

# Modified Delayed Acceptance MCMC for Quasi-Bayesian Inference with Linear Moment Conditions

Masahiro Tanaka  
Faculty of Economics  
Fukuoka University  
Fukuoka, Japan  
gspddnit45@toki.waseda.jp

**Abstract**—We develop a computationally efficient framework for quasi-Bayesian inference based on linear moment conditions. The approach employs a delayed acceptance Markov chain Monte Carlo (DA-MCMC) algorithm that uses a surrogate target kernel and a proposal distribution derived from an approximate conditional posterior, thereby exploiting the structure of the quasi-likelihood. Two implementations are introduced. DA-MCMC-Exact fully incorporates prior information into the proposal distribution and maximizes per-iteration efficiency, whereas DA-MCMC-Approx omits the prior in the proposal to reduce matrix inversions, improving numerical stability and computational speed in higher dimensions. Simulation studies on heteroskedastic linear regressions show substantial gains over standard MCMC and conventional DA-MCMC baselines, measured by multivariate effective sample size per iteration and per second. The Approx variant yields the best overall throughput, while the Exact variant attains the highest per-iteration efficiency. Applications to two empirical instrumental variable regressions corroborate these findings: the Approx implementation scales to larger designs where other methods become impractical, while still delivering precise inference. Although developed for moment-based quasi-posteriors, the proposed approach also extends to risk-based quasi-Bayesian formulations when first-order conditions are linear and can be transformed analogously. Overall, the proposed algorithms provide a practical and robust tool for quasi-Bayesian analysis in statistical applications.

**Index Terms**—quasi-Bayesian inference, delayed acceptance Markov chain Monte Carlo, generalized method of moments

## I. INTRODUCTION

A well-known limitation of conventional Bayesian analysis is its reliance on the full specification of a probabilistic model. While this feature provides a coherent inferential framework, it also makes Bayesian inference vulnerable to model misspecification. To mitigate this issue, a growing body of research has explored quasi-Bayesian approaches that relax the requirement for an exact likelihood specification. These approaches construct alternative quasi-likelihoods based on loss functions [1]–[5] or moment conditions [6]–[8], thereby enhancing robustness while preserving the interpretability of Bayesian posterior inference.

The present study adopts a quasi-likelihood formulation derived from the generalized method of moments (GMM)

criterion [9], [10]. Within the GMM framework, the statistical model is defined through a set of moment conditions rather than an explicit likelihood function. This formulation allows inference to proceed without stringent distributional assumptions such as error normality or functional form restrictions, while still enabling probabilistic interpretation through the quasi-posterior. Posterior inference can then be carried out using simulation-based techniques such as Markov chain Monte Carlo (MCMC) [6]–[8], [11].

However, quasi-Bayesian inference based on moment conditions presents substantial computational challenges. Evaluating the quasi-posterior typically involves repeated matrix inversions and determinant calculations, which are both computationally intensive and prone to numerical instability—especially in high-dimensional settings. Existing work, such as the sampler proposed by [12], sought to improve numerical stability but often at significant computational cost. More recently, [13] introduced a delayed acceptance MCMC (DA-MCMC) algorithm [14] specifically designed for moment-based quasi-Bayesian inference. This approach demonstrated clear efficiency gains over conventional quasi-posterior simulation methods, establishing DA-MCMC as a promising direction for computation in moment-based quasi-Bayesian inference. Nonetheless, the improvement remains modest, and the fundamental computational difficulty persists, particularly when dealing with high-dimensional or weakly identified models.

Building on these developments, this paper proposes an efficient framework for quasi-Bayesian inference using a modified DA-MCMC algorithm tailored to quasi-posteriors derived from linear moment conditions. The proposed method leverages the linear structure of the moment equations to design a computationally tractable proposal distribution that closely approximates the target posterior. Two implementations of the algorithm are introduced, balancing the trade-off between computational efficiency and numerical stability. Through simulation experiments and empirical applications, we demonstrate that the proposed framework achieves substantial improvements in both sampling efficiency and computational speed.

More broadly, the proposed framework also relates to recent quasi-Bayesian approaches based on risk functions, such as that of [15]. Although their quasi-posterior formulation arises from a general decision-theoretic framework rather than a set of moment conditions, the method proposed in this paper can be applied to such models as long as the first-order condition of the risk function is linear in the parameters and can be transformed analogously to the moment condition structure considered here. This connection highlights the broader applicability of the proposed algorithm beyond the GMM-based context.

The remainder of the paper is organized as follows. Section II introduces the quasi-Bayesian framework based on moment conditions and develops the proposed DA-MCMC algorithms. Section III presents simulation studies using synthetic data to evaluate the computational performance of the proposed methods. Section IV applies the framework to real-world datasets to demonstrate its empirical relevance. Section V concludes with a summary of the findings and discusses directions for future research.

