Abstract

In this paper, we revisit the threshold regression models, an important class of models in economic analysis. For example, multiple equilibria can give rise to threshold effects. The issue of conducting a valid inference of the regression parameters $\gamma$ when the threshold parameter $\tau$ needs to be estimated remains an open question in the frequentist literature. The non-standard aspect of the estimation problem motivates the use of Bayesian methods, which can correctly reflect the finite-sample uncertainty of estimating $\tau$ upon inference of $\gamma$. Our theoretical contribution is to establish a Bernstein-von Mises type theorem (Bayesian asymptotic normality) for $\gamma$ under a wide class of priors for the parameters, which essentially indicates an asymptotic equivalence between the conventional frequentist and the Bayesian inference. Our result is beneficial to both Bayesians and frequentists. A Bayesian user can invoke our theorem to convey his or her statistical result to the frequentist researchers. For a frequentist researcher, looking at the credible interval can serve as a robustness check for the finite sample uncertainty coming from the threshold estimation, and such sensitivity analysis is natural as our result guarantees the credible interval to converge to the frequentist confidence interval. The simulation studies show that the conventional confidence intervals tend to under-cover while credible intervals offer a reasonable coverage in general. As sample size increases, both methods coincide, as predicted from our theoretical conclusion. Using the data from Durlauf and Johnson (1995) on economic growth and Paye and Timmermann (2006) on stock return prediction, we illustrate that the traditional confidence intervals on $\gamma$ might under-represent the true sampling uncertainty.

Keywords— Threshold regression, Structural break, Bernstein-von Mises theorem, Sensitivity check, Model selection

\*E-mail address: kenichi.shimizu@brown.edu
1 Introduction

In this paper, we consider the class of linear regression models with a structural break, following the notations of Bai (1997):

\[ y_i = \begin{cases} 
  w_i' \alpha + z_i' \delta_1 + \epsilon_i, & \text{for } i = 1, \ldots, \lfloor n\tau \rfloor \\
  w_i' \alpha + z_i' \delta_2 + \epsilon_i, & \text{for } i = \lfloor n\tau \rfloor + 1, \ldots, n
\end{cases} \]

In such models, the threshold parameter \( \tau \) divides the samples into two groups or regimes. The relationship between the outcome \( y_i \) and the covariate \( z_i \) is determined by the regime to which the particular
observation \( i \) belongs. There could be some covariates whose relationship to \( y_i \), measured by \( \alpha \), might stay constant between the regimes. The unknown parameters include the threshold \( \tau \) as well as the regression coefficients \( \gamma = (\alpha, \delta_1, \delta_2) \).

Such models have been important in empirical economics. For instance, Durlauf and Johnson (1995) suggested that wealthy countries and poor countries can have different growth paths. In macroeconomics and empirical finance, it is common to observe an event that affects a change in the underlying model such as a decrease in output growth volatility in the 1980s known as "the Great Moderation," oil price shocks, and labor productivity change (e.g. Paye and Timmermann (2006)). An example where threshold regression is used in applied microeconomics is the tipping-point model; for example, see Card et al. (2008). Refer to Hansen (2011) for an overview of the extensive uses of threshold regression models in economic applications.

In the literature, the conventional least square estimators \( \hat{\tau} \) and \( \hat{\gamma} \) are computed as follows: for each candidate \( \tau \), compute the sum of squared residuals of the regression and denote the minimizing choice by \( \hat{\tau} \). Plug in the value \( \tau = \hat{\tau} \) in the model and denote \( \hat{\gamma} = \hat{\gamma}(\hat{\tau}) \) to be the resulting OLS coefficient estimator, where \( \hat{\gamma}(\tau) \) is the usual OLS estimator of \( \gamma \) assuming the break location \( \tau \).

Roughly speaking, the frequentist literature on this model can be divided into two groups in terms of the assumption made about the true jump size \( \delta_0 = \delta_2 - \delta_1 \). In the first group of the literature that includes Chan (1993) and Liu et al. (1997), the jump size \( \delta_0 \) is assumed to be constant with respect to sample size \( n \). The authors show that the convergence rate of \( \hat{\tau} \) is \( n^{-1} \) and the asymptotic distribution of \( \hat{\gamma} \) is the same as that of \( \hat{\gamma}(\tau_0) \). They derived the asymptotic distribution of \( \hat{\tau} \), but it depends on nuisance parameters, and hence the construction of confidence intervals for \( \tau \) is not feasible under this assumption.

In order to derive confidence intervals for \( \tau \), the second group of the authors such as Bai (1997) and Hansen (2000) assume that \( \delta_0 \) shrinks to zero as \( n \to \infty \), but slower than \( \sqrt{n} \to \infty \). In other words, \( \delta_0 = d n^\alpha \) where \( \alpha \in (-1/2, 0) \) and \( d \in \mathbb{R}^k \). This reduces the convergence rate of \( \hat{\tau} \) but enables them to find the asymptotic distribution that is free of nuisance parameters, and hence allows them to construct confidence interval for \( \tau \). For the regression coefficients, Bai (1997) and Hansen (2000) obtain the same asymptotic result as in the first group of the literature: the asymptotic distribution of \( \hat{\gamma} \) is the same as that of \( \hat{\gamma}(\tau_0) \). Intuitively speaking, this is because \( \delta_0 \) shrinks to zero but it is asymptotically still large enough for \( \tau \) to be correctly estimated with high certainty. In the literature, this shrinking jump size framework is sometimes interpreted as representing a moderately small size of the jump parameter.

For both groups mentioned above, \( \sqrt{n} (\hat{\gamma}(\hat{\tau}) - \gamma_0) \overset{d}{\to} N(0, V_\gamma) \), where \( V_\gamma \) is the standard asymptotic covariance matrix one obtains when \( \tau = \tau_0 \). This means that the econometrician can approximate the distribution of \( \hat{\gamma} \) by the conventional normal approximation as if \( \tau \) is known with certainty. Such asymptotic result implies that \( P \left( \gamma \in \hat{\Theta}(\hat{\tau}) \right) \to 1 - \alpha \) as \( n \to \infty \) where \( \hat{\Theta}(\tau) \) is the usual asymptotic \( (1 - \alpha) \)-level confidence region for \( \gamma \) under the assumption that \( \tau \) is known.

As Hansen (2000) himself points out, with finite samples, this procedure is likely to under-represent the true sampling uncertainty. See Section 6 of this paper for a simulation illustration of this point.
classic approach such as the delta method cannot be used because \( \hat{\gamma}(\tau) \) is not differentiable with respect to \( \tau \) and \( \hat{\tau} \) is not normally distributed. In order to account for such finite sample uncertainty, Hansen (2000) proposes a Bonferroni-type correction. For any \( \rho \in (0, 1) \), let \( \hat{\Gamma}(\rho) \) be the confidence interval for \( \tau \) with asymptotic coverage \( \rho \). The proposed confidence region is \( \hat{\Omega}_\rho = \bigcup_{\tau \in \hat{\Gamma}(\rho)} \hat{\Theta}(\tau) \). Because \( \hat{\Theta}_\rho \supset \hat{\Theta}(\hat{\tau}) \), \( P \left( \gamma \in \hat{\Theta}_\rho \right) \geq P \left( \gamma \in \hat{\Theta}(\hat{\tau}) \right) \rightarrow 1 - \alpha \) as \( n \rightarrow \infty \). Hence \( \hat{\Theta}_\rho \) is more conservative. The drawback of this procedure is that the researcher needs to select \( \rho \), and the inference could be sensitive to the choice of \( \rho \) in a finite sample as the author illustrates in his simulations.

Sample-splitting method provides an alternative solution to the failures of conventional estimators and confidence sets. In a sample-splitting approach, one selects the threshold parameter based on one subset of the data, and then conducts inference using the remaining data. In Section XX, we compare...

It is worth mentioning that there is another type of assumption on \( \delta_0 \) that recently attracted attention. The third group considers the small break framework: \( \delta_0 = dn^{-1/2} \). In order words, the jump size shrinks at the same rate as the sample uncertainty diminishes. The motivation in the frequentist literature for such a framework is to better reflect the finite sample uncertainty of the estimation of \( \tau \) on the asymptotic analysis. In this framework, \( \hat{\tau} \) is not consistent. Elliott and Müller (2014) suggest a way to construct a confidence interval for the regression coefficient. Andrews et al. (2019) argue that ignoring the sampling uncertainty of estimating \( \hat{\tau} \) potentially leads to an invalid inference on the parameter of interest \( \gamma \) and propose inference that reflect the data-dependent choice of the break location. Although the type of the assumption on \( \delta_0 \) is different in our paper, conceptually, our paper contributes to this literature from a Bayesian perspective. In addition, compared to the approaches in Elliott and Müller (2014) and Andrews et al. (2019), our proposed Bayesian method is computationally easier to implement.

For a Bayesian, this non-standard aspect of the estimation problem can be dealt with quite naturally by placing, for example, a uniform prior on a reasonable range of \( \tau \) and an uninformative improper prior (or a conjugate prior with a large prior variance) on \( \gamma \) and then by computing the marginal posterior probabilities. Any finite sample uncertainty of estimating \( \tau \) is automatically reflected in the marginal posterior probability of \( \gamma \). Note that unlike the conventional frequentist methods, Bayesian inference has a valid interpretation even in finite samples as it does not rely on asymptotics. Indeed, Bayesian estimations of change-point models have been very popular in statistics literature (for example, refer to Khodadadi and Asgharian (2008)).

