# BAYESAIN TESTING FOR A PRE-SELECTED REGRESSION COEFFICIENT

*Yonggang Lu, College of Business & Public Policy, University of Alaska Anchorage, 3211 Providence Drive, Anchorage, AK 99508, 907-786-4177, afyl@uaa.alaska.edu*

## ABSTRACT

We develop a new Bayesian regression model for testing a pre-specified regression coefficient. A simple, closed form of the Bayes factor is derived that depends only on the regression *t*-statistic and the usual associated *t*-df distribution. The prior that allows this form is simple and also meaningful, requiring minimal but practically important subjective inputs. The model is applied to various business problems including event study and risk decomposition analysis.

## INTRODUCTION

Regression analysis is a widely-used tool in various business fields. Many of these studies focus on inference for a specific regression coefficient. Inference using regression models is more commonly frequentist rather than Bayesian, but frequentist methods of inference have several obvious disadvantages. Especially, the inference heavily depends on the "$p \leq .05$" types of rules for rejecting null hypotheses. This type of rule is widely used in academic research, e.g. event studies [4], and also in industrial research [2]. However, as noted by [1], the null hypothesis can be true 30% of the time when $p$=0.05, casting doubt on the utility of the *p*-value for assessing evidence. Recognizing important drawbacks of frequentist methods, Bayesian inference has attracted more and more attention, particularly recently with the modern widespread use of Markov Chain Monte Carlo (MCMC).

Despite advantages of Bayesian methods, frequentist methods still enjoy popularity, for various reasons including, in particular, the forms of the test statistics are typically simple, easily understood and easily taught, particularly to analysts with less education in mathematical statistics. In this study, we bridge this gap, providing a useful Bayesian test for a regression coefficient that has comfortable correspondences for those who may be more familiar with the frequentist methods. In particular, the test depends directly and simply on the usual frequentist *t*-statistic and its associated *t*-df distribution. On the other hand, our prior is not completely vague, and allows the researcher the benefit of using subjective inputs.

## BAYES FACTOR FPR TESTING A PRE-SPECIFIED REGRESSION COEFFICIENT

Consider the usual multiple linear regression model

$$Y = X\beta + \varepsilon, \varepsilon \sim N(0, \sigma^2 I), \tag{1}$$

where $\beta$ is a $k \times 1$coefficient vector with unknown values arranged so that the parameter of interest is $\beta_1$ (all else is nuisance), $X$ is a fixed $n \times k$ design matrix (assumed full rank), and $\varepsilon$ is the $n \times 1$ error vector. We are concerned about testing the hypotheses

$$\begin{aligned} H_0 &: \beta_1 = 0 \\ H_1 &: \beta_1 \neq 0 \end{aligned}.$$

Our Bayes factor for testing the $H_0 : \beta_1 = 0$ has the simple form

$$BF_{0:1} = \frac{T_{n-k}\left(t\ ;\ 0,1\right)}{T_{n-k}\left(t\ ;\ \tilde{n}^{1/2}\lambda, 1+\tilde{n}\sigma^2_{\beta_1}\right)}. \tag{2}$$

As usual, $BF_{0:1} < 1$ favors $H_1$, with $BF_{0:1} < .05$ indicating strong support for $H_1$, i.e. [3]. The Bayes factor above is particularly attractive for both frequentists and Bayesians because it can be easily calculated exactly, free of annoying Monte Carlo or analytic approximation error: the central and non-central $t$ densities in the numerator and denominator ($T_v\ (t\ ;\ 0,\ 1)$ and $T_v\ (t\ ;\ a,\ b)$ respectively) are widely available in software. The form of the test is specifically appealing for frequentists because it depends on comfortable quantities of the usual frequentist regression-based $t$ statistic. The form of the test is also attractive for Bayesians, because it is a Bayesian test with easily specified priors $\lambda$ and $\sigma^2_{\beta_1}$, the prior mean and variance of the effect size $\beta_1/\sigma$ under the alternative hypothesis. Since $\beta_1/\sigma$ is a dimensionless quantity, generic a priori values are easily specified, e.g., $\lambda = 0$ and $\sigma_{\beta_1} = 0.5$ are sensible choices in the absence of further information. The term $\tilde{n}$ is intuitively appealing both for frequentists and for Bayesians: it is the effective sample size for estimating $\beta_1$. Specifically, $\tilde{n}$ is the inverse of the diagonal element of $\left(X^T X\right)^{-1}$ corresponding to $\beta_1$.

Finally, many statistical software packages have available commends to evaluate a central $t$ density and a non-central $t$ density with standard scale parameter. Therefore, for example, a SAS commend used for calculating the Bayes factor is given as

BF = pdf('T',t,df)/(pdf('T',t/sqrt(postv),df,ncp)/sqrt(postv)),

where $t$ is the value of $t$-statistic, postv = $1+\tilde{n}\sigma^2_{\beta_1}$ and ncp = $\tilde{n}^{1/2}\lambda/\left(1+\tilde{n}\sigma^2_{\beta_1}\right)^{1/2}$.

## REFERENCES


[1] Berger, J. O., and Sellke, T. Testing a Point Null Hypothesis: The Irreconcilability of $p$ Values and Evidence. *Journal of the American Statistical* Association, 1987, 82, 112-122.

[2] Chen, Y., A. Desclée, B. Phelps, and Rosten, J. Fixed-income Active Returns: Characteristics, Constraints, and Competition. *Fixed Income Research Reports*, Lehman Brothers Inc., 2006.

[3] Kass, R. E., and A. E. Raftery. Bayes Factors. *Journal of the American Statistical Association*, 1995, 90, 773-795.

[4] Kothari, S. P., and Warner, J. B. Econometrics of Event Studies in B. E. Eckbo (ed.), *Handbook of Corporate Finance: Empirical Corporate Finance*, *Handbooks in Finance Series*. Holland: Elsevier, 2005.