## II. METHOD

### A. Quasi-posterior

A statistical model is inferred from the moment condition  $\mathbb{E}[\mathbf{m}_i(\boldsymbol{\theta})] = \mathbf{0}_k$ , where  $\boldsymbol{\theta}$  denotes a  $k$ -dimensional vector of unknown parameters,  $\mathbf{m}_i(\cdot)$  is a vector-valued function referred to as the moment function,  $\mathbf{0}_a$  denotes an  $a$ -dimensional zero vector. We assume that  $\mathbf{m}_i(\cdot)$  has dimension  $k$ —that is, the model is exactly identified—and that  $\mathbf{m}_i(\cdot)$  is linear in  $\boldsymbol{\theta}$ .

The proposed framework encompasses a wide range of statistical models. For instance, a standard linear regression can be written as

$$y_i = \boldsymbol{\theta}^\top \mathbf{x}_i + u_i, \quad (1)$$

where  $y_i$  denotes an outcome variable,  $\mathbf{x}_i$  is a  $k$ -dimensional vector of covariates,  $\boldsymbol{\theta}$  represents the corresponding coefficients, and  $u_i$  denotes an error term. The model is estimated based on the following moment condition:

$$\mathbb{E}[(y_i - \boldsymbol{\theta}^\top \mathbf{x}_i) \mathbf{x}_i] = \mathbf{0}_k.$$

Under this approach, the distribution of the error terms is not assumed, making it more robust to model misspecification than the standard Bayesian approach. A linear probability model for a binary outcome can be formulated analogously. Multivariate regression models, such as seemingly unrelated regression [16], [17] and the local projection model [18], [19], can also be treated in a similar manner<sup>1</sup>.

Another important class of models is the instrumental variable (IV) regression model [21], [22], which consists of two equations:

$$\begin{aligned} x_i^{\text{treat}} &= g(x_i^{\text{instr}}, \tilde{\mathbf{x}}_i) + v_i, \\ y_i &= \theta^{\text{treat}} x_i^{\text{treat}} + \tilde{\boldsymbol{\theta}}^\top \tilde{\mathbf{x}}_i + u_i, \end{aligned} \quad (2)$$

<sup>1</sup>Although a local projection model is designed for time series data, it can be regarded as a model for independent observations as long as a sufficient number of lagged responses are included [20].

where  $y_i$  denotes an outcome variable,  $x_i^{\text{treat}}$  represents a treatment variable,  $x_i^{\text{instr}}$  denotes an instrumental variable that is correlated with  $x_i^{\text{treat}}$  but affects  $y_i$  only through  $x_i^{\text{treat}}$ ,  $\tilde{\mathbf{x}}_i$  is a vector of covariates,  $g(\cdot)$  denotes an unknown function,  $\theta^{\text{treat}}$  and  $\tilde{\boldsymbol{\theta}}$  are unknown parameters, and  $v_i$  and  $u_i$  are error terms. Define

$$\mathbf{x}_i = \begin{pmatrix} x_i^{\text{treat}} \\ \tilde{\mathbf{x}}_i \end{pmatrix}, \quad \mathbf{z}_i = \begin{pmatrix} x_i^{\text{instr}} \\ \tilde{\mathbf{x}}_i \end{pmatrix}, \quad \boldsymbol{\theta} = \begin{pmatrix} \theta^{\text{treat}} \\ \tilde{\boldsymbol{\theta}} \end{pmatrix}.$$

Then, the model can be estimated using the moment condition

$$\mathbb{E}[(y_i - \boldsymbol{\theta}^\top \mathbf{x}_i) \mathbf{z}_i] = \mathbf{0}_k.$$

This approach offers several advantages over conventional Bayesian instrumental variable regression methods. It eliminates the need to estimate the first-stage regression function and imposes no assumptions on the distributional form of the error terms. Some measurement error (error-in-variables) models [23] can also be formulated in a similar manner.

We construct the quasi-likelihood based on the generalized method of moments criterion [9], [10]. Given a prior distribution  $p(\boldsymbol{\theta})$ , the quasi-posterior kernel—that is, the quasi-posterior density evaluated up to the normalizing constant—is specified as

$$\pi(\boldsymbol{\theta}) = |\mathbf{W}|^{\frac{1}{2}} \exp\left(-\frac{n}{2} \bar{\mathbf{m}}(\boldsymbol{\theta})^\top \mathbf{W} \bar{\mathbf{m}}(\boldsymbol{\theta})\right) p(\boldsymbol{\theta}),$$

where

$$\bar{\mathbf{m}}(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n \mathbf{m}_i(\boldsymbol{\theta}),$$

denotes the empirical mean of the moment function  $\mathbf{m}_i(\boldsymbol{\theta})$  and  $\mathbf{W}$  is a symmetric positive-definite weighting matrix.