In this paper, we study asymptotic behavior of Bayesian estimation of the considered model under the fixed jump size framework. Our theoretical contribution is to establish asymptotic equivalence between the frequentist method and the Bayesian approach. Specifically, we prove a Bernstein-von mises type theorem for the slope parameters \( \gamma \) which validates the frequentist interpretation of the Bayesian credible regions. Our result is beneficial to both Bayesian and frequentist researchers. A Bayesian user can invoke our theorem to convey his or her statistical result to the frequentist researchers. For a frequentist researcher, looking at the credible interval can serve as a robustness check for the finite sample uncertainty.
coming from the threshold estimation, and such sensitivity analysis is natural as our result guarantees
the credible interval to converge to the frequentist confidence interval.

Our theoretical results hold under a wide range of prior specifications. First, the prior on the regression
coefficients can be either improper uninformative or conjugate informative. Second, our assumption on
the prior for the threshold location is very mild. For example, a uniform prior satisfies the requirement.
Recently Baek (2019) investigated the same model (1). As the distribution of least-squares estimator for
τ might exhibit tri-modality for small jumps, she proposed a new point estimator \(\hat{\tau}\) based on a modified
objective function. Such modification is equivalent to specifying a certain type of prior for \(\tau\) which indeed
satisfies the requirement for our asymptotic result.

In this paper, we do not consider the shrinking jump size frameworks which were described above
because it is likely that the Bayesian method would deliver different results compared to the frequentist
counterpart. Nevertheless, in contrast to Bai (1997) and Hansen (2000), Bayesian credible intervals
for \(\tau\) are available without the assumption of the shrinking jump size. Furthermore, any finite sample
uncertainty when the true jump size is small will be automatically reflected in the posterior distribution
of \(\gamma\).

Estimation of change-point models such as structural break models is considered non-standard in a
sense that there is a non-regular parameter (e.g. threshold parameter) whose point estimator converges
faster than \(n^{-1/2}\), the rate at which the regular parameters (e.g. regression coefficients) converge. Despite
its popularity in applications, the Bayesian theoretical literature on non-regular estimation is very scarce.
To our knowledge, frequentist evaluation of Bayesian approach for structural break models has not been
studied in the literature. Ghosal and Samanta (1995) consider the general non-regular estimation problem
from Bayesian perspective and establish conditions under which Bernstein-von mises theorem holds for
the regular part of the parameter. However, their assumptions are difficult to verify in regard to our
model in consideration.

The paper is organized as follows. Section 2 introduces the model. Section 3 lists assumptions
made about the data-generating-process. Section 4 outlines the proposed Bayesian estimation. Section
5 presents the asymptotic theory of our Bayesian method. Section 6 presents simulation evidence to
assess the adequacy of the asymptotic theory. Section 7 reports empirical applications to the multiple
equilibria growth model of Durlauf and Johnson (1995) and the stock return prediction model of Paye
and Timmermann (2006) (in progress). Section 8 concludes. The mathematical proofs of propositions are
listed to the Appendix. Proofs of some intermediate lemmas can be found under Supplementary Material.
2 The model

Using the notations similar to those in Bai (1997), the model we consider is

\[ y_i = \begin{cases} 
  w_i^\prime \alpha + z_i^\prime \delta_1 + \epsilon_i, & \text{for } i = 1, \ldots, \lfloor n\tau \rfloor \\
  w_i^\prime \alpha + z_i^\prime \delta_2 + \epsilon_i, & \text{for } i = \lfloor n\tau \rfloor + 1, \ldots, n 
\end{cases} \]  

(1)

where \( w_i \) and \( z_i \) are \( d_w \times 1 \) and \( d_z \times 1 \) vectors of covariates and the random variable \( \epsilon_i \) is a regression error. \( \lfloor a \rfloor \) is the largest integer that is strictly smaller than \( a \). Note that \( \alpha \) stays unchanged across the regimes defined by the threshold parameter \( \tau \in (0, 1) \). The vectors \( \alpha, \delta_1, \delta_2, \) and \( \tau \) are unknown parameters. We assume \( \delta_1 \neq \delta_2 \), so that a change has taken place. Using the reparametrization \( x_i = (w_i^\prime, z_i^\prime) \), \( \beta = (\alpha', \delta_1')' \), and \( \delta = \delta_2 - \delta_1 \), the equations (1) can be rewritten as

\[ y_i = \begin{cases} 
  x_i^\prime \beta + \epsilon_i, & \text{for } i = 1, \ldots, \lfloor n\tau \rfloor \\
  x_i^\prime \beta + z_i^\prime \delta + \epsilon_i, & \text{for } i = \lfloor n\tau \rfloor + 1, \ldots, n 
\end{cases} \]  

(2)

Note that \( z_i \) is a subvector of \( x_i \). More generally, let \( z_i = R'x_i \), where \( R \) is a \( d_x \times d_z \) known matrix with full column rank and hence \( z_i \) is defined as a linear transformation of \( x_i \). For \( R = (0_{d_w \times d_z}, I_{d_z \times d_z})' \), we obtain the model (2). For \( R = I_{d_z} \), the pure change model is obtained.

To rewrite the model in matrix form, let us introduce further notations. Define \( Y = (y_1, \ldots, y_n)' \), \( \epsilon = (\epsilon_1, \ldots, \epsilon_n)' \), \( X = (x_1, \ldots, x_n)' \), \( X_{1\tau} = (x_1, \ldots, x_{\lfloor n\tau \rfloor}, 0, \ldots, 0)' \), \( X_{2\tau} = (0, \ldots, 0, x_{\lfloor n\tau \rfloor + 1}, \ldots, x_n)' \). Define \( Z, Z_{1\tau}, \) and \( Z_{2\tau} \) similarly. Then \( Z = XR, Z_{1\tau} = X_{1\tau}R, \) and \( Z_{2\tau} = X_{2\tau}R \). Then the equations (2) can be written as

\[ Y = X\beta + Z_{2\tau}\delta + \epsilon \]  

(3)

\[ = \chi_{\tau} \gamma + \epsilon \]  

(4)

where \( \chi_{\tau} = (X, Z_{2\tau}) \) and \( \gamma = (\beta, \delta)' \). Denote by \( S_n(\tau) \) the sum of squared residuals of the regression (3). The least-squares (LS) estimator of the threshold \( \tau \) as in Bai (1997) is defined as

\[ \hat{\tau} = \arg\min_{\tau \in (0,1)} S_n(\tau) \]

and the LS estimators for the regression coefficients \( \gamma = (\beta, \delta)' \) are:

\[ \hat{\gamma} = \hat{\gamma}(\hat{\tau}) \]

where \( \hat{\gamma}(\tau) \) denotes the usual OLS estimator assuming \( \tau \) is known.
3 Data generating process

The data are assumed to include \( n \) observations on a response and a vector of covariates: \( D^n = (Y^n, X^n) = (y_1, \ldots, y_n, x_1, \ldots, x_n) \). Conditional on \( X^n \), the response is generated according to the model (1) with the true parameters \( \theta_0 = (\beta_0, \delta_0, \sigma^2_0, \tau_0) \). We make the following assumptions about the true DGP:

A1. \( \delta_0 \neq 0 \).

A2. \( \epsilon_i \) is i.i.d. with \( E(\epsilon_i | x_i) = 0, E(\epsilon_i^2 | x_i) = \sigma^2_0 \).

A3. Assume that \( \Sigma_X = E[x_i x_i'] = \text{plim} \frac{1}{n} \sum_{i=1}^{n} x_i x_i' \) exists and is positive definite.

A4. For all \( \tau_1, \tau_2 \in (0, 1) \) with \( \tau_1 < \tau_2 \), \( \frac{1}{n} \sum_{i=1}^{n \tau_2 | n \tau_1 + 1} x_i \epsilon_i = O_p(n^{-1/2}) \) and \( \frac{1}{n} \sum_{i=1}^{n \tau_2 | n \tau_1 + 1} x_i x_i' = (\tau_2 - \tau_1) \Sigma_X + O_p(n^{-1/2}) \)

If the assumptions above hold, the theoretical results in Bai (1997) apply. The author showed that the convergence rate of \( \hat{\tau} \) is \( n^{-1} \) if \( \delta_0 \) is fixed with respect to the sample size:

\[
\hat{\tau} = \tau_0 + O_p(n^{-1})
\]

and showed that the LS regression estimator for \( \gamma = (\beta, \delta)' \) is asymptotically normal with the asymptotic covariance matrix being the same as if \( \tau_0 \) is known:

\[
\sqrt{n}(\hat{\gamma} - \gamma_0) \xrightarrow{d} N_{(d_x + d_z)}(0, \sigma^2_0 V^{-1})
\]  

(5)

where

\[
V = \text{plim} \frac{1}{n} \left( \begin{array}{c} \sum_{i=1}^{n} x_i x_i' \\ \sum_{i=1}^{n \tau_0 | n \tau_0 + 1} z_i x_i' \end{array} \right) \left( \begin{array}{c} \sum_{i=1}^{n} x_i x_i' \\ \sum_{i=1}^{n \tau_0 | n \tau_0 + 1} z_i x_i' \end{array} \right)' = \text{plim} \frac{1}{n} \chi_0 \chi_0'
\]

4 Bayesian approach

4.1 Prior

To develop a Bayesian estimation framework, we model the distribution of the regression error term by normal: \( \epsilon_i | \sigma^2 \sim N(0, \sigma^2) \). Note that the normality assumption is not made for the true DGP, so the model can be mis-specified. The prior distribution for \( \tau \) admits a density \( \pi(\tau) \) whose support is \( \Theta \) and the ratio \( \pi(\tau) / \pi(\tau') \) is assumed to be bounded for any \( \tau, \tau' \in \Theta \). For example, the uniform distribution on \( \Theta \) satisfies the requirement. Other priors that reflects researcher’s prior knowledge about \( \tau \) or some penalty on the values near the boundaries as in Baek (2019) can be incorporated as well.
For the regression coefficients \((\gamma, \sigma^2)' = (\beta, \delta, \sigma^2)'\), the prior can be either the improper uninformative prior or the Normal-Inverse-Gamma conjugate prior. That is, we specify either
\[
\pi(\gamma, \sigma^2) \propto \frac{1}{\sigma^2} \tag{6}
\]
or
\[
\begin{align*}
\sigma^2 & \sim \text{InvGamma}(a, b) \\
\gamma | \sigma^2 & \sim N(\mu, \sigma^2 H^{-1}) \tag{7}
\end{align*}
\]
The uninformative prior implies that the reference point is the frequentist point estimator, which might be of interest to a frequentist researcher who might want to compute the posterior intervals for robustness check. On the other hand, the benefits of the conjugate prior include the facts that the researcher can incorporate prior belief on the regression coefficients or can impose some regularization and that the marginal likelihood is available conditional on \(\tau\). Intuitively, the two types of priors are equivalent when \(a = b = 0\) and \(H = 0\).

