MULTICOLLINEARITY IN APPLIED ECONOMICS RESEARCH AND THE BAYESIAN LINEAR REGRESSION

Eric EISENSTAT
Department of Economics, University of California, Irvine, United States
e-mail: eric.eisenstat@gmail.com

Abstract

This article revises the popular issue of collinearity amongst explanatory variables in the context of a multiple linear regression analysis, particularly in empirical studies within social science related fields. Some important interpretations and explanations are highlighted from the econometrics literature with respect to the effects of multicollinearity on statistical inference, as well as the general shortcomings of the once fervent search for methods intended to detect and mitigate these effects. Consequently, it is argued and demonstrated through simulation how these views may be resolved against an alternative methodology by integrating a researcher's subjective information in a formal and systematic way through a Bayesian approach.

Key-words: *multiple linear regressions, classical normal regression, collinearity, multicollinearity, classical inference, subjective probability, Bayesian linear regression, prior information, posterior distributions, simulation*

JEL Classification: C11, C12

1. **Introduction**

Multicollinearity is the exotic term accorded by econometricians to denote strong linear relationships (e.g. collinearity) amongst explanatory variables in a multiple linear regression analysis. Indeed, multicollinearity and its effects on statistical inference are well explored topics in econometric literature (for an extensive overview and analysis see Judge, Hill, Griffiths, Lütkepohl, & Lee, 1988, pp. 859-881). Why is collinearity in explanatory variables of such emphasized importance in econometrics? The fundamental problem for an economist engaged in an applied study, as for almost any other social science related statistical investigation for that matter, is that strong linear relationships amongst explanatory variables pose not only a formidable impedance on statistical inference regarding individual parameters, but a vastly elusive one at that; since economists rarely have direct control over the *data generating process*, they neither control the variation in explanatory variables, nor possess the option to obtain larger and/or different samples for their application.

From a statistics perspective, the common view projects that collinearity undermines accurate inference through its effect on the standard errors of individual parameter estimates: *ceteris paribus*, stronger collinearity proportionately increases standard errors, leads to wider confidence intervals, and lower test statistics (in absolute value) in significance tests.[1] Accordingly, when an empirical study yields a particular regression parameter *statistically insignificant* (i.e. *fails to reject* the appropriate hypothesis test), there are two reasons why this occurs: (i) the *true* value of the parameter of interest is in fact zero, or (ii) the data sample is not informative enough to conclusively distinguish this parameter as *statistically significantly* different from zero. Since the latter is strongly related to the degree of collinearity in explanatory variables, a researcher particularly interested in demonstrating a statistically significant relationship may, therefore, be motivated not only to employ techniques intended to alleviate (ii), but also to search for a reasonable justification of rejecting (i) informally by arguing that the failure of the significance test is more attributable to (ii) through the presence of multicollinearity.

Countless methods of "detecting" multicollinearity and containing its effects *ex post* have been proposed. In lieu of redesigning experiments that generate the data or obtaining larger samples (options which are most often simply not available to economists), the operational "solutions" in the literature almost exclusively focus on either the systematic inclusion/exclusion of certain explanatory variables, or the reconditioning of explanatory variables such as to induce orthogonality and yield lower degrees of collinearity according to some predetermined measure, depending on a particular case of interest.

It is important to emphasize, however, that all such detection mechanisms and *ex post* data-manipulation solutions have almost unilaterally fallen victim to a criticism that consistently resounds a common theme – they are all invariably *ad hoc*. That is, there is no formal argumentative justification, neither from the perspective of probability theory nor classical statistics inference, for the general use of such proposed solutions. Simply put, colliniarity in explanatory variables is *one* feature (among several) of the data that is directly related to the amount of information provided by the sample. When the sample is not informative enough to lead to decisive conclusions, the only potential "solution" to this is to introduce more information, but if the new information does not manifest itself in the form of additional/different data, such information can only be subjective. Insofar as classical statistics inference outright rejects subjective information, however, it is not surprising that the search for operational solutions within this framework has failed to produce generally accepted techniques to combat multicollinearity: techniques incorporating subjective information within an objectivist paradigm are indeed *ad hoc*.

The fundamental premise of the present paper is that there is in fact an appropriate place for subjective information in statistical inference. With this as a basis, it is argued that properly characterizing one's subjective information and formally incorporating it into an empirical study is indeed an effective way to obtain conclusive statistical inference where the data alone may not offer definitive answers.

