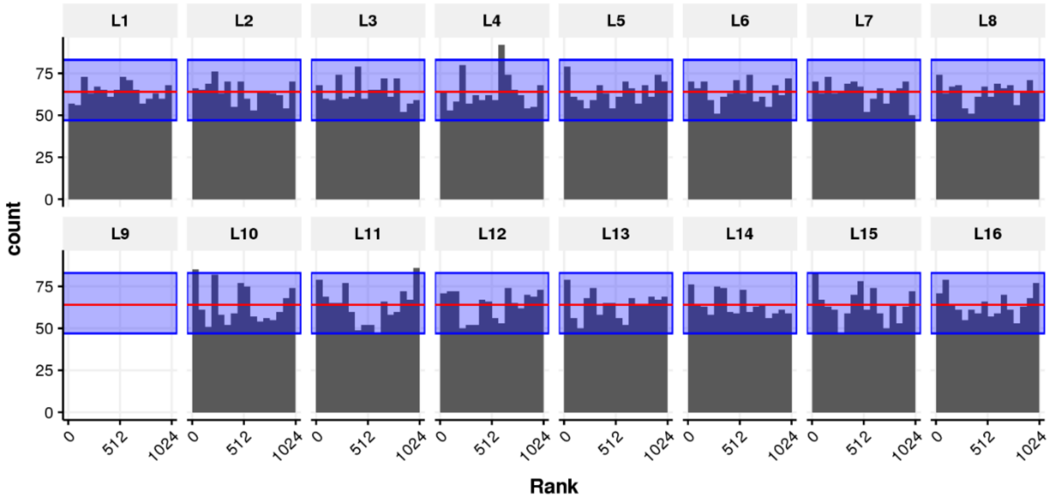

Fig. 2 Histograms of the rank of the simulated values of the 16 elements of L relative to the draws from their respective posterior distributions in 1,024 simulations. Note that the figure order is transposed, so L1–L8 are the elements of the first column of L. L9 was restricted to always be zero, so no data are shown for that element. The bin width is 64. The middle horizontal line indicates the expected bin count, with the box showing the 1st and 99th percentiles of a binomial distribution.

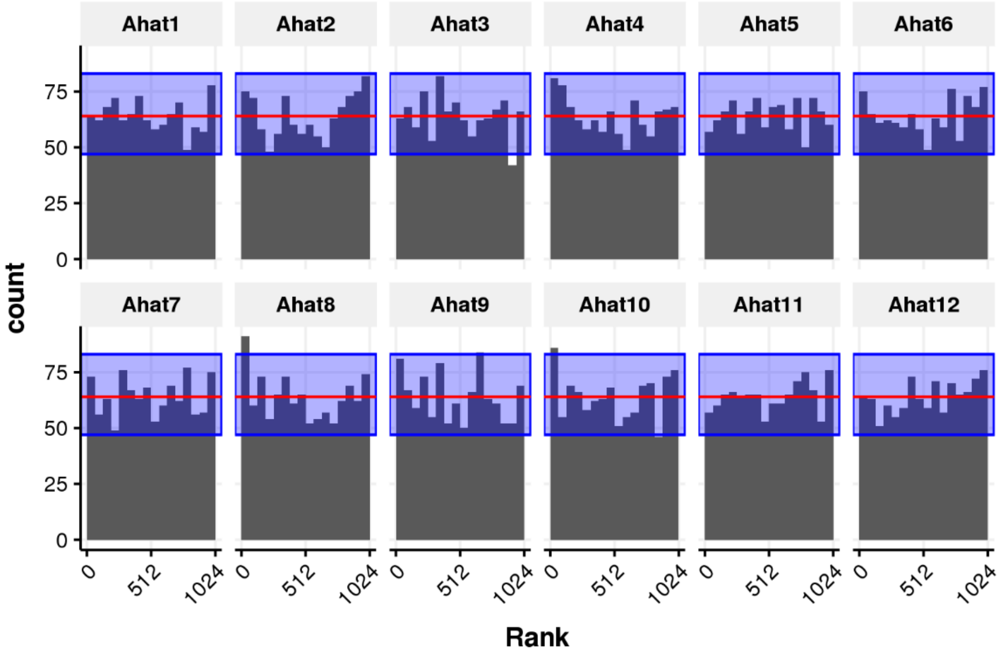

Fig. 3 Histograms of the rank of the simulated values of the 12 elements of $\widehat{A}^T$ relative to the draws from their respective posterior distributions in 1,024 simulations. The bin width is 64. The middle horizontal line indicates the expected bin count, with the box showing the 1st and 99th percentiles of a binomial distribution.