### 4.2 Posterior under the improper uninformative prior

With the improper uninformative prior \((6)\), we can show that the marginal posterior of the jump location \(\tau\) is
\[
\pi(\tau | D^n) \propto [\det(\chi_\tau \chi'_\tau)]^{-0.5} [S_n(\tau)]^{-(n-(d_x+d_z))} \times \pi(\tau) \tag{8}
\]
The marginal posterior of \(\gamma = (\beta, \delta)'\) is a mixture with the weights being the marginal posterior density \(\pi_n(\tau)\) of \(\tau\):
\[
p(\gamma | D^n) = \int_0^1 p(\gamma | \tau, D^n) \pi_n(\tau) d\tau
\]
where \(\pi_n(\tau)\) is the marginal posterior density of \(\tau\). Note that in the conventional frequentist approach, \(\tau\) is fixed at the point estimate \(\hat{\tau}\), and hence the information \(p(\gamma | \tau, D^n) \pi_n(\tau)\) for \(\tau \neq \hat{\tau}\) is essentially neglected. In contrast, such information or uncertainty is explicitly reflected in Bayesian framework. Conditional on \(\tau\), the posterior distributions of \(\gamma = (\beta, \delta)'\) is a \((d_x + d_z)\)-dimensional t-distribution centering around the OLS estimates given \(\tau\):
\[
\begin{align*}
\gamma | \tau, D^n & \sim t_{(d_x+d_z)} \left( n - (d_x+d_z), \hat{\gamma}(\tau), \frac{S_n(\tau)}{n - (d_x+d_z)} \left(\chi_\tau \chi'_\tau\right)^{-1} \right) \tag{9}
\end{align*}
\]
where \(t_k(v, \mu, \Sigma)\) is the \(k\)-dimensional t-distribution with \(v\) degrees of freedom, a location vector \(\mu \in \mathbb{R}^k\), and a \(k \times k\) shape matrix \(\Sigma\). It can be shown that the posterior for \(\sigma^2\) conditional on \(\tau\) is Inverse-Gamma.
with shape parameter \((n - (d_x + d_z))/2\) and scale parameter \(S_n(\tau)/2\):

\[
\sigma^2|\tau, D^n \sim \text{InvGamma} \left( (n - (d_x + d_z))/2, S_n(\tau)/2 \right) \tag{10}
\]

### 4.3 Posterior under the conjugate prior

With the conjugate prior (7), we can derive the posteriors for each of the parameters.

\[
\pi(\tau|D^n) \propto \left[\det(\bar{H}_\tau)\right]^{-0.5} \bar{b}_\tau^{-\bar{a}} \times \pi(\tau) \tag{11}
\]

\[
\gamma|\tau, D^n \sim t_{(d_x+d_z)} \left(2\bar{a}, \bar{b}_\tau, \bar{\mu}_\tau \right) \tag{12}
\]

\[
\sigma^2|\tau, D^n \sim \text{InvGamma} \left( \bar{a}, \bar{b}_\tau \right) \tag{13}
\]

where

\[
\begin{align*}
\bar{H}_\tau &= H + \chi_\tau \chi_\tau' \\
\bar{\mu}_\tau &= \bar{H}_\tau^{-1} \left[H\mu + \chi_\tau Y\right] \\
\bar{b}_\tau &= \bar{b} + \frac{1}{2} \left[\mu'\bar{H}_\tau \mu + \mu'Y - \mu'\bar{H}_\tau \mu\right] \\
\bar{a} &= a + \frac{n}{2}
\end{align*}
\tag{14}
\]

### 4.4 Posterior sampling

Due to the availability of the closed form conditional posterior for \(\beta, \delta\), and \(\sigma^2\) given \(\tau\), the posterior sampling is simple and fast. One can first draw \(\tau^{(1)}, \ldots, \tau^{(S)}\) from the marginal posterior of \(\tau\) as in (8) (or (11)) via for example Metropolis-Hastings algorithm. For each \(\tau^{(s)}\), one can sample posterior draws \(\gamma^{(s)}\) and \(\sigma^2^{(s)}\) from the posteriors conditional on \(\tau = \tau^{(s)}\), namely (9) (or (12)) and (10) (or (13)). A laptop with a 2.2GHz processor and 8GB RAM takes about 4.1 seconds to draw 10,000 posterior draws in the regression problem in Section 7 that has 10 slope coefficients in total.

### 5 Asymptotic theory

In this section, we investigate asymptotic behavior of the Bayesian method. Section 5.1 shows that the marginal posterior of the threshold parameter \(\tau\) contracts to the true value \(\tau_0\) at rate of \(n^{-1}\), the same rate at which the LS estimator \(\hat{\tau}\) converges. The proof is based on studying the behavior of the log ratio of the marginal posterior densities of \(\tau\). Section 5.2 establishes a Bernstein-von mises type theorem for the regression coefficients \(\gamma = (\beta, \delta)'\). The proof exploits the fact that the conditional posterior for
\( \sqrt{n} (\gamma - \gamma_0) \) given \( \tau \) is asymptotically normal, which is close to the asymptotic distribution of the OLS estimator \( \hat{\gamma} (\tau_0) \) when \( \tau \) is close to \( \tau_0 \). A bound on the KL divergence between two normal densities together with the \( n \)-consistency is used to make the argument precise.

### 5.1 \( n \)-consistency of marginal posterior of \( \tau \)

An intermediate step for proving Bernstein-von mises theorem is marginal posterior consistency of \( \tau \) at rate \( n^{-1} \). Marginal posteriors have not been studied extensively or systematically in the literature. Here, we directly analyze the form of the marginal posterior of \( \tau \). The marginal posterior density of \( \tau \) is defined as

\[
\pi_n(\tau) = \frac{\pi(\tau | Y^{(n)})}{\int \pi(\tau | Y^{(n)}) d\tau} = \frac{L_n(\tau)}{\int L_n(\tau) d\tau}
\]

where \( L_n(\tau) \) is the right hand side of the equation (8) or (11), depending on the choice of the prior. The following theorem states the main result of this subsection.

**Theorem 1** (Marginal posterior consistency of \( \tau \) at rate \( n^{-1} \)). Suppose Assumptions A1-A4 hold. Then, under both the uninformative improper prior and the conjugate prior, \( \forall \eta > 0, \epsilon > 0, \exists M > 0 \) and \( N > 0 \) such that \( n \geq N \implies \)

\[
P_{\theta_0} \left( \int_{B_{\Delta/n}^{\gamma}(\tau_0)} \pi_n(\tau) d\tau < \eta \right) > 1 - \epsilon
\]

where \( B_{\delta}(\tau_0) = (\tau_0 - \delta, \tau_0 + \delta) \).

Note that

\[
\pi_n(\tau) = \frac{L_n(\tau)}{\int L_n(\tau') d\tau'} = \frac{L_n(\tau_0)}{\int L_n(\tau') d\tau'} \frac{L_n(\tau)}{L_n(\tau_0)} = \pi_n(\tau_0) \frac{L_n(\tau)}{L_n(\tau_0)}
\]

\[
\pi_n(\tau_0) = \frac{L_n(\tau_0)}{\int L_n(\tau') d\tau'} \leq \frac{L_n(\tau_0)}{\int L_n(\tau') d\tau'} = \left[ \int_{B_{M_0/n}^{\gamma}(\tau_0)} \frac{L_n(\tau')}{L_n(\tau_0)} d\tau' \right]^{-1}
\]

for any \( M_0 > 0 \). Hence for each \( n \) and for any \( M_0 > 0 \),

\[
\int_{B_{\Delta/n}^{\gamma}(\tau_0)} \pi_n(\tau) d\tau = \pi_n(\tau_0) \int_{B_{M_0/n}^{\gamma}(\tau_0)} \frac{L_n(\tau)}{L_n(\tau_0)} d\tau \leq \left[ \int_{B_{M_0/n}^{\gamma}(\tau_0)} \frac{L_n(\tau')}{L_n(\tau_0)} d\tau' \right]^{-1} \int_{B_{M_0/n}^{\gamma}(\tau_0)} \frac{L_n(\tau)}{L_n(\tau_0)} d\tau \tag{15}
\]

Therefore, we want to find

1. an upper bound for \( \int_{B_{\Delta/n}^{\gamma}(\tau_0)} \frac{L_n(\tau)}{L_n(\tau_0)} d\tau \) and
2. a lower bound for \( \int_{B_{M_0/n}^{\gamma}(\tau_0)} \frac{L_n(\tau')}{L_n(\tau_0)} d\tau' \) for some \( M_0 > 0 \)
Note that
\[
\frac{L_n(\tau)}{L_n(\tau_0)} = \exp\left[n \left\{ \frac{1}{n} \log L_n(\tau) - \frac{1}{n} \log L_n(\tau_0) \right\} \right]
\]

