ISSN 1440-771X

MONASH University

Australia

Department of Econometrics and Business Statistics

http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/

Bayesian Bandwidth Estimation in Nonparametric Time-Varying Coefficient Models

Tingting Cheng, Jiti Gao and Xibin Zhang

February 2015

Working Paper 03/15
(revised 07/13)

Bayesian Bandwidth Estimation in Nonparametric Time–Varying Coefficient Models [1]

Tingting Cheng, Jiti Gao,[2] Xibin Zhang

Department of Econometrics and Business Statistics, Monash University, Australia

Abstract: Bandwidth plays an important role in determining the performance of nonparametric estimators, such as the local constant estimator. In this paper, we propose a Bayesian approach to bandwidth estimation for local constant estimators of time–varying coefficients in time series models. We establish a large sample theory for the proposed bandwidth estimator and Bayesian estimators of the unknown parameters involved in the error density. A Monte Carlo simulation study shows that (i) the proposed Bayesian estimators for bandwidth and parameters in the error density have satisfactory finite sample performance; and (ii) our proposed Bayesian approach achieves better performance in estimating the bandwidths than the normal reference rule and cross–validation. Moreover, we apply our proposed Bayesian bandwidth estimation method for the time–varying coefficient models that explain Okun's law and the relationship between consumption growth and income growth in the US. For each model, we also provide calibrated parametric forms of the time–varying coefficients.

Key words: Local constant estimator, bandwidth, Markov chain Monte Carlo.

JEL Classification: C11, C14, C15

[1] We thank participants at several seminars and conferences, during which earlier versions of this paper were presented, for their constructive comments and suggestions, particularly to Zongwu Cai, David Frazier, Han Hong, Maxwell King, Robert Kohn, Gael Martin, Adrian Pagan, Anastasios Panagiotelis, Peter Phillips, Peter Robinson and Farshid Vahid. This research is supported under the Australian Research Council's *Discovery Projects Scheme* (under Grant Numbers: DP1095838, DP1096374 and DP130104229).

[2] *Correspondence*: Jiti Gao, Department of Econometrics and Business Statistics, Monash University, Caulfield East, Victoria 3145, Australia. Telephone: +61 3 99031675. Fax: +61 3 99032007. Email: jiti.gao@monash.edu.

1 Introduction

Time–varying coefficient time series models have attracted much attention of econometricians and statisticians during the last two decades, following the publication of Robinson (1989). Recent studies that are most relevant to our work include Gao and Hawthorne (2006) for a semiparametric modeling of a temperature trend function, Cai (2007) for a nonparametric trending time series model, Li, Chen and Gao (2011), and Chen, Gao and Li (2012) for semiparametric trending panel data regression and its applications in modeling economic, financial and climatological data. An example that illustrates the importance of such models in finance is the capital asset pricing model (CAPM). A traditional CAPM usually assumes a constant linear relationship between an asset's return and a market portfolio's return, and such a relationship is reflected by the beta coefficient. However, some recent studies show that the beta coefficients might vary over time (see for example, Jagannathan and Wang, 1996; Ghysels, 1998; Wang, 2003).

In this paper, our investigation is focused on a time–varying coefficient model given by

$$y_t = x_t^\top \beta(\tau_t) + u_t, \quad t = 1, 2, \cdots, n, \tag{1}$$

where $x_t^\top = (x_{t1}, x_{t2}, \cdots, x_{tk})$ is a stationary time series, $\beta(\tau_t)^\top = (\beta_1(\tau_t), \beta_2(\tau_t), \cdots, \beta_k(\tau_t))$ are unknown functions of $\tau_t = \frac{t}{n}$, and u_t follows a stationary AR(1) process $u_t = \alpha u_{t-1} + v_t$, in which α ($|\alpha| < 1$) is an unknown parameter, $\{v_t\}$ is a sequence of independent and identically distributed (i.i.d.) continuous random errors with a parametrically known density function $f(v;\eta)$ characterized by a parameter vector η, a vector of unknown parameters, $\mathbb{E}_\eta[v_1] = 0$ and $0 < \mathbb{E}_\eta[v_1^2] < \infty$. It is also assumed that x_t and u_t are mutually independent.