The weighting matrix  $\mathbf{W}$  is specified as the inverse of the empirical covariance matrix of  $\mathbf{m}_i(\boldsymbol{\theta})$ :

$$\mathbf{W} = \mathbf{V}^{-1},$$

$$\mathbf{V} = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{m}_i(\boldsymbol{\theta}) - \bar{\mathbf{m}}(\boldsymbol{\theta})) (\mathbf{m}_i(\boldsymbol{\theta}) - \bar{\mathbf{m}}(\boldsymbol{\theta}))^\top.$$

This choice is desirable because the resulting estimator yields the smallest credible set, and the credible set asymptotically attains its nominal coverage—for example, a 90% credible interval is expected to contain the true parameter value with 90% probability as the sample size tends to infinity [15].

The primary challenge of this inferential approach lies in its computational cost and numerical stability. When the weighting matrix  $\mathbf{W}$  is treated as a function of  $\boldsymbol{\theta}$ , the quasi-posterior is expressed as

$$\begin{aligned} \pi(\boldsymbol{\theta}) &= |\mathbf{W}(\boldsymbol{\theta})|^{\frac{1}{2}} \times \\ &\exp\left(-\frac{n}{2} \bar{\mathbf{m}}(\boldsymbol{\theta})^\top \mathbf{W}(\boldsymbol{\theta}) \bar{\mathbf{m}}(\boldsymbol{\theta})\right) p(\boldsymbol{\theta}). \end{aligned} \quad (3)$$

Posterior simulation from this target kernel is known to be computationally inefficient and numerically unstable [24].

### B. DA-MCMC

We propose an algorithm based on the DA-MCMC method [14]. DA-MCMC is a variant of the Metropolis–Hastings algorithm that incorporates a screening process. In the first stage, given the current state  $\theta^{(t)}$ , a proposal for the new state  $\theta'$  is drawn from a proposal distribution  $q_1(\cdot|\theta^{(t)})$ . The proposal  $\theta'$  is then evaluated based on a surrogate kernel, denoted by  $\pi^*(\cdot)$ , which serves as a computationally inexpensive approximation to the target kernel  $\pi(\cdot)$ . The proposal is accepted with probability

$$\alpha_1(\theta^{(t)}, \theta') = \min \left\{ 1, \frac{q_1(\theta^{(t)}|\theta') \pi^*(\theta')}{q_1(\theta'|\theta^{(t)}) \pi^*(\theta^{(t)})} \right\}.$$

If  $\theta'$  is rejected, the current state is retained, i.e.,  $\theta^{(t+1)} = \theta^{(t)}$ . Otherwise, if accepted,  $\theta'$  advances to the second stage, where it is accepted with probability

$$\alpha_2(\theta^{(t)}, \theta') = \min \left\{ 1, \frac{q_2(\theta^{(t)}|\theta') \pi^*(\theta^{(t)})}{q_2(\theta'|\theta^{(t)}) \pi^*(\theta')} \right\},$$

where the effective second-stage proposal distribution is defined as

$$q_2(\theta'|\theta^{(t)}) = \alpha_1(\theta^{(t)}, \theta') q_1(\theta'|\theta^{(t)}).$$

In particular, [13] specifies the surrogate kernel by replacing  $W(\theta')$  with  $W(\theta^{(t)})$ , yielding

$$\pi^*(\theta') = \left| W(\theta^{(t)}) \right|^{-\frac{1}{2}} \times \exp \left( -\frac{n}{2} \bar{m}(\theta')^\top W(\theta^{(t)}) \bar{m}(\theta') \right) p(\theta').$$

As noted by [14], each iteration of DA-MCMC tends to be slightly less efficient than standard MCMC when measured by effective sample size per iteration. Nonetheless, DA-MCMC can achieve higher overall efficiency in terms of effective sample size per unit time, as it avoids the main computational bottleneck—the repeated evaluation of  $W^{-1}$  and  $|W|$ .

Previous studies [13], [14] employ a multivariate normal proposal distribution that is independent of the target kernel, which makes the DA-MCMC algorithm a close variant of the random-walk Metropolis–Hastings algorithm. While this choice makes DA-MCMC broadly applicable, it is computationally inefficient.

### C. Modified DA-MCMC

To address this limitation, the present study replaces the generic multivariate normal proposal with an approximate conditional posterior distribution that leverages the linear structure of the target kernel. To fix the idea, we focus on the linear regression (1). Define

$$\mathbf{y} = (y_1, \dots, y_n)^\top, \quad \mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^\top.$$