Case 1: Improper uninformative prior
For the case of the improper uninformative prior, from (8), we have
\[
\frac{1}{n} \log L_n(\tau) - \frac{1}{n} \log L_n(\tau_0) = \frac{1}{2n} \log \left[ \frac{\det (\chi_\tau' \chi_\tau)}{\det (\chi_{\tau_0}' \chi_{\tau_0})} \right] + \frac{n - (d_x + d_z)}{2n} \left[ \log S_n(\tau_0) - \log S_n(\tau) \right] + \frac{1}{n} \log \frac{\pi(\tau)}{\pi(\tau_0)}
\]

Note that for example \( \frac{1}{n} \sum_{i=1}^{n} x_i x_i' + \frac{1}{n} \sum_{i=n+1}^{n+\nu} x_i z_i' \) converge in probability and by continuity of determinant, its determinant converges to the determinant of the moment matrix. Hence, the quantity inside of log in the first term is \( O_p(1) \). Therefore, the first term is \( O_p(n^{-1}) \). Since the ratio \( \pi(\tau)/\pi(\tau_0) \) is bounded, the last term is \( O(n^{-1}) \). Hence, we have
\[
\frac{1}{n} \log L_n(\tau) - \frac{1}{n} \log L_n(\tau_0) = \log S_n(\tau_0) - \log S_n(\tau) + O_p(n^{-1}) = \log \left[ \frac{1}{n} S_n(\tau_0) \right] - \log \left[ \frac{1}{n} S_n(\tau) \right] + O_p(n^{-1})
\]

Case 2: Conjugate prior
For the case of the conjugate prior, from (11), we have
\[
\frac{1}{n} \log L_n(\tau) - \frac{1}{n} \log L_n(\tau_0) = \frac{1}{2n} \log \left[ \frac{\det (\bar{H}^{-1})}{\det (H^{-1})} \right] + \frac{\bar{a}}{n} \left[ \log (\bar{b}_\tau) - \log (\bar{b}_r) \right] + \frac{1}{n} \log \frac{\pi(\tau)}{\pi(\tau_0)}
\]
\[
= \log \left( \frac{1}{n} b_\tau \right) - \log \left( \frac{1}{n} b_r \right) + O_p(n^{-1})
\]

We have
\[
\frac{1}{n} b_r = \frac{b}{n} + \frac{1}{2} \left[ \frac{1}{n} \mu' H \mu + \frac{1}{n} Y' Y - \frac{1}{n} \mu' \bar{H} \mu \right]
\]
\[
= \frac{1}{2n} \left[ Y' Y + (\chi_\tau' \chi_\tau)^{-1} (\chi_\tau' Y) \right] + O_p(n^{-1})
\]
\[
= \frac{1}{2n} S_n(\tau) + O_p(n^{-1})
\]

(16)
Hence, as in the case of the improper prior, even with the conjugate prior, we have that

\[
\frac{1}{n} \log L_n(\tau) - \frac{1}{n} \log L_n(\tau_0) = \log \left[ \frac{1}{n} S_n(\tau_0) \right] - \log \left[ \frac{1}{n} S_n(\tau) \right] + O_p(n^{-1})
\]

Define \(Q_n(\tau) = \frac{1}{n} S_n(\tau)\) and \(A_n(\tau) = g(Q_n(\tau))\) where \(g(x) = -\log(x)\). Then we can write

\[
\frac{1}{n} \log L_n(\tau) - \frac{1}{n} \log L_n(\tau_0) = A_n(\tau) - A_n(\tau_0) + O_p(n^{-1})
\]  \hspace{1cm} (17)

**Definition 1.** For all \(\tau \in \Theta\),

\[
Q(\tau) = \sigma_0^2 + \begin{cases} 
(\tau_0 - \tau) \frac{(1-\tau)}{(1-\tau_0)} \delta'_0 R' \Sigma X R \delta_0, & \text{if } \tau \leq \tau_0 \\
(\tau - \tau_0) \frac{\tau_0}{\tau} \delta'_0 R' \Sigma X R \delta_0, & \text{if } \tau > \tau_0
\end{cases}
\]

\(\equiv \sigma_0^2 + \Delta(\tau)\)

Figure below shows an example of \(Q_n(\tau)\) and \(Q(\tau)\).

**Figure 1:** Example of \(Q_n(\tau)\) \(n=100\) (+), \(n=1,000\) (circle), and \(n=10,000\) (cross) and \(Q(\tau)\) (dashed).

The following propositions are used to prove Theorem 1.
Proposition 1. Under $P_{\theta_0}$, for all $\tau$,

$$Q_n(\tau) = Q(\tau) + O_p(n^{-1/2})$$

Define $A(\tau) = g(Q(\tau))$.

Proposition 2. $A(\tau)$ attains its unique maximum at $\tau_0$

Proposition 3. \(\forall \eta > 0, \forall \epsilon > 0, \exists M > 0 \text{ such that } n \geq N \implies \) 

$$P_{\theta_0}\left(\inf_{\tau \in B_{M/n}(\tau_0)} \frac{|\{A_n(\tau) - A_n(\tau_0)\} - \{A(\tau) - A(\tau_0)\}|}{|\tau - \tau_0|} < \eta\right) > 1 - \epsilon$$

Proof of theorem 1. By Proposition 2, $A(\cdot)$ attains its unique max at $\tau_0$. Note that the convex function $A(\tau)$ is not differentiable at $\tau_0$. Hence we have,

$$A(\tau) - A(\tau_0) < |\tau - \tau_0|B_1$$

$$A(\tau) - A(\tau_0) > |\tau - \tau_0|B_2$$

for some $B_1, B_2 < 0$.

By Proposition 3, given $\eta_1 > 0, \exists M > 0 :$ with $P_{\theta_0} \to 1$,

$$A_n(\tau) - A_n(\tau_0) < \eta_1 |\tau - \tau_0| + A(\tau) - A(\tau_0) < |\tau - \tau_0|\{\eta_1 + B_1\} \quad (18)$$

for all $\tau \in B_{M/n}(\tau_0)$. Similarly, given $\eta_2 > 0, \exists M_0 > 0 :$ with $P_{\theta_0} \to 1$,

$$A_n(\tau) - A_n(\tau_0) > -\eta_2 |\tau - \tau_0| + A(\tau) - A(\tau_0) > |\tau - \tau_0|\{-\eta_2 + B_2\} \quad (19)$$

for all $\tau \in B_{M_0/n}(\tau_0)$. Recall, by Eq. (17), we have

$$\frac{L_n(\tau)}{L_n(\tau_0)} = \exp\left[n\left(A_n(\tau_0) - A_n(\tau)\right) + O_p(1)\right]$$

Hence, from Eq. (18), given $\eta_1 > 0$, small compared to $-B_1$, there is $B_1' < 0$, which is independent of $M$:

we have with $P_{\theta_0} \to 1$,

$$\frac{L_n(\tau)}{L_n(\tau_0)} \leq \exp\left[n|\tau - \tau_0|B_1' + O_p(1)\right] = \exp\left[n|\tau - \tau_0|B_1'O_p(1)\right] \quad (20)$$

for all $\tau \in B_{M/n}(\tau_0)$. Note that the statement above still holds with a larger value of $M > 0$ as the area outside of the ball will be contained by that for the original $M$. Similarly, from Eq. (19), there is $B_2' < 0$
and \( M_0 > 0 \) : with \( P_{\theta_0} \to 1, \)
\[
\frac{L_n(\tau)}{L_n(\tau_0)} \geq \exp\left[ n|\tau - \tau_0|B_2' + O_p(1) \right] = \exp\left[ n|\tau - \tau_0|B_2' \right] O_p(1)
\]
(21)
for all \( \tau \in B^c_{M_0/n}(\tau_0) \). Now, by Inequality (20) and fundamental theorem of calculus,
\[
\int_{B^c_{M/n}(\tau_0)} \frac{L_n(\tau)}{L_n(\tau_0)} d\tau \leq \int_{B^c_{M/n}(\tau_0)} \exp\left[ n|\tau - \tau_0|B_1' \right] d\tau O_p(1) = \frac{1}{nB_1'} \left( e^{nB_1'} - e^{B_1'M} \right) O_p(1)
\]
Similarly, by Inequality (21),
\[
\int_{B_{M_0/n}(\tau_0)} \frac{L_n(\tau)}{L_n(\tau_0)} d\tau \geq \int_{B_{M_0/n}(\tau_0)} \exp\left[ n|\tau - \tau_0|B_2' \right] d\tau O_p(1) = \frac{1}{nB_2'} \left( e^{nB_2'} - e^{B_2'M_0} \right) O_p(1)
\]
This means, together with the bound (15),
\[
\int_{B^c_{M/n}(\tau_0)} \pi_n(\tau)d\tau \leq \left[ \int_{B_{M_0/n}(\tau_0)} \frac{L_n(\tau')}{L_n(\tau_0)} d\tau' \right]^{-1} \int_{B^c_{M/n}(\tau_0)} \frac{L_n(\tau)}{L_n(\tau_0)} d\tau \leq \frac{B_2 e^{B_2'n} - e^{B_2'M_0}}{B_1 e^{B_2'n} - e^{B_2'M_0}} O_p(1)
\]
which can be made arbitrarily small by choosing \( M > 0 \) and \( n \) sufficiently large. 