The unknown coefficient function $\beta(\tau_t)$ can be estimated by local constant method. It is generally accepted that the performance of the local constant estimator is mainly determined by the choice of bandwidth. Robinson (1989) proposed choosing bandwidth through the cross–validation (CV) method, but even in some simple settings, cross–validation may perform poorly and exhibit a large magnitude of sample variation (see Fan, Heckman and Wand, 1995, among others). Cai and Tiwari (2000), Cai (2002) and Cai (2007) proposed a nonparametric version of Akaike information criterion (AIC) to select bandwidth for local linear estimators. This method aims to derive an optimal bandwidth that minimizes a measure of penalized logarithmic mean squared error, and thus, is similar to the CV. A rule–of–thumb method for choosing bandwidth is the normal reference rule (NRR) defined as

$$h = 1.06\sigma n^{-1/5}, \tag{2}$$

where σ is the population standard deviation and is replaced by its sample measure in practice. It is a rough method for choosing bandwidths and is used in the absence of no other practical methods. In the current literature, there exist some investigations on the advantages of Bayesian sampling approaches to bandwidth estimation against some competing methods; see Zhang, King and Hyndman (2006).

Motivated by recent investigations on sampling algorithms for bandwidth estimation, we propose a Bayesian method for bandwidth estimation in the time–varying linear models given by (1), where the coefficients are estimated by local constant estimation method. The idea of using Bayesian approach to bandwidth selection is not new. However, to the best of our knowledge, there is no study available to provide theoretical support for this method. Therefore, we aim to fill this gap in the literature by establishing a large sample theory for Bayesian bandwidth estimators. In addition, we contribute to the current literature by investigating the large sample behavior of Bayesian estimators for parameters in the error density as the relevant studies in the current literature focus on the asymptotic behavior of posterior distribution rather than investigating posterior mean. For example, Walker (1969) showed that under suitable regularity conditions, as $n \to \infty$, the posterior distribution converges to a normal distribution. Based on Walker (1969), Chen (1985) further introduced three sets of conditions for asymptotic posterior normality. Phillips and Ploberger (1996) developed an asymptotic theory of Bayesian inference for stationary and nonstationary time series and provided the limiting form of the Bayesian data density for a general case of likelihoods and prior distributions. Kim (1998) considered posterior normality in the situation of nonstationary time series.

The contribution of this paper is summarized as follows.

- We propose a Bayesian approach to bandwidth estimation for local constant estimators of time–varying coefficients in time series models.

- We establish a large sample theory for a Bayesian bandwidth estimator, and this fills a gap in the current literature. We also provide a large sample theory for Bayesian estimators of a vector of unknown parameters that characterize the error density function.

- We examine the finite sample performance of the proposed Bayesian approach to parameter estimation through simulation studies.
- We compare the performance of our Bayesian bandwidth estimator in estimating time–varying coefficient functions with cross–validation and normal reference rule through simulation studies and empirical applications.

The rest of this paper is organized as follows. Section 2 briefly describes the local constant estimators of time–varying coefficients in the time series regression model and presents the large sample theory for the proposed Bayesian estimators of bandwidth and parameters in the error density. In Section 3, we present Monte Carlo simulation studies to examine the finite sample performance of our proposed method for bandwidth estimation and evaluate the accuracy of estimated coefficient function with our proposed Bayesian bandwidth estimator. In Section 4, two empirical examples are presented to illustrate the application of our proposed Bayesian method for bandwidth estimation. Section 5 concludes the paper. The key assumptions and an outline of the proof of the main results are given in Appendix A. A nonparametric specification testing procedure is detailed in Appendix B. Meanwhile, the full proofs of the main results are given in Appendix C of a supplementary document.

2 Large sample theory for Bayesian bandwidth estimation

In this section, we briefly describe the local constant estimator of the time–varying coefficient function in the time series model. We then establish large sample properties for the proposed Bayesian estimator of the bandwidth involved in a local constant kernel method, and for the unknown parameters that characterize the error density.