Then, the core component of the quasi-posterior (3) can be expressed as

$$\begin{aligned} & \exp \left( -\frac{n}{2} \bar{m}(\theta)^\top W(\theta) \bar{m}(\theta) \right) \\ &= \exp \left( -\frac{n}{2} \left[ \frac{1}{n} \mathbf{X}^\top (\mathbf{y} - \mathbf{X}\theta) \right]^\top W \left[ \frac{1}{n} \mathbf{X}^\top (\mathbf{y} - \mathbf{X}\theta) \right] \right) \\ &\propto \exp \left( -\frac{n}{2} (\theta - \hat{\theta}^\dagger)^\top G^\top W G (\theta - \hat{\theta}^\dagger) \right), \end{aligned}$$

where  $\hat{\theta}^\dagger = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$  denotes the ordinary least squares estimator for  $\theta$  and  $G = n^{-1} \mathbf{X}^\top \mathbf{X}$ . Assume that the (conditional) prior density is specified as

$$p(\theta) \propto \exp \left( -\frac{1}{2} \theta^\top Q \theta \right),$$

where  $Q$  is a symmetric matrix parameterized by a set of hyperparameters  $\tau$ . The (conditional) posterior density of  $\theta$  is then expressed as  $N(\theta|\Omega \Upsilon \hat{\theta}^\dagger, \Omega)$ , where  $\Omega = (\Upsilon + Q)^{-1}$ ,  $\Upsilon = nG^\top W G$ , and  $N(a|b, C)$  represents the probability density function of a multivariate normal distribution with mean  $b$  and covariance matrix  $C$ , evaluated at  $a$ . Using this (conditional) posterior distribution, we specify the first-stage proposal distribution as

$$q_1(\theta'|\theta^{(t)}) = N(\theta'|\Omega \Upsilon \hat{\theta}^\dagger, \Omega).$$

The direct implementation, hereafter referred to as DA-MCMC-Exact, requires repeated matrix inversions,  $\Omega = (\Upsilon + Q)^{-1}$ . The term *Exact* highlights that this version fully incorporates prior information into the proposal distribution. By doing so, this approach can make the posterior simulation highly efficient.

However, this approach suffers not only from high computational cost but also from numerical instability. Ensuring that  $\Omega$  remains invertible across the sampling space of  $(\theta, \tau)$  is often difficult. In some cases,  $Q$  acts as a regularizer for  $\Upsilon$ , as in a Tikhonov inverse, which stabilizes the simulation. In other cases, however, variability in  $Q$  introduces numerical instability—particularly when employing shrinkage priors [25]–[27] that induce substantial fluctuations in  $\tau$ .

To address this problem, we introduce an alternative implementation. This variant excludes prior information from the proposal distribution and sets  $\Omega = \Upsilon^{-1}$ , which leads to

$$q_1(\theta'|\theta^{(t)}) = N(\theta'|\hat{\theta}^\dagger, \Upsilon^{-1}).$$

This alternative implementation is hereafter referred to as DA-MCMC-Approx. The term *Approx* indicates that this version approximates the conditional posterior by omitting prior information from the proposal distribution.

The comparative performance of DA-MCMC-Exact and DA-MCMC-Approx depends on the task at hand, particularly on the relative contribution of the quasi-likelihood and the prior to the quasi-posterior. The Approx version performs well when the quasi-likelihood dominates the quasi-posterior or

when the prior is sufficiently non-informative. In contrast, its performance may deteriorate when the quasi-likelihood provides limited information, such as in small-sample settings, or when the prior exerts a strong influence on the quasi-posterior.

### III. APPLICATION TO SYNTHETIC DATA

We applied the proposed approach to infer a heteroskedastic linear regression model using synthetic data under various scenarios. This application compared the two implementations of the proposed approach with two existing benchmark methods. The first benchmark was the adaptive random-walk Metropolis–Hastings algorithm, specifically the version of [28]. The second benchmark was the DA–MCMC algorithm of [13], in which the proposal distribution follows a multivariate normal distribution whose covariance matrix is chosen adaptively. For both benchmark methods, the tuning parameters—namely target acceptance rate and learning rate—were set to the same values as those employed by [28].

The data-generating process for the synthetic data was inspired by [15]. Observations were generated from a normal distribution with non-constant variance,  $y_i \sim \mathcal{N}(\boldsymbol{\theta}^\top \mathbf{x}_i, \sigma_i^2)$ . Each covariate vector  $\mathbf{x}_i$  consisted of a constant term and exogenous random variables,

$$\mathbf{x}_i = \left(1, \tilde{\mathbf{x}}_i^\top\right)^\top, \quad \tilde{\mathbf{x}}_i \sim \mathcal{N}(\mathbf{0}_{k-1}, \mathbf{S}).$$

The covariance matrix  $\mathbf{S}$  was constructed as follows. First, a symmetric positive definite matrix was drawn from an inverse Wishart distribution with identity scale matrix and  $k + 1$  degrees of freedom,  $\mathbf{S} \sim \mathcal{IW}(\mathbf{I}_{k-1}, k + 1)$ . It was normalized to obtain a correlation matrix:

$$\mathbf{S} \leftarrow \tilde{\mathbf{S}}\tilde{\mathbf{S}}^\top, \quad \tilde{\mathbf{S}} = \text{diag}\left(s_{1,1}^{-0.5}, \dots, s_{k-1,k-1}^{-0.5}\right),$$

where  $s_{j,j}$  denotes the  $j$ th diagonal entry of  $\mathbf{S}$ . The coefficient vector was specified as  $\boldsymbol{\theta} = (1, 1, 1, 0, \dots, 0)^\top$ . The variance of the error terms depended on a subset of covariates, defined as

