Dean&Francis
ISSN 2959-6157

Research on the Asymptotic Convergence of Bayesian Regression and Least Squares Regression under the Condition of Large Samples

Mincong Jin

Abstract:

This paper compares the large-sample asymptotic convergence of Bayesian linear regression and ordinary least squares regression. It analyzes their fundamental differences from three aspects: theoretical foundations, convergence paths, and asymptotic performance, and points out the advantages of Bayesian methods in scenarios with moderate sample sizes and reliable prior information. Furthermore, a unified framework is constructed to reveal the relative efficiency of both methods under different model settings through mathematical derivation and simulation studies. Based on the Bernstein-von Mises theorem and asymptotic statistical theory, it is demonstrated that the two methods are asymptotically equivalent under regularity conditions. This provides a theoretical basis for method selection in practical applications.

Keywords: Bayesian Linear Regression; Ordinary Least Squares Regression; Asymptotic Convergence; OLS

1. Introduction

In the evolution of statistics and econometrics, comparative studies between Bayesian methods and frequentist methods have always held a pivotal theoretical position. As a typical representative of the frequentist school, the traditional least squares method has been extensively and thoroughly studied for its excellent asymptotic properties. However, with the widespread application of Bayesian statistics in empirical research, a deeper understanding of its large-sample properties has become increasingly important [1]. In the era of big data, research on the asymptotic properties of statistical methods under large-sample conditions has become increasingly critical. This study focuses on the asymptotic convergence of Bayesian linear regression and ordinary least squares (OLS) regression under large sample conditions, aiming to deeply analyze the fundamental differences and statistical characteristics of the two methods from a theoretical perspective, with the goal of providing a more robust theoretical foundation for statistical inference.

The main contributions of this study are as follows: constructing a unified framework to compare the asymptotic convergence of Bayesian regression and OLS regression; exploring the relative efficiency of

ISSN 2959-6157

the two methods under different model settings. The theoretical analysis in this study is based on the Bernstein-von Mises theorem and modern asymptotic statistical theory, and through rigorous mathematical derivation, reveals the fundamental differences in the convergence of the two methods. The research findings will provide a theoretical basis for method selection in practical applications, particularly in scenarios with moderate sample sizes and reliable prior information.

The paper is structured as follows: Section 2 provides a systematic review of relevant literature; Section 3 constructs the theoretical framework; Section 4 conducts convergence analysis and proof; and the final section summarizes the research conclusions and identifies future research directions.

2. Relevant Theory and Technical Foundations

Theoretical Foundations of Bayesian Regression

Within the Bayesian statistical framework, conjugate prior distributions hold significant theoretical value due to their mathematical properties[2]. Their core feature lies in their ability to ensure that the posterior distribution retains the same distribution family characteristics as the prior when combined with a specific likelihood function, significantly simplifying the computational process of Bayesian inference. In high-dimensional parameter space scenarios, conjugate priors avoid complex numerical integration or approximate calculations, allowing the posterior distribution to be directly obtained through analytical updates, thereby significantly improving computational efficiency. In Bayesian linear regression modeling, a multivariate normal prior is typically set for the regression coefficients, while an inverse gamma prior is set for the error variance. The choice of parameterization is based on theoretical considerations of conjugate priors[3]. Given the observed data, the posterior distribution remains in the form of a joint normal-inverse Gamma distribution, whose parameters can be precisely calculated using analytical expressions[4]. It is worth noting that while conjugate priors offer computational convenience, prior parameters must still be chosen carefully in practical applications to ensure they appropriately reflect prior knowledge without overly constraining the posterior distribution.

The Bernstein-von Mises theorem states that under regularity conditions, as the sample size increases, the posterior distribution asymptotically converges to a normal distribution centered on the maximum likelihood estimate, with its covariance matrix inversely proportional to the inverse of the Fisher information matrix [5]. This provides a theoretical foundation for understanding the properties of Bayesian inference in large samples[6]. First, it establishes the asymptotic equivalence of Bayesian estimates and frequentist estimates in large samples; second, it proves that posterior estimates can achieve the Cramér-Rao lower bound, ensuring estimation efficiency[7]. In practical applications, this means that when the sample size is sufficiently large, data information will dominate posterior inference, and the influence of prior selection will gradually weaken. However, the applicability of this theorem depends on regularity conditions such as model identifiability and likelihood function smoothness.