[1] This is, however, not generally true for linear combinations of parameters. For an example where multicollinearity leads to increased power in a hypothesis test on a sum of regression parameters see Goldberger, 1991, pp. 250-251.

This is generally accomplished by formulating subjective information in terms of *prior beliefs* and employing Bayesian methods to systematically integrate the prior beliefs with the information available in the data.

To support the proceeding argument, section 2 provides an overview of the multiple *Classical Normal Regression (CNR)* model with emphasis on the properties that are of particular interest to our discussion, as well as an in-depth, formal description of the effects of multicollinearity in this context. Accordingly, section 3 introduces the Bayesian linear regression and concludes with a simulation example that demonstrates how multicollinearity may be handled from the Bayesian perspective in empirical applications within an economic study.

2. **Classical Normal Regression and Multicollinearity**

Recall that the multiple linear regression postulates a linear relationship between a dependent variable y and k explanatory variables $x_1, \dots, x_k$ of the form

$$y_i = \beta_0 + \beta_1 x_{i1} + \sum_{j=2}^{k} \beta_j x_{ij} + \epsilon_i \tag{1}$$

where in accordance with the assumptions of the CNR model, $\epsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$. Without loss of generality, we designate β_1 to be the parameter of interest and $\beta_0, \beta_2, \dots, \beta_k$ and σ^2 to be the *nuisance* parameters. Moreover, *we* will denote the *Ordinary Least Squares (OLS)* estimators of β_j, and σ^2 with $\hat{\beta}_j$, and $\hat{\sigma}^2$, respectively.

In what follows, it is convenient to express the estimators employing the following matrix notation: let $\boldsymbol{y}$ be the $n \times 1$ vector consisting of dependent variables $\boldsymbol{y} = (y_1 \cdots y_n)'$, $\hat{\boldsymbol{\beta}}$ the $(k+1) \times 1$ vector $(\hat{\beta}_0 \cdots \hat{\beta}_k)'$, and let $\boldsymbol{X}$ denote the $n \times (k+1)$ matrix with elements $X(i,1) = 1$ and $X(i,j) = x_{i,j-1}$ for $i = 1, \dots, n$ and $j = 2, \dots, k+1$. Furthermore, define $\boldsymbol{Q} = (\boldsymbol{X}'\boldsymbol{X})^{-1}$, with $\boldsymbol{q}_j$ representing the j^{th} row of $\boldsymbol{Q}$ and q_{jh} denoting the element of $\boldsymbol{Q}$ located at row j and column h. Then,

$$\hat{\sigma}^2 = \frac{(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})}{n-k-1} \tag{2}$$

$$\hat{\beta}_1 = \boldsymbol{q}_2 \boldsymbol{X}'\boldsymbol{y} \tag{3}$$

It is well known that $\hat{\beta}_1$ and $\hat{\sigma}^2$ are stochastically independent and marginally follow the distributions

$$(n-k-1)\hat{\sigma}^2/\sigma^2 \sim \chi^2(n-k-1) \tag{4}$$

$$\hat{\beta}_1 \sim \mathcal{N}(\beta_1, \sigma^2 q_{22}) \tag{5}$$

By letting $R_{1,k}^2$ represent the *coefficient of determination* obtained from regressing x_1 on all other explanatory variables $x_2, \ldots, x_k$, and $s_{x_1}^2$ the sample variance of x_1, we may write $q_{22} = \frac{1}{n(1-R_{1,k}^2)s_{x_1}^2}$. From this expression, the conventional view on the effect of multicollinearity is immediately evident: as collinearty between x_1 and other explanatory variables increases, $R_{1,k}^2 \to 1$ and for a fixed sample size $\mathrm{var}(\hat{\beta}_1) \to \infty$. Thus, multicollinearity is typically related to imprecise estimates, which is in turn, reflected in wide confidence intervals as well as weak *power* in hypothesis tests on individual parameters.

Because the computation of confidence intervals and individual test statistics related to the parameter β_1 involve only the statistics $\hat{\beta}_1$, $\hat{\sigma}^2$, and q_{22}, it is evident that the epicentre of the multicollinearity effect lies in the variance of the individual parameter estimate under consideration. Yet the reason that a higher degree of collinearity in explanatory variables increases the variance of parameter estimates (or equivalently, the standard errors) is that larger collinearty means *less variability* in the sample, and hence, less information. Less informative data, in turn, naturally leads to more imprecise (or less certain) estimates. The central question, however, is how much collinearity is too much?