$$\sigma_i^2 = (1 + x_{i,2}^2 + x_{i,3}^2) / 3.$$

Three prior specifications were examined. The first specification, referred to as the Normal prior, assumes a normal prior with a constant unit variance,  $\boldsymbol{\theta} \sim \mathcal{N}(\mathbf{0}_k, \mathbf{I}_k)$ . The second specification, termed the NIG-homo prior, adopts a normal-inverse-gamma prior with a single common hyperparameter:

$$\boldsymbol{\theta} | \tau \sim \mathcal{N}(\mathbf{0}_k, \tau \mathbf{I}_k), \quad \tau \sim \mathcal{IG}(\nu_1, \nu_2),$$

where  $\nu_1$  and  $\nu_2$  are fixed hyperparameters and  $\mathcal{IG}(a, b)$  denotes an inverse gamma distribution with shape parameter  $a$  and rate parameter  $b$ . The third specification, referred to as the NIG-hetero prior, is a normal-inverse-gamma prior with non-common element-specific hyperparameters:

$$\theta_j | \tau_j \sim \mathcal{N}(0, \tau_j), \quad \tau_j \sim \mathcal{IG}(\nu_1, \nu_2).$$

For both the NIG-homo and NIG-hetero priors, the hyperparameters were set to  $\nu_1 = 2$  and  $\nu_2 = 1$ , making the priors

TABLE I  
RESULTS FOR SYNTHETIC DATA (1) NORMAL PRIOR

$n$	$k$	MCMC	DA-MCMC	Mod. DA-MCMC	
				Exact	Approx
(a) mESS/iter					
100	5	0.040	0.060	0.848	0.513
	20	0.004	0.010	0.412	0.090
1,000	5	0.033	0.059	0.985	0.814
	20	0.013	0.010	0.951	0.728
(b) mESS/s					
100	5	4,405	8,337	52,321	38,675
	20	139	642	9,076	4,389
1,000	5	1,644	3,999	34,736	30,800
	20	133	164	7,093	6,334

moderately informative. The hyperparameters were updated via a Gibbs sampling step.

We considered combinations of different sample sizes and numbers of covariates:  $n \in \{100, 1000\}$ ,  $k \in \{5, 20\}$ . For each experiment, a total of 200,000 draws were generated and the last 100,000 draws were retained for analysis. Performance was evaluated based on the multivariate effective sample sizes (mESS) [29]. Specifically, we computed the median values of mESS per iteration (mESS/iter) and mESS per second (mESS/s) across 100 independent runs.<sup>2</sup>

Table I summarizes the results for the Normal prior. Both DA–MCMC–Exact and DA–MCMC–Approx substantially outperformed the benchmark algorithms in terms of mESS per iteration and mESS per second across all experimental settings. The performance gap widened as the dimension  $k$  increased, indicating that the proposed methods scaled more efficiently in higher-dimensional problems. Between the two proposed implementations, DA–MCMC–Exact achieved higher mESS per iteration, reflecting its closer alignment with the true quasi-posterior distribution. In addition, DA–MCMC–Exact consistently demonstrated superior computational efficiency, as indicated by higher mESS per second. The differences between the two implementations were more pronounced for smaller samples ( $n = 100$ ) and larger dimensions ( $k = 20$ ), where the Exact version’s computational burden became more evident. Overall, DA–MCMC–Exact outperformed DA–MCMC–Approx on both the performance measures when the Normal prior was used.

Table II presents the results for the NIG-homo prior. Compared with the Normal prior case, the overall sampling efficiency declined slightly, reflecting the additional uncertainty introduced by the hyperparameter  $\tau$ . Nonetheless, both DA–MCMC–Exact and DA–MCMC–Approx continued to outperform the benchmark algorithms by substantial margins across all settings. For small samples ( $n = 100$ ) and low dimensionality ( $k = 5$ ), both implementations achieved multivariate effective sample sizes per iteration (mESS/iter) several times higher than those of the baseline methods. However, as dimensionality increased ( $k = 20$ ), efficiency gains diminished, and

<sup>2</sup>All the programs were executed in Matlab (R2025b) on an Ubuntu desktop (24.04.3 LTS) running on an AMD Ryzen Threadripper 3990X (2.9 GHz).

TABLE II  
RESULTS FOR SYNTHETIC DATA (2) NIG-HOMO PRIOR

$n$	$k$	MCMC	DA-MCMC	Mod. DA-MCMC	
				Exact	Approx
(a) mESS/iter					
100	5	0.040	0.059	0.382	0.299
	20	0.004	0.010	0.006	0.007
1,000	5	0.031	0.059	0.688	0.673
	20	0.014	0.009	0.055	0.133
(b) mESS/s					
100	5	1,698	2,673	11,682	10,460
	20	82	316	148	378
1,000	5	844	1,959	14,898	15,469
	20	133	126	559	1,587

the mESS values dropped markedly, indicating the growing difficulty of accurate sampling in higher-dimensional parameter spaces under the hierarchical prior structure. In terms of mESS per second, DA-MCMC-Approx again exhibited the best computational efficiency, particularly for  $k = 20$ , where its performance exceeded that of DA-MCMC-Exact. These results suggest that, although the hierarchical shrinkage introduced by the NIG-homo prior increases computational complexity, the proposed DA-MCMC framework remains effective and stable across a wide range of sample sizes and model dimensions.