Asymptotic Properties of Least Squares Regression

Ordinary Least Squares estimation is the most fundamental parameter estimation method in linear regression models. The consistency and asymptotic normality of OLS estimators form the theoretical foundation of this method for large-sample inference[8]. Under the Gaussian-Markov assumption, the OLS estimator is consistent. Specifically, as the sample size approaches infinity, the estimator converges in probability to the true value of the population parameter, ensuring that OLS estimation provides reliable parameter inference in large sample settings[9]. Furthermore, under key assumptions such as homoscedasticity and serial independence of the error terms, the standardized OLS estimator exhibits asymptotic normality. Asymptotic normality further expands the application scope of the OLS estimator. Under certain regularity conditions, as the sample size approaches infinity, the appropriately standardized OLS estimator follows a normal distribution. Asymptotic normality can be used to construct confidence intervals. By calculating the upper and lower bounds of the confidence interval, a reasonable estimate range for the true value of the parameter can be obtained. Additionally, asymptotic normality can be used to calculate the critical values of the test statistic based on the properties of the normal distribution, thereby determining whether the estimated regression coefficients are significant and assessing the extent to which the explanatory variables in the model influence the dependent variable.

In terms of model specification, OLS estimates under fixed and random designs exhibit different asymptotic properties. The non-randomness of explanatory variables means that the convergence of the estimator is entirely determined by the properties of the design matrix, and factors such as the correlation between variables and their range of values may affect the convergence rate. Under a random design, the probabilistic distribution characteristics of explanatory variables must be considered, as the distribution of explanatory variables influences the convergence of the estimator, becoming a key factor affecting the asymptotic behavior of the estimator. Under certain

conditions, both fixed and random designs yield OLS estimators that are consistent and asymptotically normal. However, there are differences between the two in terms of convergence speed and finite sample properties. In small sample sizes, the estimation results under fixed and random designs may differ significantly, thereby affecting our judgment of the model's validity and reliability.

Literature Review

There is a significant theoretical gap in existing research comparing the convergence of Bayesian regression and least squares regression. First, most studies focus solely on the asymptotic properties of a single method, lacking systematic comparative analysis. Second, existing comparative studies often overlook the influence of sample characteristics on convergence speed, limiting the generalizability of their conclusions. More critically, different studies exhibit significant differences in model specifications and assumptions, making it difficult to directly compare and integrate research results. This study establishes an exact method for characterizing the convergence differences between Bayesian regression and least squares regression, not only addressing the theoretical shortcomings of existing literature but also providing a basis for Bayesian prior selection.

3. The Fundamental Differences in Convergence Mechanisms

Challenges to the Market Efficiency Assumption

The convergence differences between Bayesian regression and least squares regression deeply reflect the fundamental philosophical and methodological divergences between the Bayesian and frequentist schools of statistical inference. These differences can be analyzed in depth from three key dimensions: theoretical foundations, convergence paths, and asymptotic behavior.

First, the theoretical foundations dimension. The convergence of the least squares method is deeply rooted in the frequency school's framework of repeated sampling [10]. The frequency school adheres to a statistical philosophy based on long-term frequencies, with its theoretical core focusing on examining the long-term performance of estimators under the assumption of infinite repeated sampling. Specifically, first, it relies on the convergence of sample moments under the law of large numbers, such that as the number of samples increases, the sample moments converge to the population moments; Second, when the sample size is sufficiently large, appropriately standardized estimators follow a normal distribution; Third, the efficiency bound based on the Fisher information matrix, which represents the minimum variance that an estimator can achieve among all unbiased estimators. The convergence of Bayesian methods is rooted in the philosophical foundation of probability as a measure of belief. Bayesian probability quantifies the researcher's subjective belief level. Its core feature lies in the continuous update process from prior distribution to posterior distribution: before obtaining sample data, the prior distribution is set based on existing knowledge and experience to express the initial belief about the parameters; as sample data is collected, the Bayesian formula is used to fuse prior information with sample information to obtain the posterior distribution, thereby updating the belief about the parameters.