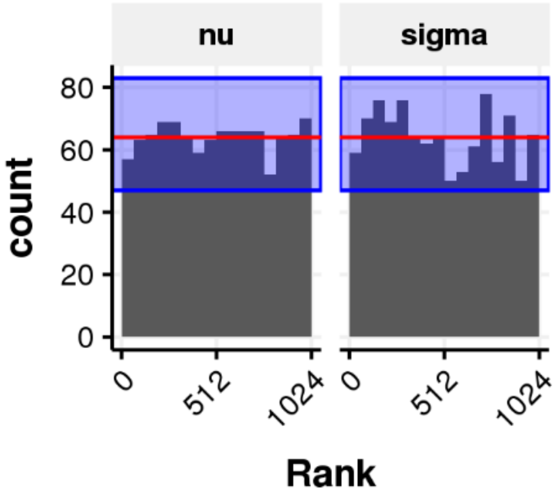

Fig. 4 Histograms of the rank of the simulated values of the parameters ν and σ relative to the draws from their respective posterior distributions in 1,024 simulations. The bin width is 64. The middle horizontal line indicates the expected bin count, with the box showing the 1st and 99th percentiles of a binomial distribution.

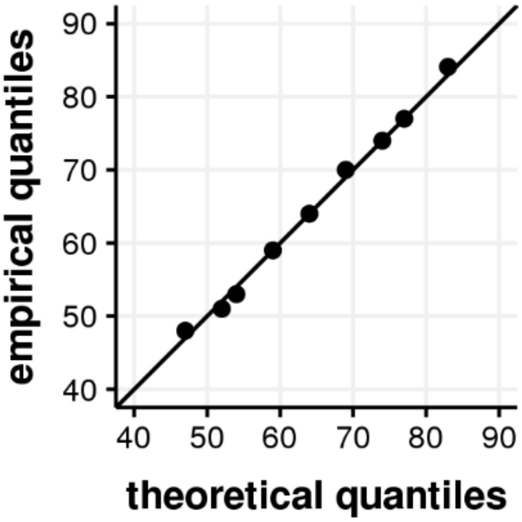

Fig. 5 Quantile-quantile plot of the observed rank histogram counts (empirical quantiles) against the quantiles of a binomial distribution. The quantiles selected were the 1, 5, 10, 25, 50, 75, 90, 95, and 99th percentile.

4. Conclusions

Here, we have provided code for a Stan program that can generate samples from the posterior of a reduced-rank regression model. As described previously, other Bayesian approaches to reduced-rank regression are available, but the present approach is relatively easy to understand and very easy to run. Interfaces to Stan are available for many popular programming languages, including Python, R, and

MATLAB, so the Stan model developed and validated here could easily be applied by other researchers. We have also demonstrated that the model is well-specified, although some tuning of the sampler and/or extreme thinning might be necessary to obtain independent draws from the posterior. We conclude with a few comments about the limitations and possible extensions to this Bayesian reduced-rank regression algorithm.

In this program, the number of latent dimensions, k, is specified ahead of time. When k is set to some value larger than is needed to account for the data, the posteriors of all the elements of some columns of A should be centered on zero. Since k can be thought of as specifying the number of columns of A with nonzero elements, this pattern of results can be interpreted roughly as the model inferring a value of k smaller than the chosen value. This may result in more severe shrinkage to the nonzero columns, but the model can then be rerun with a different choice of k. Formal model comparisons, for example using Bayes Factors (Kass and Raftery 1995) or approximate leave-one-out cross-validation (Vehtari et al. 2017), could be used to select the best value of k.

A further limitation of the present implementation is that it might be insensitive to situations with many covariates, especially when a relatively small number of covariates have any appreciable impact on the outcomes. Frequentist regularization techniques, such as ridge regression, find sparse solutions by penalizing the model likelihood by a term related to the magnitude of the estimated coefficients. In the Bayesian framework, a number of approaches focused on the choice for the prior distribution are proposed to obtain sparse solutions (Gelman et al. 2013; Betancourt 2018). One particularly well-suited approach is the regularized horseshoe prior (Piironen and Vehtari 2017), sometimes also called the *Finnish horseshoe*, which uses the following prior for regression coefficients: $b_i \sim \mathcal{N}(0, \lambda_i \tau)$, where λ_i is a per-coefficient shrinkage parameter, and τ is a global shrinkage parameter. With appropriately selected hyperpriors, this has the effect of forcing most coefficients to be close to zero but leaving others free to take on relatively large values.

Using a regularized horseshoe prior on Λ in the present model would enforce row-sparsity on β, meaning that all the entries in some rows would be shrunk toward zero, leaving other whole rows free to be relatively large. In some cases, this might be desirable, but in general, a goal would be to obtain a solution that was sparse throughout. Applying such a prior would alter the structure of L and $\hat{A}^T$, rendering priors derived from the assumptions of identically distributed rows inappropriate, so appropriate alternative priors would have to be identified. An exception is when $k = 1$, then the regularized horseshoe prior can be used in the present model without apparent problems. Work to apply the regularized horseshoe to our reduced-rank regression method under less restrictive conditions is ongoing.