2.1 Local constant kernel method

The principle idea of local constant estimation method is that if $\beta(\tau)$ is continuous, it behaves like a reasonable constant over its small neighbourhood. So we approximate $\beta(\tau)$ by a constant a in the neighborhood of τ and minimize a locally weighted sum of squares to obtain a local constant kernel estimator, $\widehat{\beta}(\tau;h)$, of the form:

$$\widehat{\beta}(\tau;h) = \arg\min_{a} \sum_{t=1}^{n} (y_t - x_t^\top a)^2 K_h(\tau_t - \tau), \tag{3}$$

where $K_h(u) = K(u/h)/h$, $K(\cdot)$ is a kernel function and $h > 0$ is the bandwidth satisfying $h \to 0$ and $nh \to \infty$ as $n \to \infty$. Solving (3), we can obtain that

$$\widehat{\beta}(\tau;h) = \left(\sum_{t=1}^{n} x_t x_t^\top K_h(\tau_t - \tau)\right)^{-1} \sum_{t=1}^{n} x_t y_t K_h(\tau_t - \tau). \tag{4}$$

Note that it is also possible to use a local linear kernel method. Since both the establishment and the proof of the main results based on the local constant kernel method are already very technical, we just use the local constant kernel method in this paper. From (4), it is clear that we need to specify a kernel function $K(\cdot)$ and choose a bandwidth h before we may study and compute the estimate of $\beta(\tau)$. It is generally accepted in the current literature that the choice of $K(\cdot)$ is less important than that of h. In this paper, we propose a Bayesian method to estimate the bandwidth, for which we will establish a large sample theory in the following section.

2.2 Bayesian bandwidth estimation

Denote $X_n = (x_1, x_2, \cdots, x_n)^\top$ and $Y_n = (y_1, y_2, \cdots, y_n)^\top$, where $x_1, x_2, \cdots, x_n$ are observed values of the regressor and $y_1, y_2, \cdots, y_n$ are observed values of the response in (1). Let $\theta = (\alpha, \eta^\top)^\top$. For notational simplicity, we will use $f(v;\theta)$ to denote $f(v;\eta)$ throughout the rest of this paper.

Let $h = a_n\lambda$ with $\{a_n\}$ being a sequence of real numbers satisfying $a_n \to 0$ as $n \to \infty$ and λ being a random variable. Since $\{v_t\}$ is assumed to be i.i.d., we introduce the following likelihood function:

$$L_n(\lambda, \theta) = \prod_{t=1}^{n} f(\widehat{v}_t;\theta) = \prod_{t=1}^{n} f(\widehat{u}_t - \alpha\widehat{u}_{t-1};\theta),$$

where $\widehat{u}_t = y_t - x_t^\top \widehat{\beta}(\tau_t;h)$ in which $\widehat{\beta}(\cdot\,;h)$ is being defined in (4). The logarithm of likelihood function is then $l_n(\lambda,\theta) = \log L_n(\lambda,\theta)$.

Let $\pi_\lambda(\lambda)$ be the density function of λ and $\pi_\theta(\theta)$ be the density function of θ. Then, we estimate $(\lambda, \theta^\top)^\top$ by

$$\mathbb{E}\left[(\lambda,\theta^\top)^\top | X_n, Y_n\right] = \frac{\int (\lambda,\theta^\top)^\top e^{l_n(\lambda,\theta)} \pi_\lambda(\lambda)\pi_\theta(\theta) d\lambda d\theta}{\int e^{l_n(\lambda,\theta)} \pi_\lambda(\lambda)\pi_\theta(\theta) d\lambda d\theta} = \begin{pmatrix} \mathbb{E}_\star[\lambda | X_n, Y_n] \\ \mathbb{E}_\star[\theta | X_n, Y_n] \end{pmatrix}.$$

Since the Bayesian estimates $\left(\mathbb{E}_\star[\lambda|X_n,Y_n], \mathbb{E}_\star[\theta^\top|X_n,Y_n]\right)^\top$ do not have closed–form expressions, we therefore approximate them by $\widehat{\lambda}_{mn}$ and $\widehat{\theta}_{mn}$, respectively, where $\widehat{\lambda}_{mn} = \frac{1}{m}\sum_{j=1}^m \lambda_{jn}$ and $\widehat{\theta}_{mn} = \frac{1}{m}\sum_{j=1}^m \theta_{jn}$, in which λ_{jn} and θ_{jn} denote the respective jth posterior draws and m denotes the number of Markov chain Monte Carlo (MCMC) iterations. Consequently, we can obtain the Bayesian bandwidth estimator $\widehat{h}_{mn}$ by $\widehat{h}_{mn} = a_n\widehat{\lambda}_{mn}$.