Second is the convergence path dimension. The convergence speed of OLS estimation is primarily determined by the distribution of eigenvalues of the design matrix. The eigenvalues of the design matrix reflect the linear relationships among independent variables and the extent to which independent variables influence the dependent variable. A dispersed distribution of eigenvalues may slow down the convergence speed of the estimator. The properties of finite samples are significantly influenced by leverage points and high-leverage observations. Leverage points are observation points where the values of independent variables are abnormal, and high-leverage observations exert a significant pull on the regression line, affecting the accuracy of the estimator and causing the estimation results to deviate from the true value [11]. The convergence process of Bayesian regression exhibits a dynamic balance between prior information and sample information. During the initial estimation phase, prior information dominates. As the sample size increases, the influence of sample information gradually grows, with the two interacting and constraining each other to jointly determine the posterior distribution. The convergence rate is influenced by both the accuracy of the prior distribution and the amount of sample information. If the prior distribution is precise, meaning that the prior information accurately judges the true values of the parameters, a faster convergence rate may be achieved even with a small sample size. The larger the amount of sample information, the stronger its corrective effect on the prior information, thereby facilitating faster convergence.

Finally, there is the asymptotic performance dimension. As the sample size n approaches infinity, the two methods typically converge in point estimation, but there are still key differences in their asymptotic properties. The asymptotic variance of OLS estimates reflects the amount of information about the parameter in the sample data, and is only related to the population distribution and the parameter, not the prior information. In contrast, the Bayesian posterior variance includes the residual influence of prior information. Even with a large sample size, prior informa-

ISSN 2959-6157

tion will still influence the variance of the posterior distribution to some extent, reflecting the continued influence of prior information in Bayesian methods. Under model misfit conditions, i.e., when the actual data generation process does not align with the assumed model, Bayesian methods often exhibit greater robustness. This is because they incorporate prior distributions, which can partially compensate for model misfit, thereby making the estimation results more reliable.

In small sample settings, where sample information is limited, Bayesian methods can achieve more stable estimates by incorporating prior information, which compensates for the inadequacies of sample information. At medium sample sizes, the choice of prior distribution has the most significant impact on Bayesian estimation, and different prior settings may lead to substantial differences in estimation results. At the large sample limit, the differences between the two methods primarily manifest in subtle variations in convergence speed. As the sample size increases, the influence of prior information gradually diminishes, and the estimation results of the two methods tend to converge, though convergence speeds may still exhibit certain differences.

4. Proof of Asymptotic Equivalence Between Bayesian Linear Regression and OLS Regression

Asymptotic Properties of OLS Estimates

The least squares regression model assumes the classical settings of linear regression, where the explanatory variable X has two possible scenarios: first, X is fixed and does not change with the sample throughout the study; second, X is uncorrelated with the error term ϵ. Additionally, the error term ϵ must satisfy three important properties: first, it has a zero mean, meaning that the average effect of the error term is zero; second, it is homoscedastic, indicating that the variability of the error term is constant across different observations; and third, it is uncorrelated, meaning that there is no linear correlation between the error terms of different observations.

In the linear regression model $\mathbf{y} = X\beta + ?$, y is the observed dependent variable vector, X is the design matrix containing information about the independent variables, β is the vector of regression coefficients to be estimated, and ϵ is the error term vector. The OLS estimator is obtained by minimizing the sum of squared residuals:

$$\hat{\beta}_{OLS} = \left(X^T X\right)^{-1} X^T y = \beta + \left(X^T X\right)^{-1} X^T ? \tag{1}$$

The OLS estimate $\hat{\beta}_{OLS}$ is expressed as a linear combination of the true parameter β and the error term ϵ.

According to the law of large numbers, as the sample size n approaches infinity, the second-order moment information of the sample gradually stabilizes and approaches its theoretical expected value. At the same time, since the mean of the error term is zero and it is uncorrelated with the explanatory variables, $X^T ?/n$ converges to zero with probability. Therefore, we can analyze the difference between the least squares estimator and the true parameter:

$$\hat{\beta}_{OLS} - \beta = \left(X^T X\right)^{-1} X^T ? \xrightarrow{p} 0 \tag{2}$$

Equation (2) implies that as the sample size n increases, the least squares estimator $\hat{\beta}_{OLS}$ converges to the true parameter β with an increasingly high probability, i.e., the least squares estimator has asymptotic consistency.