5. References

Betancourt M. Bayes sparse regression. 2018 Mar [accessed 2018 Sep 13]. https://betanalpha.github.io/assets/case_studies/bayes_sparse_regression.html

Borga M, Landelius T, Knutsson H. A unified approach to pca, pls, mlr and cca. Linköping University, Department of Electrical Engineering; 1997. Report No.: LiTH-ISY-R-1992.

Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, Riddell A. Stan : a probabilistic programming language. J Stat Softw. 2017;76(1). https://doi.org/10.18637/jss.v076.i01.

Cook SR, Gelman A, Rubin DB. Validation of software for Bayesian models using posterior quantiles. J Comput Graph Stat. 2006;15(3);675–692. https://doi.org/10.1198/106186006X136976.

Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian data analysis. 3rd ed. Boca Raton (FL): Chapman and Hall/CRC; 2013.

Gelman A, Simpson D, Betancourt M. The prior can generally only be understood in the context of the likelihood. Entropy. 2017;19(10):555. https://doi.org/10.3390/e19100555.

Geweke J. Bayesian reduced rank regression in econometrics. J Econometrics. 1996;75(1):121–146. https://doi.org/10.1016/0304-4076(95)01773-9.

Geweke J, Zhou G. Measuring the pricing error of the arbitrage pricing theory. Rev Financ Stud. 1996;9(2):557–587. https://doi.org/10.1093/rfs/9.2.557.

Goh G, Dey DK, Chen K. Bayesian sparse reduced rank multivariate regression. J Multivariate Anal. 2017;157:14–28. https://doi.org/10.1016/j.jmva.2017.02.007.

Kass RE, Raftery AE. Bayes factors. J Am Stat Assoc. 1995;90(430):773–795. https://doi.org/10.2307/2291091.

Leung D, Drton M. Order-invariant prior specification in Bayesian factor analysis. Stat Probabil Lett. 2016;111:60–66. https://doi.org/10.1016/j.spl.2016.01.006.

Piironen J, Vehtari A. Sparsity information and regularization in the horseshoe and other shrinkage priors. Electron J Stat. 2017;11(2):5018–5051. https://doi.org/10.1214/17-EJS1337SI.

Stan Development Team. Stan modeling language user’s guide and reference manual. Ver. 2.17.0. 2017 [accessed 2019 Jun 27]. http://mc-stan.org.

Talts S, Betancourt M, Simpson D, Vehtari A, Gelman A. Validating Bayesian inference algorithms with simulation-based calibration. 2018 Apr 18 [accessed 2019 Jun 19]. ArXiv:1804.06788 [Stat]. http://arxiv.org/abs/1804.06788.

Vehtari A, Gelman A, Gabry J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Stat Comput. 2017;27(5):1413–1432. https://doi.org/10.1007/s11222-016-9696-4.

Vidaurre D, van Gerven MA, Bielza C, Larrañaga P, Heskes T. Bayesian sparse partial least squares. Neural Comput. 2013;25(12):3318–3339.

1 DEFENSE TECHNICAL
(PDF) INFORMATION CTR
DTIC OCA

1 CCDC ARL
(PDF) FCDD RLD CL
TECH LIB

1 GOVT PRINTG OFC
(PDF) A MALHOTRA

1 CCDC ARL
(PDF) FCDD RLH B
T DAVIS
BLDG 5400 RM C242
REDSTONE ARSENAL AL
35898-7290

1 USA ARMY G1
(PDF) DAPE HSI M SAMS
300 ARMY PENTAGON
RM 2C489
WASHINGTON DC 20310-0300

1 USAF 711 HPW
(PDF) 711 HPW/RH K GEISS
2698 G ST BLDG 190
WRIGHT PATTERSON AFB OH
45433-7604

1 USN ONR
(PDF) ONR CODE 341 J TANGNEY
875 N RANDOLPH STREET
BLDG 87
ARLINGTON VA 22203-1986

1 USA NSRDEC
(PDF) RDNS D D TAMILIO
10 GENERAL GREENE AVE
NATICK MA 01760-2642

1 OSD OUSD ATL
(PDF) HPT&B B PETRO
4800 MARK CENTER DRIVE
SUITE 17E08
ALEXANDRIA VA 22350

<u>ABERDEEN PROVING GROUND</u>

11 CCDC ARL
(PDF) FCDD RLH
J LANE
Y CHEN
P FRANASZCZUK
K MCDOWELL
K OIE
FCDD RLH BD
D HEADLEY
FCDD RLH FA
B FILES
A DECOSTANZA
FCDD RLH FB
A EVANS
FCDD RLH FC
J GASTON
FCDD RLH FD
A MARATHE