To that end, it should be noted that the effect of collinearity on the variance of parameter estimates is significant *only* relative to the sample size. In fact, if the joint explanatory vector $(x_{i1} \cdots x_{ik})$ is independent across observations i, the effect (on the variance of an individual parameter estimate) of an *increase* in the sample size by one observation is equivalent to the effect of a *decrease* in the auxiliary coefficient of determination $R_{j,k}^2$ by $\frac{1}{n}(1 - R_{j,k}^2)$. Intuitively, both collinearity and sample size may be viewed as two very similar factors that determine the variability in the sample, which is the primary source of information offered by the data for statistical inference. Hence, the extent of either effect (high collinearity or low sample size) on individual parameter inference must be interpreted accordingly.

A succinct and elegant interpretation of the severity of collinearity is best offered by the esteemed econometrician Arthur S. Goldberger (Goldberger, 1991, p. 252):

To say that "standard errors are inflated by multicollinearity" is to suggest that they are artificially, or spuriously, large. But in fact they are appropriately large: the coefficient estimates actually would vary a lot from sample to sample. This may be regrettable but it is not spurious.

Note that from a purely classical objectivist perspective that obstinately refutes all prior information in statistical inference, this claim is undisputable. That is, one certainly cannot commit inferential exclusivity to a set of data, and upon receiving vague inference from that data, dismiss this vagueness on the grounds that the data is "poorly conditioned." Within the bonds of data exclusivity, one simply has no way of judging to what extent is a confidence interval "unreasonably" wide, since there exists no basis for comparison in succinctly defining "unreasonably." Indeed, a larger degree of collinearity reduces the ability of the data to identify the statistical significance of individual parameters, as does a lower sample size, lower sum of squared residuals in explanatory variables, etc., and without an ulterior source of information this data deficiency is incircumventable; different techniques of manipulating the data can only exploit more efficiently (or less efficiently) the variability already present, but will never induce more variability.

On the other hand, it is entirely reasonable to gauge the degree to which "standard errors are inflated" by admitting prior information because the prior belief provides exactly the basis for comparison lacking in a purely objectivist paradigm. More specifically, one may justifiably claim that a confidence interval is "too wide" if it extends into regions where *a priori* the researcher assigns a low degree of probability in the sense that the information offered by the data regarding the parameter of interest contradicts the prior information. In addition, the extent of this "contradiction" may be sensibly measured when the prior information is formalized in a probabilistic manner. It is in this sense that prior information provides an additional instrument, which in the presence of collinearity provides the crucial supplement to variability lacking in the data.

As a matter of fundamental principle, classical statistics inference offers little in terms of accommodation for subjective prior beliefs, and incorporating such beliefs in this paradigm, even when formulated probabilistically, is an awkward exercise at best. Bayesian methods, on the other hand, are well known to be the most efficient way of systematically combining prior information with the data in generating robust statistical inference (for introductory Bayesian texts, see (Koop, 2003) and (Gelman, Carlin, Stern, & Rubin, 2003)). To that end, we provide a simple demonstration of how prior information may be employed in alleviating adverse effects of collinearity within the Bayesian linear regression framework (Koop, 2003, pp. 15-85), (Gelman, Carlin, Stern, & Rubin, 2003, pp. 351-385), (Poirier, 1995, pp. 524-580).

3. **Bayesian Linear Regression**

In concept, Bayesian inference differs fundamentally from classical inference in the following sense: the focus of Bayesian inference is on what the parameter is *most likely to be*, whereas the most common concern of classical inference is on what the parameter is *definitely not*. Nevertheless, there are strong practical parallels between the two approaches. For example, given a particular significance level α the $(1-\alpha)\%$ *posterior probability interval* (commonly constructed as the *highest posterior density (HPD)* interval) bears a close resemblance to the $(1-\alpha)\%$ *confidence interval* for either individual parameters or a combination of parameters, while the *mode* of the posterior distribution is comparable to the parameter *estimate* generated by classical techniques. More importantly, as the sample size increases, both the posterior modes and posterior probability intervals converge to the corresponding *Maximum Likelihood* estimates and confidence intervals (Poirier, 1995, pp. 306-307).