5.2 Bernstein-von Mises Theorem for the regression coefficients
The marginal posterior of \((\beta, \delta)'\) is a mixture weighted by \(\pi_n(\tau)\). Furthermore, due to Theorem 1, we can focus our attention on the values of \(\tau\) in a \(n^{-1}\) neighborhood of \(\tau_0\):
\[
\int_0^1 p(\beta, \delta|\tau, D^n) \pi_n(\tau) d\tau = \int_{B_{M/n}(\tau_0)} p(\beta, \delta|\tau, D^n) \pi_n(\tau) d\tau + o_p(1)
\]
Conditional on \(\tau = \tau_0\), the standard result of OLS applies:
\[
\sqrt{n} (\hat{\tau}(\tau_0) - \tau_0) = \left[ \frac{\sqrt{n} (\hat{\beta}(\tau_0) - \beta_0)}{\sqrt{n} (\hat{\delta}(\tau_0) - \delta_0)} \right] \overset{d}{\to} N_{(d_{x+d_z})} (0, \sigma_0^2 V^{-1})
\]
(22)
where \( V \) is defined in (22). We are now ready to establish the following Bernstein-von mises type result:

\[
d_{TV} \left( \pi \begin{pmatrix} \sqrt{n} (\beta - \hat{\beta}(\tau_0)) \\ \sqrt{n} (\delta - \hat{\delta}(\tau_0)) \end{pmatrix} \bigg| D^n \right), N_{(d_{x+d_z})} (0, \sigma_0^2 V^{-1}) \to 0
\]
in \( P_{\theta_0} \) probability.
Proof of theorem 2. Define \( z = \sqrt{n} (\gamma - \hat{\gamma}(\tau_0)) = (\sqrt{n} (\beta - \hat{\beta}(\tau_0)), \sqrt{n} (\delta - \hat{\delta}(\tau_0)))' \).

\[
d_{TV} \left( \pi(z|D^n), N_{(d_x + d_z)}(0, V) \right) = \int |\pi(z|D^n) - \phi(z; 0, V)| dz \\
\leq \int \int |\pi(z|\tau, D^n) - \phi(z; 0, V)| dz d\pi(\tau|D^n) = \int d_{TV} \left( \pi(z|\tau, D^n), N_{(d_x + d_z)}(0, V) \right) d\pi(\tau|D^n) \\
= \int_{B_{M/n}(\tau_0)} d_{TV} \left( \pi(z|\tau, D^n), N_{(d_x + d_z)}(0, V) \right) d\pi(\tau|D^n) + o_p(1)
\]

where the last equality is due to theorem 1.

Case 1: Improper uninformative prior

Let us first consider the case with the improper uninformative prior (6). From (9), asymptotically, the posterior of \( \gamma = (\beta, \delta)' \) conditional on \( \tau \) is normal:

\[
\gamma|\tau, D^n \sim N_{(d_x + d_z)} \left( \hat{\gamma}(\tau), \frac{S_n(\tau)}{n - (d_x + d_z)} (\chi\tau\chi)' \right) \\
\implies z|\tau, D^n \sim N_{(d_x + d_z)} \left( \sqrt{n} (\hat{\gamma}(\tau) - \hat{\gamma}(\tau_0)), \frac{nS_n(\tau)}{n - (d_x + d_z)} (\chi\tau\chi)' \right)
\]

The total variation distance is bounded above by 2 times square root of the KL divergence. Note that for any \( p \)-dimensional normal distributions \( N(\mu_1, \Sigma_1), N(\mu_2, \Sigma_2) \), we have

\[
KL \left( N(\mu_1, \Sigma_1), N(\mu_2, \Sigma_2) \right) \leq \frac{\left| \text{det} (\Sigma_2^{-1}) - \text{det} (\Sigma_1^{-1}) \right|}{\min \left( \text{det} (\Sigma_1^{-1}), \text{det} (\Sigma_2^{-1}) \right)} + p \|\Sigma_2^{-1} - \Sigma_1^{-1}\|_\infty \|\Sigma_1\|_\infty + \|\mu_1 - \mu_2\|_2^2 \|\Sigma_2^{-1}\|_2
\]

where \( \|\Sigma\|_\infty = \max_{ij} |\Sigma_{ij}| \) is the largest element of \( \Sigma \) in the absolute value and \( \|\Sigma\|_2 = \sup_{\mu} \|\Sigma\mu\|_2/\|\mu\|_2 \) is a matrix norm induced by the standard norm on \( \mathbb{R}^p, \|\mu\|_2 = \sum_{i=1}^p \mu_i^2 \).

In our case, \( \mu_1 = \sqrt{n} (\hat{\gamma}(\tau) - \hat{\gamma}(\tau_0)), \mu_2 = 0, \Sigma_2 = \sigma_0^2 V^{-1} \), and

\[
\Sigma_1 = \frac{nS_n(\tau)}{n - (d_x + d_z)} (\chi\tau\chi)'^{-1} \\
= \frac{nQ_n(\tau)}{n - (d_x + d_z)} \left( \frac{1}{n} \chi\tau\chi \right)'^{-1} \\
\equiv V_n^{-1}(\tau)
\]
Note that since $|\tau - \tau_0| < \frac{M}{n}$,

\[
\Sigma_1 - \Sigma_2 = \frac{n}{n - (d_x + d_z)} \left( Q_n(\tau) - Q_n(\tau_0) \right) \hat{V}_n^{-1}(\tau) = o_p(1)
\]

\[
+ \frac{n}{n - (d_x + d_z)} Q_n(\tau_0) \left( \hat{V}_n^{-1}(\tau) - \hat{V}_n^{-1}(\tau_0) \right) = o_p(1)
\]

\[
+ \frac{n}{n - (d_x + d_z)} (Q_n(\tau_0) - \sigma_0^2) \hat{V}_n^{-1}(\tau_0) = O_p(n^{-1/2})
\]

\[
+ \frac{n}{n - (d_x + d_z)} \sigma_0^2 \left( \hat{V}_n^{-1}(\tau_0) - V^{-1} \right) = o_p(1)
\]

\[
+ \left( \frac{n}{n - (d_x + d_z)} - 1 \right) \sigma_0^2 V^{-1} = o_p(1)
\]

which implies

\[
\Sigma_2^{-1} - \Sigma_1^{-1} = o_p(1)
\]

Hence $II = o_p(1)$. By continuity of $det$ and $min$ functions, we also have that $I = o_p(1)$ for $\tau \in B_{M/n}(\tau_0)$.

Lastly, to show $III = o_p(1)$, note that $Y = X\beta_0 + Z_{2\tau_0}\delta_0 + \epsilon = X\beta_0 + Z_{2\tau}\delta_0 + \epsilon^*$, where $\epsilon^* = (Z_{2\tau_0} - Z_{2\tau})\delta_0 + \epsilon$. This implies

\[
\sqrt{n} \left( \hat{\beta}(\tau) - \beta_0 \right) = \sqrt{n} \left( \hat{\delta}(\tau) - \delta_0 \right) = \left[ \frac{1}{n} \begin{pmatrix} X'X & X'Z_{2\tau} \\ Z_{2\tau}'X & Z_{2\tau}'Z_{2\tau} \end{pmatrix} \right]^{-1} \frac{1}{\sqrt{n}} \begin{pmatrix} X'\epsilon^* \\ Z_{2\tau}'\epsilon^* \end{pmatrix}
\]

\[
= \left[ \frac{1}{n} \begin{pmatrix} X'X & X'Z_{2\tau} \\ Z_{2\tau}'X & Z_{2\tau}'Z_{2\tau} \end{pmatrix} \right]^{-1} \frac{1}{\sqrt{n}} \begin{pmatrix} X'\epsilon + X'(Z_{2\tau_0} - Z_{2\tau})\delta_0 \\ Z_{2\tau}'\epsilon + Z_{2\tau}'(Z_{2\tau_0} - Z_{2\tau})\delta_0 \end{pmatrix}
\]

It can be shown that for $|\tau - \tau_0| < \frac{M}{n}$,

\[
\frac{1}{\sqrt{n}} X'(Z_{2\tau_0} - Z_{2\tau}) = o_p(1)
\]

\[
\frac{1}{\sqrt{n}} Z_{2\tau_0}'(Z_{2\tau_0} - Z_{2\tau}) = o_p(1)
\]

\[
\frac{1}{n} X'Z_{2\tau} - \frac{1}{n} X'Z_{2\tau_0} = o_p(1)
\]

\[
\frac{1}{n} Z_{2\tau}'Z_{2\tau} - \frac{1}{n} Z_{2\tau_0}'Z_{2\tau_0} = o_p(1)
\]
which implies
\[
\sqrt{n} (\hat{\gamma}(\tau) - \gamma_0) = \left[ \frac{1}{n} \begin{pmatrix} X' X & X' Z_{2\tau_0} \\ Z'_{2\tau_0} X & Z'_{2\tau_0} Z_{2\tau_0} \end{pmatrix} \right]^{-1} \frac{1}{\sqrt{n}} \begin{pmatrix} X' \epsilon \\ Z'_{2\tau} \epsilon \end{pmatrix} + o_p(1)
\]

By central limit theorem, it can be shown that the left hand side above has the same limit distribution as that of \( \sqrt{n} (\hat{\gamma}(\tau_0) - \gamma_0) \) for \( |\tau - \tau_0| < \frac{M}{n} \). This means that \( III = o_p(1) \). Finally,
\[
d_{TV} \left( \pi(z|\tau,Y^{(n)}), N_{(d_z+d_z)}(0,V) \right) \leq 2\sqrt{o_p(1)} = o_p(1)
\]

where we know that the last term is \( o_p(1) \) from the proof for the case of improper prior. By definition,
\[
\bar{\mu}_\tau = \left[ \frac{1}{n} \bar{H} + \frac{1}{n} \chi_\tau \chi_\tau' \right]^{-1} \left[ \frac{1}{n} \bar{H} \mu + \frac{1}{n} \chi_\tau Y \right] = \hat{\gamma}(\tau) + O_p(n^{-1})
\]

so the first term in (23) is also \( o_p(1) \). Also note that for each \( \tau \in B_{M/n}(\tau_0) \)
\[
(n\bar{b}_\tau/\bar{a})\bar{H}_\tau^{-1} - \sigma^2_0 V = \left[ (n\bar{b}_\tau/\bar{a})\bar{H}_\tau^{-1} - \frac{nS_n(\tau)}{n - (d_x + d_z)(\chi_\tau\chi_\tau)} \right] + \left[ \frac{nS_n(\tau)}{n - (d_x + d_z)(\chi_\tau\chi_\tau)^{-1}} - \sigma^2_0 V \right]
\]

where we know that the second term is \( o_p(1) \) from the proof for the case of improper prior. For the first
term, we have

$$(n \bar{b}_r / \bar{a}) \bar{H}_r^{-1} = \frac{b_r}{\bar{a} + n/2} \left[ \frac{1}{n} H + \frac{1}{n} \chi \chi' \right]^{-1}$$

$$= \frac{\frac{1}{n} b_r}{\bar{a}/n + 1/2} \left[ \frac{1}{n} H + \frac{1}{n} \chi \chi' \right]^{-1}$$

From (16), we know that $\frac{1}{n} b_r = \frac{1}{2n} S_n(\tau) + O_p(n^{-1})$, so we have

$$(n \bar{b}_r / \bar{a}) \bar{H}_r^{-1} = S_n(\tau) \left[ \chi \chi' \right]^{-1} + O_p(n^{-1})$$

Therefore, the first term in (24) is also $o_p(1)$. The same argument from the proof for the improper prior implies the desired result.