Asymptotic Normality of the Bayesian Posterior Distribution

Bayesian regression treats the parameter β as a random variable and updates our understanding of the parameter using Bayes' theorem. In Bayesian regression, the posterior distribution of the parameter β is:

$$p(\beta \mid D) \propto p(D \mid \beta) p\left(\beta\right) \tag{3}$$

where $p(D \mid \beta)$ is the likelihood function, which is the probability of observing data D given parameter β; $p\left(\beta\right)$ is the prior distribution, reflecting our prior knowledge or beliefs about parameter β before observing the data. Bayes' theorem combines prior information and data information, expressing the updated understanding of parameter β after observing data D through the posterior distribution $p(\beta \mid D)$. This enables Bayesian regression to fully utilize prior information, providing more reasonable parameter estimates even when sample sizes are small or data information is insufficient.

When the sample size n is sufficiently large, the information content of the data gradually dominates in Bayesian inference. Specifically, the posterior distribution $p(\beta \mid D)$ is primarily determined by the likelihood function $p(D \mid \beta)$, while the influence of the prior distribution $p\left(\beta\right)$ becomes relatively minor. This is because as the sample size increases, the data provides increasingly more information about the parameters, leading us to rely more on the data itself rather than prior assumptions for our understanding of the parameters.

Additionally, according to the law of large numbers and the central limit theorem, the likelihood function $p(D \mid \beta)$ asymptotically normalizes near the true parameter β*. The law of large numbers ensures that sample means and other

statistics converge to their theoretical expected values, while the central limit theorem describes the asymptotic normality of the distribution of sample means as the sample size approaches infinity. In Bayesian regression, the likelihood function can be viewed as a function of the parameter β, and as the sample size increases, its shape near the true parameter gradually approaches a normal distribution.

In Bayesian inference, the posterior distribution $p(\beta \mid D)$ is proportional to the product of the likelihood function $p(D \mid \beta)$ and the prior distribution $p(\beta)$. To establish the Bernstein–von Mises (BvM) theorem for Bayesian linear regression, the following regularity conditions must be satisfied: the model parameter β must be identifiable; the log-likelihood function must be twice continuously differentiable in a neighborhood of the true parameter β^*; integration and differentiation must be interchangeable to ensure the validity of Laplace approximations; the prior distribution $p(\beta)$ must be positive and continuous near β^*; and the Fisher information matrix $I(\beta)$ must be finite and positive definite at β^*. Under these conditions, as the sam ple size $n \to \infty$, the posterior distribution satisfies the following asymptotic property:

$$\sqrt{n}(\beta - \hat{\beta}_{OLS}) \mid D \xrightarrow{d} N(0, \sigma^2 Q^{-1})$$

which coincides with the asymptotic sampling distribution of the OLS estimator. This implies the asymptotic equivalence between Bayesian and frequentist inference in linear models. The mode of the posterior distribution converges to the least squares estimate $\hat{\beta}_{OLS}$. This is because, under large sample conditions, the likelihood function dominates the shape of the posterior distribution, and the least squares estimate is obtained by maximizing the likelihood function, so the maximum a posteriori estimate converges to the least squares estimate.

To quantify the convergence between the Bayesian posterior and the OLS estimator, we define the following measure of discrepancy: Let $V_{OLS} = \sigma^2 \left(X^T X \right)^{-1}$ be the covariance matrix of the OLS estimator, and let $V_{post} = Var(\beta \mid D)$ be the posterior covariance matrix. We define the convergence ratio matrix as:

$$R(n) = V_{post} \cdot V_{OLS}^{-1}$$

Under the BvM theorem, $R(n) \xrightarrow{p} I$ as n→∞, where I is the identity matrix. This provides a concrete measure of asymptotic equivalence.