Note that the latter fact reflects exactly the previously outlined intuition regarding the effect of collinearity relative to sample size. Insofar as the effect of collinearity is most apparent in smaller samples and diminishes proportionately as n

increases, it is crucial that whatever instrument is adapted to offset the effects of collinearity in smaller samples reduces in relative importance as the sample size grows. Employing prior information through Bayesian techniques achieves just that: prior information is most influential on the posterior distribution, and hence most effective in combating collinearity, when n is small, while this influence is proportionately reduced as n increases and vanishes altogether as $n \to \infty$.

We illustrate the Bayesian approach in this context through a simple simulation example based on some well-known results of the Bayesian linear regression. Accordingly, suppose the model of interest is

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i, \;\; \epsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2) \qquad (6)$$

and the *true* parameter values are

$$\beta_0 = 15, \qquad \beta_1 = 1, \qquad \beta_2 = -4, \qquad \sigma^2 = 2$$

Using a *pseudo-random* number generator, we simulate three data samples of $\{y_i, x_{i1}, x_{i2}\}$ for $n = 100$, $n = 1{,}000$, and $n = 10{,}000$, respectively, and compare the inference one would obtain under the Bayesian framework to that of the classical framework. Since our primary interest lies in the influence of collinearity on statistical inference, the simulated data is generated to yield a relatively high sample correlation between the explanatory variables while each pair $\{x_{i1}, x_{i2}\}$ is sampled independently. Moreover, the experiment is designed such that the correlation between x_1 and x_2 emits a stronger effect on the precision of the estimates of β_1 relative to β_2. Summary statistics of the simulated datasets are reported in Table 1.

Table 1

Simulation Data Summary Statistics

	n	average	standard deviation	correlation y	x_1	x_2
y	100	−70.424	2.411	1.000	−0.733	−0.840
x_1	100	5.033	0.258		1.000	0.891
x_2	100	22.596	0.554			1.000
y	1,000	−69.976	2.421	1.000	−0.703	−0.815
x_1	1,000	4.987	0.246		1.000	0.893
x_2	1,000	22.487	0.559			1.000
y	10,000	−70.007	2.470	1.000	−0.702	−0.811
x_1	10,000	5.003	0.247		1.000	0.893
x_2	10,000	22.504	0.551			1.000

Now, consider how an econometrician might approach the task of estimating this model aware only of the descriptive properties of the data and operating under the assumption that the linear model is *correctly specified* as given in (6). Assume further that the econometrician is in possession of the smallest sample ($n = 100$) and is concerned that the *strength* of collinearity relative to this sample size may lead to uninterestingly vague inference regarding her primary parameters of interest β_1 and β_2. On the other hand, her theoretical training endows her with some key intuition regarding the values of these parameters. She summarizes her beliefs as follows:

1) centered at $\beta_1 = 2$, $\beta_2 = -2$;
2) symmetric (i.e. $\beta_1 = 0$ is just as likely as $\beta_1 = 4$, etc.);
3) highly unlikely that $\beta_1 > 5$ or $\beta_2 < -5$.

These beliefs may be formalized in terms of *prior probability distributions* regarding β_1 and β_2. Consequently, we shall proceed with a general form of the prior distribution given by

$$\beta_j \mid \sigma^2 \overset{nid}{\sim} \mathcal{N}\left(m_{j+1}, \sigma^2 v_{j+1,j+1}^2\right), \quad \sigma^2 \sim \mathcal{IG}\left(\frac{g}{2}, \frac{h}{2}\right) \qquad (7)$$

where $\mathcal{IG}(\cdot)$ denotes the *inverse gamma distribution* (for example, see (Gelman, Carlin, Stern, & Rubin, 2003, pp. 573-577)). It can be shown that the implied *marginal* distribution of β_j is

$$\frac{\beta_j - m_{j+1}}{v_{j+1,j+1}\sqrt{h/g}} \sim t(g) \qquad (8)$$

where $t(g)$ *student-t distribution* with g degrees of freedom, and therefore, all three prior beliefs described above may be accommodated in (8) by appropriately setting the parameters m_{j+1} and $v_{j+1,j+1}^2$. Specifically, let $m_2 = 2$, $m_3 = -2$, $v_{2,2}^2 = v_{3,3}^2 = \frac{3g}{h} / \Psi_g^{-1}(c)$, where $\Psi_g^{-1}(\cdot)$ denotes the *cumulative distribution function (cdf)* of the *student-t* distribution with g degrees of freedom This ensures that the modes of the distributions for β_1 and β_2 are 2 and -2, respectively, while $\Pr(\beta_1 > 5) = \Pr(\beta_2 \leq -5) = c$, where c may be set to any reasonably small value, (e.g. the ensuing results are based on $c = 0.01$). The symmetry condition is, of course, automatically satisfied since the student-t distribution is naturally symmetric.