6 Simulation

6.1 Comparison with the conventional frequentist method

Consider the following simple model: $y_i = \delta_0 1(i > \lfloor n \tau_0 \rfloor) + \epsilon_i$ with $\epsilon_i \sim N(0, 1)$. Assume $\tau_0 = 0.5$. We consider different values of sample size $n = 50, 100, 250, $ and $500$ and the true jump size $\delta_0 = 0.25, 0.5, $ and $1.0$.

We study the behavior of Bayesian approach in repeated experiments. For each pair of $n$ and $\delta_0$, we generate 1,000 data sets. We report the coverage of the 95% credible interval for the jump size $\delta$ (Table 1), the mean interval length (Table 3), and the mean absolute value of the difference between the posterior mode of $\delta$ and $\delta_0$ (Table 2).

We also present the performance of the conventional frequentist estimator. In specific, we computed the least-square estimator $\hat{\tau}$, the jump size estimator $\hat{\delta}(\hat{\tau})$ conditional on $\tau = \hat{\tau}$, and its 95% confidence interval based on the conventional asymptotic theory. This would what the researchers do based on the method in Hansen (2000). As mentioned in introduction section, he proposes a Bonferroni-type correction of CIs, but the tuning parameter has to be selected, so we do not present Bonferroni-type corrected CIs.

1. Table 1 shows that for small $n$ and/or small $\delta_0$, the conventional CI significantly under-covers. On the other hand, Bayesian credible interval has a relatively reasonable coverage.

2. From Table 2, we can see that except for a few cases, the conventional method and Bayesian method provide similar bias.
<table>
<thead>
<tr>
<th>( \delta_0 = )</th>
<th>Least-squares</th>
<th>Bayesian</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>0.50</td>
<td>1.00</td>
</tr>
<tr>
<td>0.25</td>
<td>0.50</td>
<td>1.00</td>
</tr>
</tbody>
</table>

n = 50 | 0.63 | 0.83 | 0.95 |
| 0.97 | 0.97 | 0.97 |

n = 100 | 0.71 | 0.91 | 0.95 |
| 0.96 | 0.96 | 0.96 |

n = 250 | 0.82 | 0.94 | 0.95 |
| 0.96 | 0.96 | 0.95 |

n = 500 | 0.91 | 0.95 | 0.95 |
| 0.95 | 0.96 | 0.95 |

Table 1: Coverage of the conventional 95% confidence interval (left) and the 95% credible interval (right).

<table>
<thead>
<tr>
<th>( \delta_0 = )</th>
<th>Least-squares</th>
<th>Bayesian</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>0.50</td>
<td>1.00</td>
</tr>
<tr>
<td>0.25</td>
<td>0.50</td>
<td>1.00</td>
</tr>
</tbody>
</table>

n = 50 | 0.56 | 0.37 | 0.23 |
| 0.3 | 0.28 | 0.24 |

n = 100 | 0.35 | 0.2 | 0.16 |
| 0.21 | 0.19 | 0.16 |

n = 250 | 0.18 | 0.1 | 0.1 |
| 0.13 | 0.11 | 0.1 |

n = 500 | 0.09 | 0.07 | 0.07 |
| 0.08 | 0.07 | 0.07 |

Table 2: |Bias| of \( \hat{\delta}(\hat{\tau}) \) and the posterior mode of \( \delta \)

3. From Table 3, we can see that the primary reason for the under-coverage of the confidence intervals seems to be that the interval tends to be too short. On the other hand, credible intervals provide a better coverage since they are wider, which would be the result of correctly reflecting the sampling uncertainty of threshold estimation.

4. Finally, looking at Tables 1, 2, and 3, we see that as \( n \) increases, the discrepancy between the Bayesian and the frequentist results decreases, as expected from the proven Bernstein-von mises theorem from the previous section.
<table>
<thead>
<tr>
<th>$\delta_0$</th>
<th>Least-squares</th>
<th>Bayesian</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.25 0.50 1.00</td>
<td>0.25 0.50 1.00</td>
</tr>
<tr>
<td>$n = 50$</td>
<td>1.31 1.25 1.12</td>
<td>1.64 1.57 1.35</td>
</tr>
<tr>
<td>$n = 100$</td>
<td>0.92 0.84 0.79</td>
<td>1.14 1.02 0.85</td>
</tr>
<tr>
<td>$n = 250$</td>
<td>0.56 0.51 0.49</td>
<td>0.68 0.56 0.5</td>
</tr>
<tr>
<td>$n = 500$</td>
<td>0.38 0.35 0.35</td>
<td>0.44 0.36 0.35</td>
</tr>
</tbody>
</table>

Table 3: Length of the conventional 95% confidence interval (left) and the 95% credible interval (right).

6.2 Comparison to the sample-splitting method [In progress]

7 Application

In this section, we consider two empirical applications: economic growth model in Durlauf and Johnson (1995) and stock return prediction in Paye and Timmermann (2006). In both papers, the estimation is done using the conventional frequentist approach: they first computed $\hat{\tau}$ and then estimated the regression coefficients by fixing $\tau$ at $\hat{\tau}$.

7.1 Growth and multiple equilibria: Durlauf and Johnson (1995)

In this section, we apply our method on the data from Durlauf and Johnson (1995) which is also considered in Hansen (2000). The authors suggest that cross-section growth behavior may be determined by initial conditions. They explore this hypothesis using the Summers-Heston data set, reporting results obtained from a regression tree approach, which is an extension of threshold regressions.

In one of the specifications in their paper, the authors divide the countries into two groups by the literacy rate in the base year:

$$\ln\left(\frac{Y}{L}\right)_{i,1985} - \ln\left(\frac{Y}{L}\right)_{i,1960} = \begin{cases} \delta_1^{(1)} + \delta_1^{(2)} \ln(Y/L)_{i,1960} + \delta_1^{(3)} \ln(I/Y)_{i} + \delta_1^{(4)} \ln(n_i + g + \delta) + \delta_1^{(5)} \ln(school)_{i} + \epsilon_i, & \text{if } Lit_{i,1960} \leq \tau \\ \delta_2^{(1)} + \delta_2^{(2)} \ln(Y/L)_{i,1960} + \delta_2^{(3)} \ln(I/Y)_{i} + \delta_2^{(4)} \ln(n_i + g + \delta) + \delta_2^{(5)} \ln(school)_{i} + \epsilon_i, & \text{if } Lit_{i,1960} > \tau \end{cases}$$

where for each country $i$, $(Y/L)_{i,t}$ is real GDP per member of the population aged 15-64 in year $t$, $(I/Y)_{i}$ is investment to GDP ratio, $n_i$ is growth rate of working-age population, $school_i$ is fraction of working-age population enrolled in secondary school, and $Lit_{i,1960}$ is the literacy rate in the base year 1960. The variables not indexed by $t$ are annual averages over the period 1960-1985. Following the authors, we set $g + \delta = 0.05$, where $g$ is the growth rate of technology and $\delta$ is the depreciation rate of both human and physical capitals. We standardize the covariates to make the comparison of the estimated coefficients
easier.

The prior on $\tau$ that we use is $Uniform(0.05, 0.95)$. The posterior mean of $\tau$ is 53.7% with 95% credible interval being [31% 68%]. See Figure 2 for the trace plot and the posterior density. The Least Squares method produces a similar point estimate, which is $\hat{\tau} = 53\%$.

In Figure 3, we display the posterior densities of the slope coefficients as well as 95% credible intervals. The intervals in red are the confidence intervals computed assuming the point estimate $\hat{\tau}$ being fixed which would be what the conventional frequentist method would produce. We see that in general, as it was expected, the Bayesian credible intervals are wider than the frequentist counterparts, which would be a result of correctly reflecting the finite sample uncertainty of the unknown $\tau$. Importantly, this can have a qualitative consequence on statistical importance of some parameter(s). For example, for the slope coefficients $\delta_1^{(2)}$ on $\ln(Y/L)_i, 1960$ and $\delta_1^{(3)}$ on $\ln(I/Y)_i$, the confidence interval does not include 0 while the Bayesian credible interval does. Hence, it is possible that ignoring the finite sample uncertainty of the threshold estimation stage could lead to an invalid inference.

Figure 2: Trace plots (left) and posterior density (right) for $\tau$
Figure 3: Posterior density of $\beta$’s. 95% credible interval (blue) and 95% confidence interval assuming $\hat{\tau}$ (red).