Before we establish an asymptotic distribution for $\widehat{\lambda}_{mn}$, we introduce the following notation. Let $\gamma(v_t;\theta) = \frac{f^{(1)}(v_t;\theta)}{f(v_t;\theta)}$ where $f^{(1)}(v_t;\theta)$ is the first derivative of $f(v_t;\theta)$. Let $\gamma_n(v_t) = \int(\mathbb{E}[\theta]-\theta)\gamma(v_t;\theta)A_n(\theta)\pi(\theta)d\theta$ and $\phi_n(v_t) = \int\gamma(v_t;\theta)A_n(\theta)\pi(\theta)d\theta$ with $A_n(\theta) = \frac{e^{n\mathbb{E}_\theta[\log f(v_1;\theta)]}}{\int e^{n\mathbb{E}_\theta[\log f(v_1;\theta)]}\pi_\theta(\theta)d\theta}$ and $\mathbb{E}_\theta[\log f(v_1;\theta)] = \int \log f(v_1;\theta)f(v;\theta)dv$. Denote $G_n(\theta) = \prod_{t=1}^n f(v_t;\theta)$ and $g_n(\theta) = \log G_n(\theta)$. Let $f_n(\theta|X_n,Y_n) = G_n(\theta)\pi_\theta(\theta)/(\int G_n(\theta)\pi_\theta(\theta)d\theta)$ and $\mathbb{E}[\theta|X_n,Y_n] = \int\theta f_n(\theta|X_n,Y_n)d\theta$. Let $\widehat{\theta}_n$ denote the maximum likelihood estimator of θ and $\Delta_n(\widehat{\theta}_n) = -g_n^{(2)}(\widehat{\theta}_n)$, where $g_n^{(2)}(\widehat{\theta}_n)$ is the second derivative of $g_n(\cdot)$ evaluated at the maximum likelihood estimator $\widehat{\theta}_n$. Denote $C_n = \left\{\theta : \Delta_n^{1/2}(\widehat{\theta}_n)\|\widehat{\theta}_n - \theta\| < c\right\}$ and $D_n = \left\{\theta : \Delta_n^{1/2}(\widehat{\theta}_n)\|\widehat{\theta}_n - \theta\| \geq c\right\}$, where c is a positive constant.

Let $\beta^{(i)}(\tau)$ denote the ith order derivative of $\beta(\tau)$. Throughout this paper, we use $\mathbb{E}_\theta[Z]$ to denote the conditional expectation of Z given θ. Let $\Sigma_x = \mathbb{E}\left[x_1x_1^\top\right]$, $w_t = x_t^\top\Sigma_x^{-1}x_t$, $d_{1n} = \mathbb{E}[v_1\phi_n(v_1)]$, $d_{2n} = \mathbb{E}[v_1\gamma_n(v_1)]$, $b_{1n} = a_n^{-1}\mathbb{E}[w_1]\ d_{1n}$, $b_{0n} = 1 - K(0)\mathbb{E}[\lambda^{-1}]b_{1n}$ and $\delta_n^2 = \mathbb{E}[w_1^2]\ \mathbb{E}\left[\left(v_1\phi_n(v_1) - d_{1n}\right)^2\right]$. Let $d_1 = -K(0)\left(1 - \mathbb{E}[\lambda]\mathbb{E}[\lambda^{-1}]\right)$.

We then establish the main asymptotic distributions in Theorems 1–3 below.

Theorem 1. Let Assumptions 1–4 listed in Appendix A1 hold. If, in addition, $\frac{n\ (a_n^2+d_{1n}^2)}{m\ \delta_n^2} \to 0$ as $m\to\infty$ and $n\to\infty$, then we have

$$\sqrt{n}\,\delta_n^{-1}a_n\left(b_{0n}\left(\widehat{\lambda}_{mn} - \mathbb{E}[\lambda]\right) - d_1b_{1n}\right) \longrightarrow_D \mathcal{N}(0, d_1^2).$$

Remark 1: Theorem 1 establishes an asymptotic normality for $\widehat{\lambda}_{mn}$ with a rate of convergence $\sqrt{n}\,\delta_n^{-1}a_n$. Note that Theorem 1 implies that the rate of convergence for the bandwidth estimator is $\sqrt{n}\,\delta_n^{-1}$. Note also that b_{1n} disappears asymptotically when $a_n^{-1}d_{1n}\to 0$ as $n\to\infty$. In this case, $\widehat{\lambda}_{mn}$ is an unbiased estimator of $\mathbb{E}[\lambda]$ with a rate of $\sqrt{n}\,\delta_n^{-1}a_n$. When $d_{1n}\to C_\phi \neq 0$ for $|C_\phi| < \infty$, $\widehat{\lambda}_{mn}$ is a biased estimator of $\mathbb{E}[\lambda]$ with a rate of $\sqrt{n}\,\delta_n^{-1}$.