Note that Bayesian methods require that prior distributions be properly specified for *all* parameters. Since, the researcher is neither particularly interested in β_0 and σ^2, nor does she posses very specific beliefs regarding their values, she may specify g, h, m_1, and $v_{1,1}^2$ in such way that results in *mildly-informative* prior distributions for β_0 and σ^2. Such mild beliefs, for example, are sufficiently represented with the following values: $g = 2$, $h = 2$, $m_1 = 0$, $v_{1,1}^2 = 100$.

The essence of Bayesian inference is the focus on *updating* one's prior belief by the observed data sample. This, in turn, requires the construction of the *likelihood* function, which is operationally expressed as the distribution of the dependent variable conditional on the parameters:

$$\boldsymbol{y} \mid \boldsymbol{\beta}, \sigma^2 \sim \mathcal{N}(\boldsymbol{X\beta}, \sigma^2 \boldsymbol{I}_n) \qquad (9)$$

where $\boldsymbol{I}_n$ denotes the $n \times n$ *identity matrix*. Using (7) and (9), the *joint posterior* distribution is obtained through Bayes' Rule as

$$p(\boldsymbol{\beta},\sigma^2 \mid \boldsymbol{y}) = \frac{p(\boldsymbol{\beta},\sigma^2)\, p(\boldsymbol{y} \mid \boldsymbol{\beta},\sigma^2)}{p(\boldsymbol{y})} \quad (10)$$

and contains all information necessary to carry out statistical inference on the parameters. While exact analytic expressions for posterior distributions are, in general, intractable (most often, posterior inference is based on *simulating* from the posterior distribution), the Bayesian linear regression model yields fairly simple and practically straightforward posteriors.

Consequently, define the following notation: let $\boldsymbol{m} = (m_1\ m_2\ m_3)'$, $\boldsymbol{V} = \mathrm{diag}\{(v_{1,1}^2\ v_{2,2}^2\ v_{3,3}^2)\}$ (i.e. a 3×3 diagonal matrix), and

$$\begin{aligned}
\widehat{\boldsymbol{V}} &= (\boldsymbol{V}^{-1} + \boldsymbol{X}'\boldsymbol{X})^{-1} \\
\widehat{\boldsymbol{m}} &= \widehat{\boldsymbol{V}}(\boldsymbol{V}^{-1}\boldsymbol{m} + \boldsymbol{X}'\boldsymbol{y}) \\
\hat{h} &= h + (\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{m}})'(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{m}}) + (\widehat{\boldsymbol{m}} - \boldsymbol{m})'\boldsymbol{V}^{-1}(\widehat{\boldsymbol{m}} - \boldsymbol{m} \\
\hat{g} &= g + n
\end{aligned} \quad (11)$$

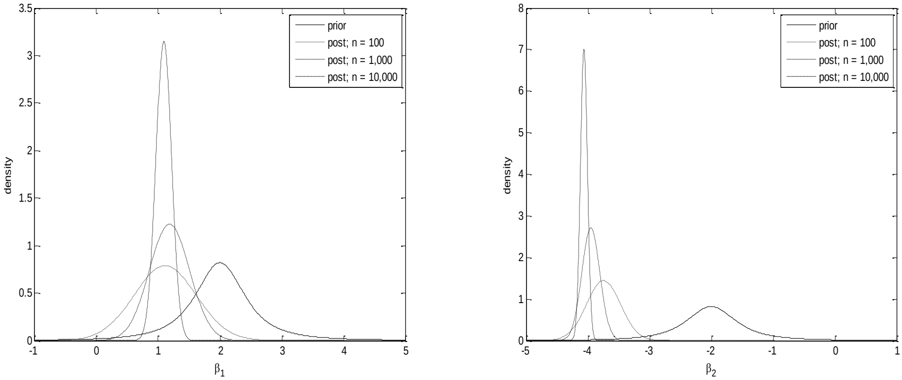
The joint posterior distribution of all model parameters in our case is then given by