Figure 4: Posterior density of $\sigma$

7.2 Stock return prediction: Paye and Timmermann (2006) [In progress]

Paye and Timmermann (2006) investigates the instability in models of ex-post predictable components in stock returns. The model that they consider is:

$$\text{Ret}_t = \begin{cases} 
\delta^{(1)}_1 + \delta^{(2)}_1 \text{Div}_{t-1} + \delta^{(3)}_1 \text{Tbill}_{t-1} + \delta^{(4)}_1 \text{Spread}_{t-1} + \delta^{(5)}_1 \text{Def}_{t-1} + \epsilon_t, & \text{if } t \leq \lfloor T\tau \rfloor \\
\delta^{(1)}_2 + \delta^{(2)}_2 \text{Div}_{t-1} + \delta^{(3)}_2 \text{Tbill}_{t-1} + \delta^{(4)}_2 \text{Spread}_{t-1} + \delta^{(5)}_2 \text{Def}_{t-1} + \epsilon_t, & \text{if } t > \lfloor T\tau \rfloor 
\end{cases}$$

where $\text{Ret}_t$ denotes the excess stock return during month $t$, $\text{Div}_{t-1}$ is the lagged dividend yield, $\text{Tbill}_{t-1}$ is the lagged short term interest rate, $\text{Spread}_{t-1}$ is the lagged spread between the short term and the long term interest rates, and $\text{Def}_{t-1}$ is the lagged U.S. default premium.
8 Conclusion and future direction

In this paper, we established a Bernstein von-mises type theorem for the regression coefficients in linear regression models with a structural break. A frequentist researcher can look at Bayesian credible intervals for the regression coefficients as a robustness check to see whether the finite sample uncertainty coming from the break location estimation affects the inference on the regression coefficients. Such sensitivity analysis is natural as our theoretical result guarantees the credible interval to converge to the conventional confidence intervals that the frequentist researcher would use otherwise.

Extending the result to multiple number of breaks should be straightforward. When the number of breaks is unknown, the researcher can either conduct model comparisons based on marginal likelihoods, which are available under the conjugate prior, or can compute the posterior probabilities of the number of breaks using methods such as reversible jump MCMC or dynamic programing techniques.

I am currently working on several types of extensions of this paper. First, the homoscedasticity assumption could be too strong in some applications, and hence extending the results to the case of heteroscedasticity would be of interest. Second, researchers might want to relax the linearity assumption for the regression function in some cases. I am currently working on Bayesian non-parametric regressions with structural breaks.

A Proof of Propositions

A.1 Proof of Proposition 1

**Proposition 1.** Under $P_{\theta_0}$, for all $\tau$,

$$Q_n(\tau) = Q(\tau) + O_p(n^{-1/2})$$

**Proof of Proposition 1.** Let $\tau \in (0, 1)$ be given. Let $M = I - X(X'X)^{-1}X'$. We have the following identity: $S_n(\tau) = \bar{S}_n - V_n(\tau)$ (Amemiya (1985), Bai (1997)), where $\bar{S}_n$ is the sum of squared residuals from regressing $Y$ on $X$ alone and $V_n(\tau) = \tilde{\delta}(\tau)(Z'_{2\tau}MZ_{2\tau})\tilde{\delta}(\tau)$. By Frisch-Waugh Theorem, the OLS estimate of $\delta$ in Eq. (3) is equivalent to that in the model $MY = MZ_{2\tau}\delta + M\epsilon$. Note the true model is $MY = MZ_{2\tau_0}\delta_0 + M\epsilon$. Hence,

$$\tilde{\delta}(\tau) = (Z'_{2\tau}MZ_{2\tau})^{-1}Z'_{2\tau}MY$$

$$= (Z'_{2\tau}MZ_{2\tau})^{-1}Z'_{2\tau}MZ_{2\tau_0}\delta_0 + (Z'_{2\tau}MZ_{2\tau})^{-1}Z'_{2\tau}M\epsilon$$

$$= (Z'_{2\tau}MZ_{2\tau})^{-1}Z'_{2\tau}MZ_{2\tau_0}\delta_0 + (Z'_{2\tau}MZ_{2\tau})^{-1}Z'_{2\tau}M\epsilon$$
Hence,

\[ V_n(\tau) = \delta_0'(Z'_{2\tau}MZ_{2\tau_0})'(Z'_{2\tau}MZ_{2\tau})^{-1}(Z'_{2\tau}MZ_{2\tau_0})\delta_0 \\
+ 2\delta_0'(Z'_{2\tau}MZ_{2\tau_0})'(Z'_{2\tau}MZ_{2\tau})^{-1}(Z'_{2\tau}M\epsilon) \\
+ (Z'_{2\tau}M\epsilon)'(Z'_{2\tau}MZ_{2\tau})^{-1}(Z'_{2\tau}M\epsilon) \]

This means that

\[ \frac{1}{n} V_n(\tau) = \frac{1}{n} \delta_0'(Z'_{2\tau}MZ_{2\tau_0})'(Z'_{2\tau}MZ_{2\tau})^{-1}(Z'_{2\tau}MZ_{2\tau_0})\delta_0 + O_p(n^{-1/2}) \]

Also we have

\[ \bar{S}_n(\tau) = Y'MY = \delta_0'Z'_{2\tau_0}MZ_{2\tau_0}\delta_0 + 2\delta_0Z_{2\tau_0}M + \epsilon'M\epsilon \]

which implies

\[ \frac{1}{n} \bar{S}_n(\tau) = \frac{1}{n} \delta_0'Z'_{2\tau_0}MZ_{2\tau_0}\delta_0 + \sigma_0^2 + O_p(n^{-1/2}) \]

By the above identity,

\[ Q_n(\tau) = \frac{1}{n} \bar{S}_n(\tau) - \frac{1}{n} V_n(\tau) \]

\[ = \sigma_0^2 + \frac{1}{n} \delta_0' \left\{ (Z'_{2\tau_0}MZ_{2\tau_0}) - (Z'_{2\tau}MZ_{2\tau_0})'(Z'_{2\tau}MZ_{2\tau})^{-1}(Z'_{2\tau}MZ_{2\tau_0}) \right\} \delta_0 + O_p(n^{-1/2}) \]

Note that

\[ (Z'_{2\tau_0}MZ_{2\tau_0}) = Z'_{2\tau_0}Z_{2\tau_0} - Z'_{2\tau_0}X(XX')^{-1}X'Z_{2\tau_0} \]

\[ = R'X'_{2\tau_0}X_{2\tau_0} - R'(X'_{2\tau_0}X)(XX')^{-1}(X'X_{2\tau_0})R \]

\[ = R'X'_{2\tau_0}X_{2\tau_0} - R'(X'_{2\tau_0}X_{2\tau_0})(XX')^{-1}(X'_{2\tau_0}X_{2\tau_0})R \]

Hence,

\[ \frac{1}{n} (Z'_{2\tau_0}MZ_{2\tau_0}) = (1 - \tau_0)R'S_XR - (1 - \tau_0)^2 R'S_XR + O_p(n^{-1/2}) \]

\[ = \tau_0(1 - \tau_0)R'S_XR + O_p(n^{-1/2}) \]

Similarly,

\[ \frac{1}{n} (Z'_{2\tau}MZ_{2\tau}) = \tau(1 - \tau)R'S_XR + O_p(n^{-1/2}) \]
Proof of Proposition 2.

\[ (Z'_{2\tau}MZ_{2\tau_0}) = Z'_{2\tau}Z_{2\tau_0} - Z'_{2\tau}X(X'X)^{-1}X'Z_{2\tau_0} \]
\[ = R'X'_{2\tau}X_{2\tau_0}R - R'(X'_{2\tau}X)(X'X)^{-1}(X'X_{2\tau_0})R \]
\[ = R'X'_{2\tau_0}X_{2\tau_0}R - R'(X'_{2\tau}X_{2\tau})(X'X)^{-1}(X'_{2\tau_0}X_{2\tau_0})R \]

which implies that
\[
\frac{1}{n}(Z'_{2\tau}MZ_{2\tau_0}) = (1 - \tau_0)R'S_XR + (1 - \tau)(1 - \tau_0)R'S_XR + O_p(n^{-1/2})
\]
\[ = \tau(1 - \tau_0)R'S_XR + O_p(n^{-1/2}) \]

Therefore,
\[
\frac{1}{n}(Z'_{2\tau}MZ_{2\tau_0})(Z'_{2\tau}MZ_{2\tau})^{-1}(Z'_{2\tau}MZ_{2\tau_0}) = \left[ \frac{(1 - \tau_0)}{\tau(1 - \tau)(1 - \tau_0)} \right] R'S_XR + O_p(n^{-1/2})
\]
\[ = \frac{\tau(1 - \tau_0)^2}{1 - \tau} \]

Finally,
\[
\frac{1}{n}\left\{ (Z'_{2\tau_0}MZ_{2\tau_0}) - (Z'_{2\tau}MZ_{2\tau_0})(Z'_{2\tau}MZ_{2\tau})^{-1}(Z'_{2\tau}MZ_{2\tau_0}) \right\}
\]
\[ = \left[ \tau_0(1 - \tau_0) - \frac{\tau(1 - \tau_0)^2}{1 - \tau} \right] R'S_XR + O_p(n^{-1/2})
\]
\[ = (\tau_0 - \tau)\frac{1 - \tau_0}{1 - \tau} R'S_XR + O_p(n^{-1/2}) \]

\[ \square \]

A.2 Proof of Proposition 2

Proposition 2. A(\tau) attains its unique maximum at \tau_0

Proof of Proposition 2. By definition, \( Q(\tau_0) = \sigma_0^2 \). Note that \( \delta_0^2 R'S_XR \delta_0 > 0 \). This is because (1) \( R \) has full column rank, (2) \( \delta_0 \neq 0 \), and (3) \( \Sigma_X \) is assumed to be positive definite. Hence, \( \Sigma_X > \sigma_0^2 \) \( \forall \tau \neq \tau_0 \). Recall that \( A(\tau) = g(Q(\tau)) \) where \( g(x) = -log(x) \). Hence \( A(\tau) = -log(\sigma_0^2) \) if \( \tau = \tau_0 \) and \( A(\tau) < -log(\sigma_0^2) \) otherwise. \( \square \)