Theorem 2. (i) Let Assumptions 1–3 and 5 listed in Appendix A1, If, in addition, $\frac{na_n^2}{\Delta_n(\widehat{\theta}_n)\sigma_n^2} \to_P 0$ and

$\frac{n\,(a_n^2+d_{2n}^2)}{m\,\sigma_n^2} \to 0$ as $m \to \infty$ and $n \to \infty$, then we have

$$\sqrt{n}\,\sigma_n^{-1} a_n \left(\widehat{\theta}_{mn} - \mathbb{E}[\theta] - b_{2n}\right) \to_D \mathcal{N}(0, K^2(0)\mathbb{E}^2[\lambda^{-1}] I_p),$$

where $b_{2n} = a_n^{-1} K(0)\mathbb{E}[\lambda^{-1}]\,\mathbb{E}[w_1]\,d_{2n}$, $\sigma_n^2 = \mathbb{E}[w_1^2]\mathbb{E}\left[\left(v_1\gamma_n(v_1) - d_{2n}\right)^2\right]$, I_p denotes the $p \times p$ identity matrix, and p denotes the number of parameters in the vector θ.

(ii) Let Assumptions 1(i), 2(i), 3(i) and 5(ii)(iii)(iv) hold. If, in addition, $\gamma(v;\eta) = \frac{f^{(1)}(v;\eta)}{f(v;\eta)} \equiv \gamma(v)$ does not depend on η, then as $\frac{\Delta_n(\widehat{\theta}_n)}{m} \to_P 0$, we have as $m \to \infty$ and $n \to \infty$

$$\sqrt{\Delta_n(\widehat{\alpha}_n)}\;(\widehat{\alpha}_{mn} - \mathbb{E}[\alpha]) \to_D \mathcal{N}(0, \Sigma_0), \tag{5}$$

where Σ_0 is a positive definite matrix as defined in Assumption 5(iv).

Remark 2: (i) Theorem 2(i) shows that asymptotic normality is achievable for Bayesian estimator $\widehat{\theta}_{mn}$ with a rate of convergence of $\sqrt{n}\,\sigma_n^{-1} a_n$. To the best of our knowledge, this is a new finding about Bayesian estimation, particularly for a class of posterior means, as the current relevant literature has only established some asymptotic properties for the posterior distribution; see Walker (1969), Chen (1985) and Kim (1998) etc. Theorem 2(i) also shows that the conventional $\sqrt{n}$–rate of convergence is achievable when $\frac{a_n}{\sigma_n} \to C_0 \neq 0$.

(ii) As expected, Theorem 2(ii) shows that α can be consistently estimated with the conventional $\sqrt{n}$–rate of convergence when $\gamma(v;\eta) = \frac{f^{(1)}(v;\eta)}{f(v;\eta)} \equiv \gamma(v)$ does not depend on η and $\frac{\Delta_n(\widehat{\theta}_n)}{n} \to \Delta_\alpha > 0$, where Δ_α is a positive definite matrix.

An asymptotically normal distribution for the local constant estimator $\widehat{\beta}(\tau; \widehat{h}_{mn})$ with bandwidth estimated by our proposed Bayesian method is given in the following theorem.

Theorem 3. Let Assumptions 1–4 listed in Appendix A1 hold. As $m \to \infty$ and $n \to \infty$, we have

$$\sqrt{n a_n \cdot \mathbb{E}[\lambda]}\;\left(\widehat{\beta}(\tau; \widehat{h}_{mn}) - \beta(\tau)\right) \longrightarrow_D \mathcal{N}(0, \Sigma_\beta),$$

where $\Sigma_\beta = \sigma_u^2 \cdot \int K^2(v)dv \cdot \Sigma_x^{-1}$ with $\sigma_u^2 = \mathbb{E}[u_1^2]$.