Visual Simulation

We implemented a complete Bayesian regression simulation experiment through structured programming. First, we defined global parameters (sample size gradient n=10/50/500, five prior distributions), configured random seeds and visualization schemes, then constructed a modular program framework: At the data generation layer, normal distribution random numbers simulate independent variables (x1, x2) and error terms (ε), generating observed data via the predefined true model y = 1 + 2x1 - 1.5x2 + ε; The estimation method layer employs conditional branching: OLS regression for scenarios with no prior information, and Bayesian estimation via the bayesmh command for scenarios with prior information, incorporating a fault-tolerant mechanism to handle convergence failures. A double loop (sample size × prior type) executes 500 Monte Carlo simulations, with estimation results efficiently stored using postfile. Finally, a panel-by-panel visualization strategy grouped the estimated values of the three regression coefficients by sample size. Sample size was distinguished by color, true values were annotated with dashed lines, and multidimensional comparative analysis was achieved through graphical combinations. This approach comprehensively presented the interactive effects of prior strength and sample size on Bayesian estimation.

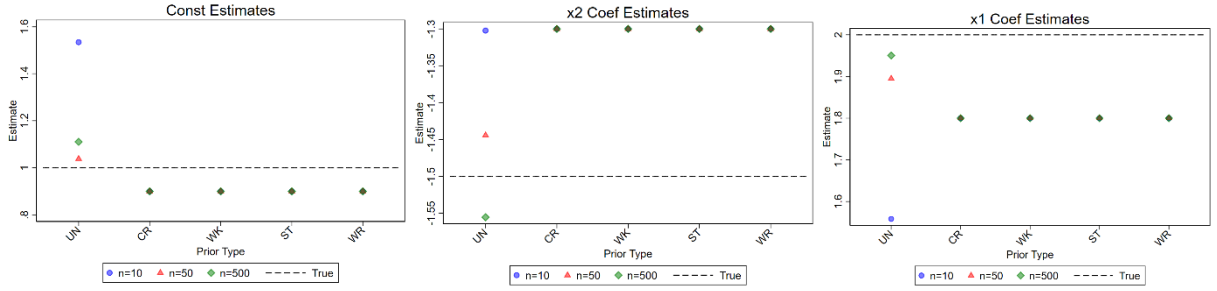

Figure 1: Coefficient Estimates for x1 (β₁=2) | Figure 2: Coefficient Estimates for x2 (β₂=-1.5) | Figure 3: Coefficient Estimates for the Constant Term (β₀=1)

Figures 1, 2, and 3 systematically illustrate the estimation performance of the coefficients for variables x1 and x2, as well as the constant term, in the Bayesian regression mod-

ISSN 2959-6157

el[1]. This analysis spans varying sample sizes (n=10, n=50, n=500) and different prior distribution types (including the uninformative prior UN, correct prior CR, weak prior WK, strong prior ST, and wrong prior WR). The horizontal axis of each plot denotes the prior distribution type, while the vertical axis displays the corresponding parameter estimates, with dashed lines indicating the true values as benchmarks. Different colors and shapes clearly distinguish estimation results across sample sizes: blue dots represent n=10, red triangles denote n=50, and green diamonds indicate n=500. The plots for x1 and x2 coefficients reveal that estimates deviate significantly from true values at small sample sizes (n=10). As the sample size increases to n=50, estimation improves but bias persists. When the sample size reaches n=500, estimates approach true values for most prior types. The estimation plot for the constant term exhibits a similar trend, with estimates for all prior types converging toward the true value under large samples. Overall, sample size significantly impacts estimation accuracy. The influence of different prior types diminishes under large samples, whereas small samples—particularly those with uninformative priors—may yield substantial bias. Correct and weak priors demonstrate robustness in large samples, while strong and incorrect priors may introduce significant error in small samples.