A.3 Proof of Proposition 3

Proposition 3. \( \forall \eta > 0, \forall \epsilon > 0, \exists M > 0 \) and \( N > 0 \) such that \( n \geq N \implies \)
\[
P_{\theta_0}\left( \inf_{\tau \in B_{nM}^\tau(\tau_0)} \left\{ \frac{|A_n(\tau) - A_n(\tau_0)|}{|\tau - \tau_0|} \right\} < \eta \right) > 1 - \epsilon
\]

25
Proof of Proposition 3. Recall that $A_n(\tau) = g(Q_n(\tau))$ and $A(\tau) = g(Q(\tau))$ where $g(x) = -\log(x)$. By Taylor approximation, there is $c$ between $x$ and $a$: 

$$g(x) - g(a) = g'(a)(x - a) + \frac{1}{2}g''(c)(x - a)^2$$

Hence, for each $\tau \in B^c_{M/n}(\tau_0)$, there is $c_n$ between $Q_n(\tau)$ and $Q(\tau)$:

$$g(Q_n(\tau)) - g(Q(\tau)) = g'(Q(\tau)) (Q_n(\tau) - Q(\tau)) + \frac{1}{2}g''(c_n) (Q_n(\tau) - Q(\tau))^2$$

$$= g'(Q(\tau)) O_p(n^{-1/2}) + O_p(n^{-1})$$

where we used Proposition 1. Similarly, there is $c_{0n}$ between $Q_n(\tau_0)$ and $Q(\tau_0)$:

$$g(Q_n(\tau_0)) - g(Q(\tau_0)) = g'(Q(\tau_0)) O_p(n^{-1/2}) + O_p(n^{-1})$$

Note that $g''(c_n) = \frac{1}{c_n^3}$ and $g''(c_{0n}) = \frac{1}{c_{0n}^3}$ are bounded with probability tending to one because for each $\tau$, $Q_n(\tau) \overset{p}{\rightarrow} Q(\tau)$, and $Q(\tau)$ is bounded.

Now,

$$\{A_n(\tau) - A_n(\tau_0)\} - \{A(\tau) - A(\tau_0)\} = \{A_n(\tau) - A(\tau)\} - \{A_n(\tau_0) - A(\tau_0)\}$$

$$= \left\{g(Q_n(\tau)) - g(Q(\tau))\right\} - \left\{g(Q_n(\tau_0)) - g(Q(\tau_0))\right\}$$

$$= \left\{g'(Q(\tau)) - g'(Q(\tau_0))\right\} O_p(n^{-1/2}) + O_p(n^{-1})$$

$$= \left[ -\frac{1}{\sigma_0^2} + \Delta(\tau) \right] \left(-\frac{1}{\sigma_0^2 + \Delta(\tau_0)}\right) O_p(n^{-1/2}) + O_p(n^{-1})$$

$$= \left[ \frac{1}{\sigma_0^2} - \frac{1}{\sigma_0^2 + \Delta(\tau)} \right] O_p(n^{-1/2}) + O_p(n^{-1})$$

In general, there is $B > 0$ such that $\frac{1}{b} - \frac{1}{b+x} \leq Bx$ for $b, x > 0$. Hence, $\frac{1}{\sigma_0^2} - \frac{1}{\sigma_0^2 + \Delta(\tau)} \leq B\Delta(\tau) \leq B'|\tau - \tau_0|$ where the last inequality holds for some $B' > 0$ due to the shape of $Q(\tau)$.

Finally,

$$\frac{|\{A_n(\tau) - A_n(\tau_0)\} - \{A(\tau) - A(\tau_0)\}|}{|\tau - \tau_0|} \leq B' O_p(n^{-1/2}) + \frac{1}{|\tau - \tau_0|} O_p(n^{-1}) \leq O_p(n^{-1/2}) + \frac{1}{M} O_p(1)$$

The desired result is established by taking $M$ large enough.
B Derivation of Posterior distributions

B.1 Improper uninformative prior

In this section, we derive the posterior in Section 4.2 under the improper uninformative prior (6). The posterior is

\[
p(\theta|D^n) \propto \left(\frac{1}{\sigma}\right)^n \exp\left[ -\frac{1}{2\sigma^2} \left\{ \sum_{i=1}^{n} (y_i - \chi'_\tau,\gamma)^2 \right\} \right] \frac{1}{\sigma^2} \pi(\tau)
\]

\[
\propto \left(\frac{1}{\sigma}\right)^n \exp\left[ -\frac{1}{2\sigma^2} \left\{ S_n(\tau) + (\gamma - \hat{\gamma}(\tau))^\prime \chi'_\tau (\gamma - \hat{\gamma}(\tau)) \right\} \right] \frac{1}{\sigma^2} \pi(\tau)
\]

(25)

where \(S_n(\tau)\) is the sum of squared residuals given \(\tau\). Integrating the right hand side with respect to \(\gamma\), we obtain

\[
\left(\frac{1}{\sigma^2}\right)^{n-(d_x+d_z)+1} \exp\left[ -\frac{1}{2\sigma^2} S_n(\tau) \right] [\det (\chi'_\tau)]^{-0.5} \pi(\tau)
\]

by the property of the multivariate normal density. Integrating the above with respect to \(\sigma^2\) over the positive part of the real line and using the change of variable \(\phi = 1/\sigma^2\), we get the marginal posterior for \(\tau\)

\[
\pi(\tau|D^n) \propto [\det (\chi'_\tau)]^{-0.5} [S_n(\tau)]^{-(n-(d_x+d_z))/2} \times \pi(\tau)
\]

To obtain the conditional posterior for \(\gamma\) given \(\tau\), we integrate (25) with respect to \(\sigma^2\). With the change of variable \(\phi = 1/\sigma^2\), we can show that

\[
\pi(\gamma|\tau, D^n) \propto \left[ 1 + \frac{1}{n-(d_x+d_z)} (\gamma - \hat{\gamma}(\tau))^\prime \frac{1}{S_n(\tau)/(n-(d_x+d_z))} \chi'_\tau (\gamma - \hat{\gamma}(\tau)) \right]^{-n/2}
\]

which means that

\[
\gamma|\tau, D^n \sim t_{(d_x+d_z)} \left( n - (d_x + d_z), \hat{\gamma}(\tau), \frac{S_n(\tau)}{n-(d_x+d_z)} (\chi'_\tau)^{-1} \right)
\]

Finally, to obtain the conditional posterior for \(\sigma^2\) given \(\tau\), we integrate (25) with respect to \(\gamma\) and can show that

\[
\pi(\phi|\tau, D^n) \propto \phi^{n-(d_x+d_z)+1} \exp\left[ -\frac{S_n(\tau)}{2\phi} \right]
\]

which means that \(\phi|\tau, D^n \sim Gamma((n - (d_x + d_z))/2, S_n(\tau)/2)\) or equivalently,

\[
\sigma^2|\tau, D^n \sim InvGamma((n - (d_x + d_z))/2, S_n(\tau)/2)
\]
B.2 Conjugate prior

In this section, we derive the posterior in Section 4.3 under the conjugate prior (7). It can be shown that the posterior is

\[
p(\theta | D^n) \propto p(D^n | \theta) \pi(\theta)
\]

\[
\propto \left(\frac{1}{\sigma}\right)^n \exp\left[-\frac{1}{2\sigma^2} \left\{ \sum_{i=1}^{n} (y_i - \chi'_{\tau,i} \gamma)^2 \right\}\right] \frac{1}{\sigma^2} \pi(\tau)
\]

\[
\times \left(\frac{1}{\sigma^2}\right)^{\bar{a}+p/2+1} \exp\left[-\frac{1}{\sigma^2} \left\{ \bar{b} + \frac{1}{2} (\gamma - \mu)' H (\gamma - \mu) \right\}\right]
\]

\[
\propto \left(\frac{1}{\sigma^2}\right)^{\bar{a}+p/2+1} \exp\left[-\frac{1}{\sigma^2} \left\{ \bar{b}_\tau + \frac{1}{2} (\gamma - \bar{\mu}_\tau)' \bar{H}_\tau (\gamma - \bar{\mu}_\tau) \right\}\right]
\]

where \( \bar{a}, \bar{b}_\tau, \bar{\mu}_\tau, \) and \( \bar{H}_\tau \) are defined in Eq. (14). Note that the distribution for \( \gamma, \sigma^2 | \tau, D^n \) is also a normal-inverse-gamma distribution. By the same approach we used under the improper prior, we can show

\[
\pi(\tau | D^n) \propto \left[ \text{det} \left( \bar{H}_\tau \right) \right]^{-0.5} \bar{b}_\tau^{-\bar{a}} \times \pi(\tau)
\]

and

\[
\sigma^2 | \tau, D^n \sim \text{InvGamma} (\bar{a}, \bar{b}_\tau)
\]

Finally, to derive the posterior of \( \gamma \) conditional on \( \tau \), we use the well-known property that the integral of a normal-inverse-gamma distribution with respect to \( \sigma^2 \) is a t-distribution to conclude that

\[
\gamma | \tau, D^n \sim t_{(d_x + d_z)} (2\bar{a}, \bar{\mu}_\tau, (\bar{b}_\tau/\bar{a}) \bar{H}_\tau^{-1})
\]
Acknowledgements

I am grateful for Andriy Norets, my dissertation advisor, for guidance and encouragement through the work on this project. I thank for valuable comments and suggestions from Eric Renault, Susanne Schen- nach, Jesse Shapiro, Kenneth Chay, Toru Kitagawa, Adam McCloskey, Dimitris Korobilis, and Florian Gunsilius as well as participants in the Brown University Econometrics Seminar. This work was supported by the Economics department dissertation fellowship at Brown University. All remaining errors are mine.

References