Remark 3: Theorem 3 shows that with our proposed Bayesian bandwidth, an asymptotic normality for the local constant estimator of the coefficient function, $\widehat{\beta}(\cdot\,; h)$, is achievable. We can find that the local constant estimator has a rate of convergence at an order of $\sqrt{n\,a_n}$, which is equivalent to the rate of $\sqrt{nh}$ for the case where h is being treated as a fixed bandwidth.

Before we give an outline of the proof of Theorems 1–3 in Appendix A and then the full proofs in the supplementary material, we examine the finite sample performance of our Bayesian estimators by simulation studies in Section 3.

3 Simulation

The purposes of the Monte Carlo simulation study are as follows. First, with 1000 replications under different data generating processes, we examine the finite sample performance of the proposed Bayesian approach to parameter estimation. Second, we compare the performance of our proposed Bayesian approach to bandwidth estimation with cross–validation method and normal reference rule by computing mean squared error of the estimated time–varying coefficient functions.

3.1 Parameter estimation

We consider the following time–varying coefficient time series model

$$y_t = x_t^\top \beta(\tau_t) + u_t, \quad t = 1, 2, \cdots, n \tag{6}$$

where $x_t = (1, x_{t2})^\top$, $\beta(\tau) = (\beta_1(\tau), \beta_2(\tau))^\top$, $\beta_1(\tau) = 0.2\exp(-0.7 + 3.5\,\tau)$ and $\beta_2(\tau) = 2\tau + \exp(-16(\tau - 0.5)^2) - 1$. In order to generate samples, we generate x_{t2} through an $AR(1)$ model given by $x_{t2} = 0.5x_{(t-1)2} + \epsilon_x$ with ϵ_x being generated from $\mathcal{N}(0, 0.5^2)$. We generate the error term of (6) from a stationary AR(1) process $u_t = \alpha u_{t-1} + v_t$ with α being generated from $\alpha = 0.5 + \epsilon_1$, where ϵ_1 comes from a truncated normal distribution with mean zero and standard deviation 0.5, and values of ϵ_1 are bounded within $(-0.4, 0.4)$, and v_t is generated independently from a specified parametric distribution.

We first consider the following two distributions for v_t.

[DGP1] Gaussian error $v_t \sim \mathcal{N}\left(0, \sigma_g^2\right)$: σ_g is generated from $\sigma_g = 0.7 + \epsilon_2$, where ϵ_2 is generated from the truncated $\mathcal{N}(0, 0.5^2)$ and are restricted between -0.4 and 0.4.

[DGP2] A mixture of two Gaussian distributions: $v_t \sim w\mathcal{N}\left(\mu_1, \sigma_1^2\right) + (1-w)\mathcal{N}\left(\mu_2, \sigma_2^2\right)$, where w is the weight assigned to the first component density, μ_1 and μ_2 are the mean parameters for these two component densities, and σ_1^2 and σ_2^2 are the variance parameters for these two component densities.

- w is generated according to $w = 0.5 + \epsilon_3$, where ϵ_3 is generated from the truncated $\mathcal{N}(0, 0.5^2)$ and is restricted within $(-0.4, 0.4)$.
- σ_1 is generated from $\sigma_1 = 0.7 + \epsilon_4$, and σ_2 is generated from $\sigma_2 = 1.1 + \epsilon_5$, where both ϵ_4 and ϵ_5 are independently generated from the truncated $\mathcal{N}(0, 0.5^2)$ and are restricted between -0.3 and 0.3.
- μ_1 is generated according to $\mu_1 = 0.2 + \epsilon_6$, where ϵ_6 is generated from the truncated $\mathcal{N}(0, 0.5^2)$ and is bounded between -0.3 and 0.3.
- μ_2 is computed by solving the equation $\mu_2 = -w \times \mu_1 / (1.0 - w)$.