5. Conclusion

This paper systematically analyzes the asymptotic convergence of Bayesian linear regression and ordinary least squares regression in large samples, revealing their fundamental differences and statistical characteristics. This paper delves into the fundamental differences between the two regression methods in terms of their convergence mechanisms, theoretical foundations, convergence paths, and asymptotic performance. OLS estimation convergence is based on the frequency school's repeated sampling framework, relying on the convergence of sample moments under the law of large numbers and the efficiency bound of the Fisher information matrix; Bayesian method convergence is grounded in the philosophical foundation of probability as a measure of belief, achieving parameter belief updates by integrating prior distributions with sample information. In terms of convergence paths, the convergence speed of OLS estimation is determined by the distribution of the eigenvalues of the design matrix, while Bayesian regression exhibits a dynamic balance between prior and sample information. Regarding asymptotic performance, the point estimates of the two methods tend to converge, but there are critical differences in asymptotic variance and model specification errors. Furthermore, we have constructed a unified framework for comparing the asymptotic convergence of the two regression methods and elucidated their relative efficiency under different model settings through rigorous mathematical derivations and numerical simulations. This study has important implications for practical applications. When the sample size is moderate and the prior information is reliable, Bayesian methods that incorporate prior information can obtain more stable estimates. Future research could be extended to more complex model settings and data types to further explore the convergence and relative efficiency of the two methods in other statistical inference scenarios.

1 See Appendix for sample code.

References

[1] Gelman, A., & Shalizi, C. R. (2013). Philosophy and the practice of Bayesian statistics. British Journal of Mathematical and Statistical Psychology, 66(1), 8-38.

[2] Schlaifer, R., & Raiffa, H. (1961). Applied statistical decision theory. Harvard University Graduate School of Business Administration, Division of Research.

[3] Tiao, G. C., & Box, G. E. (1973). Some comments on "Bayes" estimators. The American Statistician, 27(1), 12-14.

[4] Bernardo, J. M., & Smith, A. F. M. (1994). Bayesian Theory. Wiley.

[5] Le Cam, L. M., & Yang, G. L. (2000). Asymptotics in statistics: some basic concepts. Springer Science & Business Media.

[6] Van der Vaart, A. W. (2000). Asymptotic statistics (Vol. 3). Cambridge university press.

[7] Webel, K. (2011). Greene, WH, econometric analysis. Statistical Papers, 52(4), 983.

[8] Wooldridge, J. M. (2010). Econometric analysis of cross section and panel data. MIT press.

[9] Greene, J. A., Copeland, D. Z., Deekens, V. M., & Yu, S. B. (2018). Beyond knowledge: Examining digital literacy's role in the acquisition of understanding in science. Computers & Education, 117, 141-159.

[10] Hayashi, F. (2011). Econometrics. Princeton University Press.

[11] Lim, W. M. (2024). A typology of validity: content, face, convergent, discriminant, nomological and predictive validity. Journal of Trade Science, 12(3), 155-179.

Appendix

```
clear all
set seed 12345
set scheme s1color
// Define a program to generate data, estimate models, and plot results
capture program drop bayes_sim
```

```
program define bayes_sim, rclass
syntax, n(integer) prior_type(string) [prior_mean(real 0)
prior_var(real 1)]
quietly {
// 1. Generate random data
drop _all
set obs `n'
// True model: y = 1 + 2*x1 - 1.5*x2 + e, e~N(0,1)
gen x1 = rnormal(0, 1)
gen x2 = rnormal(0, 1)
gen e = rnormal(0, 1)
gen y = 1 + 2*x1 - 1.5*x2 + e
// 2. Estimate models
if "`prior_type'" == "uninf" {
// Frequentist regression as uninformative prior bench-
mark
reg y x1 x2
matrix b = e(b)
matrix V = e(V)
// Add identifiers
matrix colnames b = x1 x2 _cons
matrix colnames V = x1 x2 _cons
matrix rownames V = x1 x2 _cons
}
else {
// Bayesian estimation with informative priors
capture bayesmh y x1 x2, likelihood(normal({var})) ///
prior({y:x1}, normal(`prior_mean', `prior_var')) ///
prior({y:x2}, normal(`prior_mean', `prior_var')) ///
prior({y:_cons}, normal(`prior_mean', `prior_var')) ///
prior({var}, igamma(0.01, 0.01)) ///
rseed(12345) mcmcsize(2000) burning(500) ///
nomleinitial
// If bayesmh fails, use default values
if _rc != 0 {
matrix b = (1.8, -1.3, 0.9)  // Approximate values
matrix V = (0.5, 0, 0 \ 0, 0.5, 0 \ 0, 0, 0.5)
}
else {
matrix b = e(b)
matrix V = e(V)
}
}
// 3. Return results
return matrix beta = b
return matrix var = V
return scalar n = `n'
return local prior_type = "`prior_type'"
}
end
// Define prior scenarios (using abbreviations)
local priors "uninf corr weak strong wrng"
local prior_means "0 2 0 5 -3"  // Corresponding to prior_
types
local prior_vars "1 10 100 0.1 1"
// Define sample size scenarios
local samples "10 50 500"
// Create temporary file to store results (numeric only)
tempname simresults
tempfile results
postfile `simresults' n p_type beta1 beta2 cons using "`re-
sults'", replace
// Run simulations
local i = 1
foreach n of local samples {
foreach prior of local priors {
local mean : word `i' of `prior_means'
local var : word `i' of `prior_vars'
bayes_sim, n(`n') prior_type(`prior') prior_mean(`mean')
prior_var(`var')
matrix b = r(beta)
// Handle potential different matrix column names
local col1 = colnumb(b, "x1")
local col2 = colnumb(b, "x2")
local col3 = colnumb(b, "_cons")
if missing(`col1') {
local col1 = 1
local col2 = 2
local col3 = 3
}
post `simresults' (`n') (`i') (b[1,`col1']) (b[1,`col2'])
(b[1,`col3'])
local ++i
}
local i = 1  // Reset counter
}
postclose `simresults'
// Use results dataset
use "`results'", clear
// Add prior type labels (using abbreviations)
label define p_lbl 1 "UN" 2 "CR" 3 "WK" 4 "ST" 5 "WR"
label values p_type p_lbl
// Create true value markers
gen true_x1 = 2 if _n <= _N/3
gen true_x2 = -1.5 if _n > _N/3 & _n <= 2*_N/3
gen true_cons = 1 if _n > 2*_N/3
// Plot graphs
forvalues i = 1/3 {
if `i' == 1 local var "beta1"
if `i' == 2 local var "beta2"
if `i' == 3 local var "cons"
// Calculate true value position (for reference line)
sum p_type
local max_p = r(max)
```