$$\boldsymbol{\beta} \mid \sigma^2, \boldsymbol{y} \sim \mathcal{N}(\widehat{\boldsymbol{m}}, \sigma^2\widehat{\boldsymbol{V}}), \quad \sigma^2 \mid \boldsymbol{y} \sim \mathcal{IG}\left(\frac{\hat{g}}{2}, \frac{\hat{h}}{2}\right) \quad (12)$$

whereas the *marginal* posterior distributions of interest are obtained as

$$\frac{\beta_j - \widehat{m}_{j+1}}{\hat{v}_{j+1,j+1}\sqrt{\hat{h}/\hat{g}}} \mid \boldsymbol{y} \sim t(\hat{g}) \quad (13)$$

The marginal posterior distributions for each of the three cases (of varying sample size) under examination are plotted along with the corresponding prior distributions for the parameters β_1 and β_2 in Figure 1, Panels (A) and (B), respectively. More specifically, Panel (A) illustrates the evolution of the posterior distribution of β_1 from the prior as the sample size grows while Panel (B) depicts the analogous phenomenon for β_2. The intuition regarding the influence of prior information on posterior inference as n increases is immediately evident. In both cases, with each increasing sample size, the posterior distribution *collapses* around the mode, which in turn, converges to the true parameter value.



Panel (A): β_1 **Panel (B)**: β_2

Fig. 1. Prior and posterior distributions for β_1 and β_2

Observe, however, that the *collapsing* effect is distinctly slower for β_1 in comparison to β_2. This is precisely a reflection of the influence of collinearity, which is by design more influential in the posterior distribution of $\beta_1$than that of β_2. In fact, a closer examination of Panel (A) reveals that the posterior of β_1 for $n = 100$ is not noticeably less *dispersed* relative to its prior distribution, but rather only exhibits a shift in *location* towards the true value. A more illuminating interpretation of the latter may be formulated as follows: the posterior distribution of β_1 at $n = 100$ reflects a *joint* effort on the part of the prior information and the data whereby the information from the data is incorporated into more accurately centering the posterior while the prior maintains the dispersion contained by substituting for the lack of certainty projected by highly collinear data with *a priori* information. As a result, even with a relatively low sample size (i.e. relative to the degree of correlation in the explanatory variables), posterior inference regarding β_1 is sufficiently informative.

The important trade-off is, of course, that this gain in precision at $n = 100$ is strongly reliant on the prior beliefs, and hence, accentuates the importance of introducing prior information cautiously and in a manner that is convincingly justifiable. On the other hand, as the sample size grows and the information projected by the data gains in vigor, the need for the prior to contain the posterior precision diminishes and its role in determining the shape of the posterior distribution is marginalized. This is clearly reflected in Panel (A), by the progressive reduction of the posterior dispersion at $n = 1{,}000$ and $n = 10{,}000$ where the abundance of available observations overcomes the small sample deficiencies resulting from collinearity.

Table 2 and Table 3 further reinforce this intuition in a numerical comparison of Bayesian and classical inference that would be conventionally employed in interpreting the results for each of the three sample size levels. A quick overview of Table 3, which summarizes typical classical quantities of interest, reveals the diminishing effect of collinearity in increasing n: for both parameters β_1 and β_2, as n increases regression estimates converge to the true values, standard errors decrease, confidence intervals shrink, and significance test statistics grow (in absolute value). Additionally, this phenomenon is accelerated for quantities related to β_2, in evident parallel with the influence of prior information on posterior distributions, and is likewise explained by the more prominent influence of collinearity on the precision of β_1 estimates.

In this sense, the fact that the analogous posterior quantities detailed in Table 2 converge to their classical counterparts is unsurprising. In fact, at $n = 10{,}000$ the posterior modes of β_1 and β_2 (which for the student-t distribution are equivalent to the respective posterior means and medians) are nearly identical to the OLS estimates $\hat{\beta}_1$ and $\hat{\beta}_2$, respectively. Similarly, posterior standard deviations are approximately equivalent to the standard errors, as are the 95% posterior probability intervals to the 95% confidence intervals.