The results for the NIG-hetero prior are summarized in Table III. Similar to the NIG-homo case, the overall efficiency decreased relative to the Normal prior, reflecting the increased complexity of sampling when each coefficient was assigned an individual variance parameter. Nonetheless, both DA-MCMC-Exact and DA-MCMC-Approx clearly outperformed the benchmark methods across all scenarios. In terms of mESS per iteration, DA-MCMC-Exact tended to yield slightly higher values, especially in lower-dimensional settings ( $k = 5$ ), indicating that the richer hierarchical structure did not prevent effective exploration of the posterior distribution. However, in higher dimensions ( $k = 20$ ), the efficiency gap between DA-MCMC-Exact and DA-MCMC-Approx narrowed, with the latter showing a modest advantage in mESS per second due to its reduced computational burden. Overall, DA-MCMC-Approx achieved a favorable balance between efficiency and stability, even under the more flexible, heterogeneous prior structure. These results confirm that the proposed framework remains robust when extended to priors imposing coefficient-specific shrinkage, such as those used in high-dimensional regression and sparse modeling contexts.

Across all prior specifications and experimental settings, the proposed DA-MCMC algorithms consistently outperformed the benchmark methods in both sampling efficiency and computational speed. The DA-MCMC-Exact variant achieved the highest per-iteration efficiency, whereas DA-MCMC-Approx offered superior overall performance measured by effective sample size per second. The relative advantage of the Approx version became more pronounced as dimensionality increased or sample size grew, underscoring its scalability and numerical stability. Taken together, these results demonstrate that the

TABLE III  
RESULTS FOR SYNTHETIC DATA (3) NIG-HETERO PRIOR

$n$	$k$	MCMC	DA-MCMC	Mod. DA-MCMC	
				Exact	Approx
(a) mESS/iter					
100	5	0.039	0.059	0.355	0.328
	20	0.004	0.010	0.082	0.032
1,000	5	0.031	0.059	0.645	0.673
	20	0.015	0.010	0.526	0.531
(b) mESS/s					
100	5	1,493	2,429	9,993	10,850
	20	75	286	1,488	1,135
1,000	5	834	1,849	13,225	14,686
	20	122	118	3,628	4,196

proposed framework provides a flexible and computationally efficient tool for quasi-Bayesian inference across a wide range of model and prior configurations.

#### IV. APPLICATION TO REAL DATA

We applied the proposed approach to infer an IV regression model (2). The inference procedure followed the same framework as that used for the linear regression model, with the only distinction being the transformation of the quasi-likelihood. Specifically, the exponential term in the quasi-likelihood was modified as follows:

$$\begin{aligned} & \exp\left(-\frac{n}{2}\bar{\mathbf{m}}(\boldsymbol{\theta})^\top \mathbf{W}(\boldsymbol{\theta})\bar{\mathbf{m}}(\boldsymbol{\theta})\right) \\ &= \exp\left(-\frac{n}{2}\left[\frac{1}{n}\mathbf{Z}^\top(\mathbf{y} - \mathbf{X}\boldsymbol{\theta})\right]^\top \mathbf{W}\left[\frac{1}{n}\mathbf{Z}^\top(\mathbf{y} - \mathbf{X}\boldsymbol{\theta})\right]\right) \\ &\propto \exp\left(-\frac{n}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}^\dagger)^\top \mathbf{G}^\top \mathbf{W} \mathbf{G}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}^\dagger)\right), \end{aligned}$$

where  $\hat{\boldsymbol{\theta}}^\dagger = (\mathbf{Z}^\top \mathbf{X})^{-1} \mathbf{Z}^\top \mathbf{y}$  and  $\mathbf{G} = n^{-1} \mathbf{Z}^\top \mathbf{X}$ . Note that when the IV regression is exactly identified,  $\hat{\boldsymbol{\theta}}^\dagger$  coincides with the two-stage least squares estimator [30]. We employed the NIG-hetero prior with the same hyperparameters as in Section III.

We applied the IV regression to two real datasets. The first dataset, denoted as AJR, is originally compiled by [31], [32]<sup>3</sup>. They investigated how the risk of expropriation affected gross domestic product per capita. To address potential endogeneity in this relationship, European settler mortality was used as an instrumental variable. The specification also includes several control variables: a constant term, the latitude of each country (and its square), and dummy variables indicating whether the country is located in Africa or Asia, as well as whether it belongs to the group of former British colonies (Australia, Canada, New Zealand, and the United States). The sample consists of  $n = 64$  observations, and the number of moment conditions is  $k = 10$ .