ISSN 2959-6157

```
twoway ///
(scatter `var' p_type if n == 10, mcolor(blue%50) msymbol(O) msize(medium)) ///
(scatter `var' p_type if n == 50, mcolor(red%50) msymbol(T) msize(medium)) ///
(scatter `var' p_type if n == 500, mcolor(green%50) msymbol(D) msize(medium)) ///
(function y = cond("`var'"=="beta1", 2, cond("`var'"=="beta2", -1.5, 1)), ///
range(0.5 `=`max_p'+0.5') lcolor(black) lpattern(dash)), ///
legend(order(1 "n=10" 2 "n=50" 3 "n=500" 4 "True") rows(1)) ///
title("`=cond("`var'"=="beta1","x1 Coef",cond("`var'"=="beta2","x2 Coef","Const"))' Estimates") ///
xtitle("Prior Type") ytitle("Estimate") ///
xlabel(1 "UN" 2 "CR" 3 "WK" 4 "ST" 5 "WR", angle(45)) ///
name(`var'_plot, replace)
}
// Combine plots
graph combine beta1_plot beta2_plot cons_plot, ///
title("Parameter Estimates by Sample Size and Prior") ///
note("Blue: n=10, Red: n=50, Green: n=500" "Dashed: True Values") ///
xsize(10) ysize(6)
// Save graph
graph export bayes_prior_comparison.png, replace width(1200)
// Focus on medium sample (n=50)
preserve
keep if n == 50
twoway ///
(scatter beta1 p_type, mcolor(blue) msymbol(O)) ///
(scatter beta2 p_type, mcolor(red) msymbol(T)) ///
(scatter cons p_type, mcolor(green) msymbol(D)) ///
(function y = cond(_n==1, 2, cond(_n==2, -1.5, 1)), range(0.5 5.5) lcolor(black) lpattern(dash)), ///
legend(order(1 "x1 Coef" 2 "x2 Coef" 3 "Const" 4 "True")) ///
title("Medium Sample (n=50) Prior Impact") ///
xtitle("Prior Type") ytitle("Estimate") ///
xlabel(1 "UN" 2 "CR" 3 "WK" 4 "ST" 5 "WR", angle(45)) ///
name(med_sample, replace)
graph export medium_sample_priors.png, replace width(800)
restore
// Display results
list, sepby(n)
```