According to the existing theoretical results in nonparametric estimation, h is proportional to $n^{1/(4+p)}$, where p is the dimension of variables, with which we are smoothing. In this simulation study, as the unknown coefficient is a function of time, so we set $a_n = n^{-1/5}$. Our objective is to choose λ by our proposed Bayesian method. Based on each DGP, we can obtain likelihood function. For the choice of prior, we choose inverse gamma prior denoted as $IG(1, 0.05)$ for h^2, σ_g^2, σ_1^2 and σ_2^2, uniform prior $U(-0.95, 0.95)$ for AR coefficient α, uniform prior $U(0.05, 0.95)$ for weight parameter w in DGP2 and Gaussian prior $\mathcal{N}(0, 9)$ for μ_1 in DGP2. Then by Bayes theorem, we can obtain posterior, from which we can use Metropolis algorithm to sample parameters.

We choose sample size n to be 200, 600 and 1200, respectively. For each data generating process (DGP) and each sample size, we generate 1000 samples.

For DGP1, we estimate parameters $\theta_g = \left(\theta_{g1}, \theta_{g2}\right)^\top = \left(\alpha_g, \sigma_g^2\right)^\top$, where α_g and σ_g^2 are the AR coefficient and variance of v_t, respectively. For DGP2, we estimate parameters $\theta_m = (\theta_{m1}, \theta_{m2})^\top = \left(\alpha_m, \sigma_m^2\right)^\top$, where α_m and σ_m^2 denote AR coefficient and variance of v_t, respectively. Note that σ_m^2 is computed by $\sigma_m^2 = w \times (\mu_1^2 + \sigma_1^2) + (1 - w) \times (\mu_2^2 + \sigma_2^2)$ given that $w\mu_1 + (1 - w)\mu_2 = 0$. We use our proposed Bayesian method to estimate θ_g and θ_m. The estimates are denoted as $\widehat{\theta}_g = \left(\widehat{\theta}_{g1}, \widehat{\theta}_{g2}\right)^\top = (\widehat{\alpha}_g, \widehat{\sigma}_g^2)^\top$ and $\widehat{\theta}_m = \left(\widehat{\theta}_{m1}, \widehat{\theta}_{m2}\right)^\top = (\widehat{\alpha}_m, \widehat{\sigma}_m^2)^\top$.

In each case, λ was also estimated. Since it is involved in h, its performance is incorporated in the estimation of the coefficient functions in Section 3.2 below.

To examine the finite sample performance of the Bayesian method, we compute the biases and the mean squared errors (MSEs) for each of the components of $\widehat{\theta}_g$ and $\widehat{\theta}_m$ as follows: for $i = g, m$

and $j = 1,2$

$$\text{Bias}_{ij} = \frac{1}{1000}\sum_{r=1}^{1000}(\widehat{\theta}_{ij}(r) - \theta_{ij}(r)) \text{ and } \text{MSE} = \frac{1}{1000}\sum_{r=1}^{1000}(\widehat{\theta}_{ij}(r) - \theta_{ij}(r))^2,$$

where $\theta_{ij}(r)$ is the value of θ_{ij} in the r–th replication and $\widehat{\theta}_{ij}(r)$ is the estimated value of $\widehat{\theta}_{ij}$ in the r–th replication. The results of parameter estimates are presented in Table 1. From Table 1, we find that with the sample size increasing, the bias and the MSE of the AR coefficient and variance of v_t in both DGPs decrease. This indicates that our proposed Bayesian approach to parameter estimation has very good finite sample performance.

Table 1: Bias and MSE of parameter estimates based on 1000 replications

		DGP 1		DGP 2	
	n	$\widehat{\alpha}_g$	$\widehat{\sigma}_g^2$	$\widehat{\alpha}_m$	$\widehat{\sigma}_m^2$
Bias	200	−0.1193	0.0277	−0.1005	−0.1460
	600	−0.0308	0.0112	−0.0295	−0.0709
	1200	−0.0140	0.0067	−0.0130	−0.0259
MSE	200	0.0281	0.0040	0.0230	0.0462
	600	0.0029	0.0011	0.0024	0.0190
	1200	0.0009	0.0005	0.0008	0.0088

3.2 Estimation of time–varying coefficient functions

As our objective is to obtain better estimates for unknown coefficient functions, we examine the performance of our proposed Bayesian method to bandwidth estimation by checking the accuracy of $\widehat{\beta}(\cdot;\ h)$. We considered DGP1, DGP2 and the following DGP3:

[DGP3] Centralised Chi–squared distribution: $v_t \sim \frac{\chi^2(2)-2}{2}$.

For each DGP, we assume the following structure for the error term.