The second dataset, denoted as Movies, originates from [33]. We used the version provided in Chapter 12 of [34]<sup>4</sup>.

<sup>3</sup><https://www.openicpsr.org/openicpsr/project/112564/version/V1/view>.

<sup>4</sup>[https://www.princeton.edu/~mwatson/Stock-Watson\\_4E/Stock-Watson-Resources-4e.html](https://www.princeton.edu/~mwatson/Stock-Watson_4E/Stock-Watson-Resources-4e.html).

TABLE IV  
RESULTS FOR REAL DATA

Data	$n$	$k$	MCMC	DA-MCMC	Mod. DA-MCMC	
					Exact	Approx
(a) mESS/iter						
AJR	64	10	0.004	0.007	0.015	0.004
Movies	516	36	0.004	0.004	–	0.145
(b) mESS/s						
AJR	64	10	85	278	509	224
Movies	516	36	25	42	–	3,257

The data cover 516 weekends between 1995 and 2004 and record the number of assaults across selected U.S. counties, along with national attendance figures for highly violent films. We estimated the relationship between weekend assault counts and film attendance using an instrumental variables approach, where predicted attendance serves as an instrument for observed attendance. The specification also includes a comprehensive set of control variables, such as fixed effects for year and month, indicators for holiday weekends, and multiple weather-related covariates. The dimension of the inferential problem is characterized by  $n = 516$  and  $k = 36$ .

The results are summarized in Table IV. In both empirical applications, the proposed DA-MCMC methods outperformed the benchmark algorithms in sampling efficiency. For the AJR dataset, DA-MCMC-Exact achieved the highest mESS per iteration (mESS/iter) and per second (mESS/s), reflecting strong computational efficiency in a moderately sized, exactly identified model. DA-MCMC-Approx also performed competitively, providing substantial improvement over the baseline methods at a lower computational cost.

For the Movies dataset, the advantage of DA-MCMC-Approx became especially pronounced. The Exact version was computationally infeasible in this higher-dimensional setting ( $k = 36$ ), whereas DA-MCMC-Approx achieved exceptionally high efficiency, yielding an mESS/s more than an order of magnitude greater than that of the benchmark algorithms. These results demonstrate the scalability and robustness of the Approx implementation in complex, high-dimensional empirical problems.

Overall, the empirical analyses reinforce the findings from the synthetic data experiments. Both implementations of the proposed DA-MCMC framework produced substantial gains in sampling efficiency over conventional MCMC and DA-MCMC methods, even in realistic econometric settings. The DA-MCMC-Exact variant provided the most accurate inference in low- to moderate-dimensional models, whereas the DA-MCMC-Approx variant proved considerably more scalable and computationally stable in higher-dimensional applications. These results highlight the practical versatility of the proposed approach for quasi-Bayesian inference in diverse empirical contexts.

## V. DISCUSSION

This paper has proposed a computationally efficient framework for quasi-Bayesian inference based on the DA-MCMC

algorithm. By exploiting the linear structure of moment conditions, the method enables the construction of proposal distributions that closely approximate the conditional posterior, enhancing both mixing and computational performance. Two implementations—DA-MCMC-Exact and DA-MCMC-Approx—were developed to achieve a balance between computational efficiency and numerical stability. Simulation studies using synthetic data showed substantial improvements in sampling efficiency relative to standard MCMC and DA-MCMC algorithms, while empirical applications to real datasets reinforced the scalability and robustness of the proposed approach in practical settings.

Future research could extend the framework to nonlinear and overidentified models, refine the surrogate kernel, and explore integration with modern variational and sequential inference techniques. Overall, the proposed DA-MCMC framework offers a versatile and computationally tractable tool for quasi-Bayesian analysis in complex statistical modeling.