Where the posterior inference differs most significantly from classical inference, however, is in terms of the parameter β_1 for $n = 100$. Here, it is worthwhile to note that the 95% classical confidence interval extends over negative values of β_1. Indeed, the limited sample is not informative enough to identify β_1 as statistically significantly different from zero at the 5% significance level in the classical context (this is equivalently verified by the corresponding significance test failing to reject the null hypothesis $H_0 :\ \beta_1 = 0$ against $H_A :\ \beta_1 \neq 0$). On the Bayesian side, Table 2 illustrates that with a sample of $n = 100$ observations, the *lower bound* of the 95% posterior probability interval is notably *greater than* the hypothesized null value $\beta_1 = 0$. In an analogous statement of significance, therefore, our Bayesian inference allows us to confidently proclaim β_1 as statistically significantly different from zero, given our prior beliefs.[2]

Table 2

Bayesian/Posterior Inference

	n	mean / median / mode	standard deviation	95% Probability Interval (HPD) lower bound	95% Probability Interval (HPD) upper bound	zero outside interval
β_1	100	1.113	0.510	0.111	2.115	yes
	1,000	1.182	0.326	0.544	1.821	yes
	10,000	1.089	0.127	0.840	1.337	yes
β_2	100	−3.747	0.279	−4.294	−3.200	yes
	1,000	−3.951	0.147	−4.240	−3.662	yes
	10,000	−4.064	0.057	−4.175	−3.952	yes

[2] In fact, the Bayesian paradigm defines a formal methodology of Bayesian Hypothesis Testing which is based on Bayesian Posterior Odds and is generally unrelated in terms of inference to the HPD interval approach demonstrated here; for more details see (Poirier, 1995, pp. 376-392, 540-551). However, the technical and conceptual complexities involved in a satisfactory discussion of posterior odds are beyond the scope of our purpose. We only mention here that in the simplified setting of our example, and particularly insofar as our focus is on comparing posterior inference to classical inference, the HPD interval approach is sufficiently appropriate.

Table 3

Classical/Frequentist Inference

	n	estimate	standard deviation	95% Confidence Interval lower bound	95% Confidence Interval upper bound	significance test t	significance test significantly different from zero
β_1	100	0.743	1.135	−1.510	2.996	0.655	no
	1,000	1.201	0.399	0.419	1.983	3.013	yes
	10,000	1.090	0.130	0.836	1.345	8.392	yes
β_2	100	−3.965	0.528	−5.012	−2.917	−7.511	yes
	1,000	−4.001	0.175	−4.346	−3.657	−22.802	yes
	10,000	−4.069	0.058	−4.183	−3.955	−69.964	yes

4. **Conclusion**

The simulation example of the previous section serves to illustrate a simple case where incorporating subjective information through Bayesian methods yields a more conclusive statistical inference relative to its strictly objective, classical counterpart. This should be by no means misinterpreted as suggesting that subjective beliefs offer a general solution to the multicollinearity problem. A better way to perceive the role of subjective information in a particular application is to ask the question what kind of prior belief(s) would yield the affirmative results sought by the researcher? If such beliefs are justifiable through theoretical considerations related to the subject of interest, then Bayesian inference may offer convincing support for definitive empirical claims that cannot be asserted by objective inference alone, be it due to multicollinearity or other data sample related deficiencies. If the said beliefs, on the other hand, cannot be justified on a theoretical basis, conclusions drawn based on *ad hoc* claims will once again be subject to doubt and inevitable rejection.

REFERENCES

- Gelman, A., Carlin, J.B., Stern, H.S., & Rubin, D.B. (2003), *Bayesian Data Analysis* (2nd ed.). Boca Raton: Chapman & Hall/CRC.
- Goldberger, A. S. (1991), *A Course in Econometrics*, Cambridge: Harvard University Press.
- Greene, W.H. (2003), *Econometric Analysis* (5th ed.), New Jersey: Prentice Hall.
- Johnston, J., & DiNardo, J. (1997), *Econometric Methods* (4th ed.), New York: McGraw-Hill.
- Judge, G.G., Hill, R.C., Griffiths, W.E., Lütkepohl, H., & Lee, T.-C. (1988), *Introduction to the Theory and Practice of Econometrics* (2nd ed.), New York: John Wiley & Sons.
- Koop, G. (2003), *Bayesian Econometrics*, Chichester: John Wiley & Sons, Ltd.
- Poirier, D.J. (1988), *Frequentist and Subjectivist Perspectives on the Problems of Model Building in Economics*, “Journal of Economic Perspectives”, 2, 121-144.
- Poirier, D.J. (1995), *Intermediate Statistics and Econometrics: A Comparative Approach*, Cambridge: The MIT Press.