## REFERENCES

- [1] T. Zhang, “Information-theoretic upper and lower bounds for statistical estimation,” *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1307–1321, 2006.
- [2] T. Zhang, “From  $\epsilon$ -entropy to KL-entropy: Analysis of minimum information complexity density estimation,” *Annals of Statistics*, vol. 34, no. 5, pp. 2180–2210, 2006.
- [3] W. Jiang and M. A. Tanner, “Gibbs posterior for variable selection in high-dimensional classification and data mining,” *Annals of Statistics*, vol. 36, no. 5, pp. 2207–2231, 2008.
- [4] P. G. Bissiri, C. C. Holmes, and S. G. Walker, “A general framework for updating belief distributions,” *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 78, no. 5, pp. 1103–1130, 2016.
- [5] N. Syring and R. Martin, “Gibbs posterior concentration rates under sub-exponential type losses,” *Bernoulli*, vol. 29, no. 2, pp. 1080–1108, 2023.
- [6] J.-Y. Kim, “Limited information likelihood and Bayesian analysis,” *Journal of Econometrics*, vol. 107, no. 1–2, pp. 175–193, 2002.
- [7] V. Chernozhukov and H. Hong, “An MCMC approach to classical estimation,” *Journal of Econometrics*, vol. 115, no. 2, pp. 293–346, 2003.
- [8] G. Yin, “Bayesian generalized method of moments,” *Bayesian Analysis*, vol. 4, no. 2, pp. 191–207, 2009.
- [9] L. P. Hansen, “Large sample properties of generalized method of moments estimators,” *Econometrica*, vol. 50, no. 4, pp. 1029–1054, 1982.
- [10] A. R. Hall, *Generalized Method of Moments*. Oxford University Press, 2004.
- [11] H. Hong, H. Li, and J. Li, “BLP estimation using Laplace transformation and overlapping simulation draws,” *Journal of Econometrics*, vol. 222, no. 1, pp. 56–72, 2021.
- [12] G. Yin, Y. Ma, F. Liang, and Y. Yuan, “Stochastic generalized method of moments,” *Journal of Computational and Graphical Statistics*, vol. 20, no. 3, pp. 714–727, 2011.
- [13] M. Tanaka, “Delayed acceptance markov chain monte carlo for robust bayesian analysis,” *Springer Proceeding in Mathematics and Statistics*, forthcoming.
- [14] J. A. Christen and C. Fox, “Markov chain Monte Carlo using an approximation,” *Journal of Computational and Graphical Statistics*, vol. 14, no. 4, pp. 795–810, 2005.
- [15] D. T. Frazier, C. Drovandi, and R. Kohn, “Calibrated generalized Bayesian inference,” *arXiv preprint arXiv:2311.15485*, 2024.
- [16] A. Zellner, “An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias,” *Journal of the American Statistical Association*, vol. 57, no. 298, pp. 348–368, 1962.
- [17] D. G. Fiebig, “Seemingly unrelated regression,” *A Companion to Theoretical Econometrics*, pp. 101–121, 2001.

- [18] Ò. Jordà, "Estimation and inference of impulse responses by local projections," *American Economic Review*, vol. 95, no. 1, pp. 161–182, 2005.
- [19] Ò. Jordà and A. M. Taylor, "Local projections," *Journal of Economic Literature*, vol. 63, no. 1, pp. 59–110, 2025.
- [20] J. L. Montiel Olea and M. Plagborg-Møller, "Local projection inference is simpler and more robust than you think," *Econometrica*, vol. 89, no. 4, pp. 1789–1823, 2021.
- [21] J. D. Angrist and A. B. Krueger, "Instrumental variables and the search for identification: From supply and demand to natural experiments," *Journal of Economic Perspectives*, vol. 15, no. 4, pp. 69–85, 2001.
- [22] S. Burgess, D. S. Small, and S. G. Thompson, "A review of instrumental variable estimators for Mendelian randomization," *Statistical Methods in Medical Research*, vol. 26, no. 5, pp. 2333–2355, 2017.
- [23] W. A. Fuller, *Measurement Error Models*. John Wiley & Sons, 2009.
- [24] M. Tanaka, "Adaptive MCMC for generalized method of moments with many moment conditions," *arXiv preprint arXiv:1811.00722*, 2021.
- [25] T. Park and G. Casella, "The Bayesian Lasso," *Journal of the American Statistical Association*, vol. 103, no. 482, pp. 681–686, 2008.
- [26] C. M. Carvalho, N. G. Polson, and J. G. Scott, "The horseshoe estimator for sparse signals," *Biometrika*, vol. 97, no. 2, pp. 465–480, 2010.
- [27] A. Bhattacharya, D. Pati, N. S. Pillai, and D. B. Dunson, "Dirichlet–Laplace priors for optimal shrinkage," *Journal of the American Statistical Association*, vol. 110, no. 512, pp. 1479–1490, 2015.
- [28] M. Vihola, "Robust adaptive Metropolis algorithm with coerced acceptance rate," *Statistics and Computing*, vol. 22, no. 5, pp. 997–1008, 2012.
- [29] D. Vats, J. M. Flegal, and G. L. Jones, "Multivariate output analysis for Markov chain Monte Carlo," *Biometrika*, vol. 106, no. 2, pp. 321–337, 2019.
- [30] W. Greene, *Econometric Analysis, 8th Edition*. Pearson, 2017.
- [31] D. Acemoglu, S. Johnson, and J. A. Robinson, "The colonial origins of comparative development: An empirical investigation," *American Economic Review*, vol. 91, no. 5, pp. 1369–1401, 2001.
- [32] D. Acemoglu, S. Johnson, and J. A. Robinson, "The colonial origins of comparative development: An empirical investigation: Reply," *American Economic Review*, vol. 102, no. 6, pp. 3077–3110, 2012.
- [33] G. Dahl and S. DellaVigna, "Does movie violence increase violent crime?," *Quarterly Journal of Economics*, vol. 124, no. 2, pp. 677–734, 2009.
- [34] J. H. Stock and M. W. Watson, *Introduction to Econometrics, 4th Edition*. Pearson, 2020.