[Case 1] Assume $u_t = \alpha u_{t-1} + v_t$ and $v_t \sim \mathcal{N}(0,\sigma^2)$.

[Case 2] Assume $u_t = \alpha u_{t-1} + v_t$ and $v_t \sim w\mathcal{N}(\mu_1,\sigma_1^2) + (1-w)\mathcal{N}(\mu_2,\sigma_2^2)$.

For each case, we estimate λ and further get bandwidth by $h = a_n\lambda$. We then compute $\widehat{\beta}(\cdot;\ h)$ using the estimated bandwidth. We plotted the median of estimated coefficient functions $\widehat{\beta}_1(\cdot;\ h)$

and $\widehat{\beta}_2(\cdot;\ h)$ based on 1000 replications with sample size 600 in Figure 1. The plots with sample size 200 and 1200 are available in Appendix D of the supplementary document. From Figure 1, we find that under each DGP, estimated curves for $\beta_1(\tau)$ and $\beta_2(\tau)$ are very close to the corresponding true function curves.

In addition, we measure the accuracy of the estimated bandwidths by computing mean squared error for $\widehat{\beta}_1(\cdot;\ h)$ and $\widehat{\beta}_2(\cdot;\ h)$ separately.

$$\mathrm{MSE}_{\beta}(\widehat{h}_{mn}) = \frac{1}{1000}\frac{1}{n}\sum_{r=1}^{1000}\sum_{t=1}^{n}\left(\widehat{\beta}_j^r(\tau_t;\widehat{h}_{mn}^r) - \beta_j(\tau_t)\right)^2,$$

where $\widehat{\beta}_j^r(\tau_t;\widehat{h}_{mn}^r)$ is the estimate of $\beta_j(\tau_t)$ for the rth replication and $j = 1,2$ in which $\widehat{h}_{mn}^r = a_n\widehat{\lambda}_{mn}^r$ and $\widehat{h}_{mn}^r$ is the estimate of bandwidth for the rth replication.

We also compare our Bayesian method with cross–validation (CV) method. The optimal bandwidth by CV can be obtained by solving the following equation.

$$h_{\mathrm{CV}} = \arg\min_h \frac{1}{n}\sum_{t=1}^{n}\left(y_t - x_t^\top \widehat{\beta}^{-1}(\tau_t;h)\right),$$

where $\widehat{\beta}^{-1}(\tau_t;h)$ denotes the leave–one–out local constant estimator of $\beta(\tau_t)$. After we obtained the optimal h_{CV}, we can compute $\widehat{\beta}(\tau_t;h_{\mathrm{CV}})$ and then get the corresponding mean square error $\mathrm{MSE}_{\beta}(h_{\mathrm{CV}})$.

Table 2: Mean squared errors of $\widehat{\beta}_1(\cdot;\ h)$ and $\widehat{\beta}_2(\cdot;\ h)$.

		$\widehat{\beta}_1(\cdot;\ h)$				$\widehat{\beta}_2(\cdot;\ h)$			
	n	Bayes_1	Bayes_2	CV	NRR	Bayes_1	Bayes_2	CV	NRR
DGP1	200	0.0752	0.0780	0.1297	0.0793	0.1095	0.1123	0.1944	0.1149
	600	0.0285	0.0284	0.0448	0.0303	0.0365	0.0364	0.0678	0.0462
	1200	0.0162	0.0161	0.0224	0.0172	0.0186	0.0185	0.0334	0.0253
DGP2	200	0.1194	0.1167	0.2145	0.1223	0.1861	0.1811	0.3501	0.2073
	600	0.0439	0.0430	0.0777	0.0518	0.0599	0.0584	0.1230	0.0839
	1200	0.0255	0.0254	0.0400	0.0304	0.0299	0.0296	0.0597	0.0453
DGP3	200	0.1409	0.1034	0.2486	0.1488	0.2188	0.1409	0.3858	0.2262
	600	0.0478	0.0443	0.0879	0.0587	0.0601	0.0504	0.1312	0.0890
	1200	0.0277	0.0262	0.0453	0.0339	0.0328	0.0288	0.0673	0.0502

Figure 1: Estimated curves for $\beta_1(\tau)$ and $\beta_2(\tau)$ under DGP1–3 in Case 1 (the first two rows) and Case 2 (the last two rows) with sample size $n = 600$

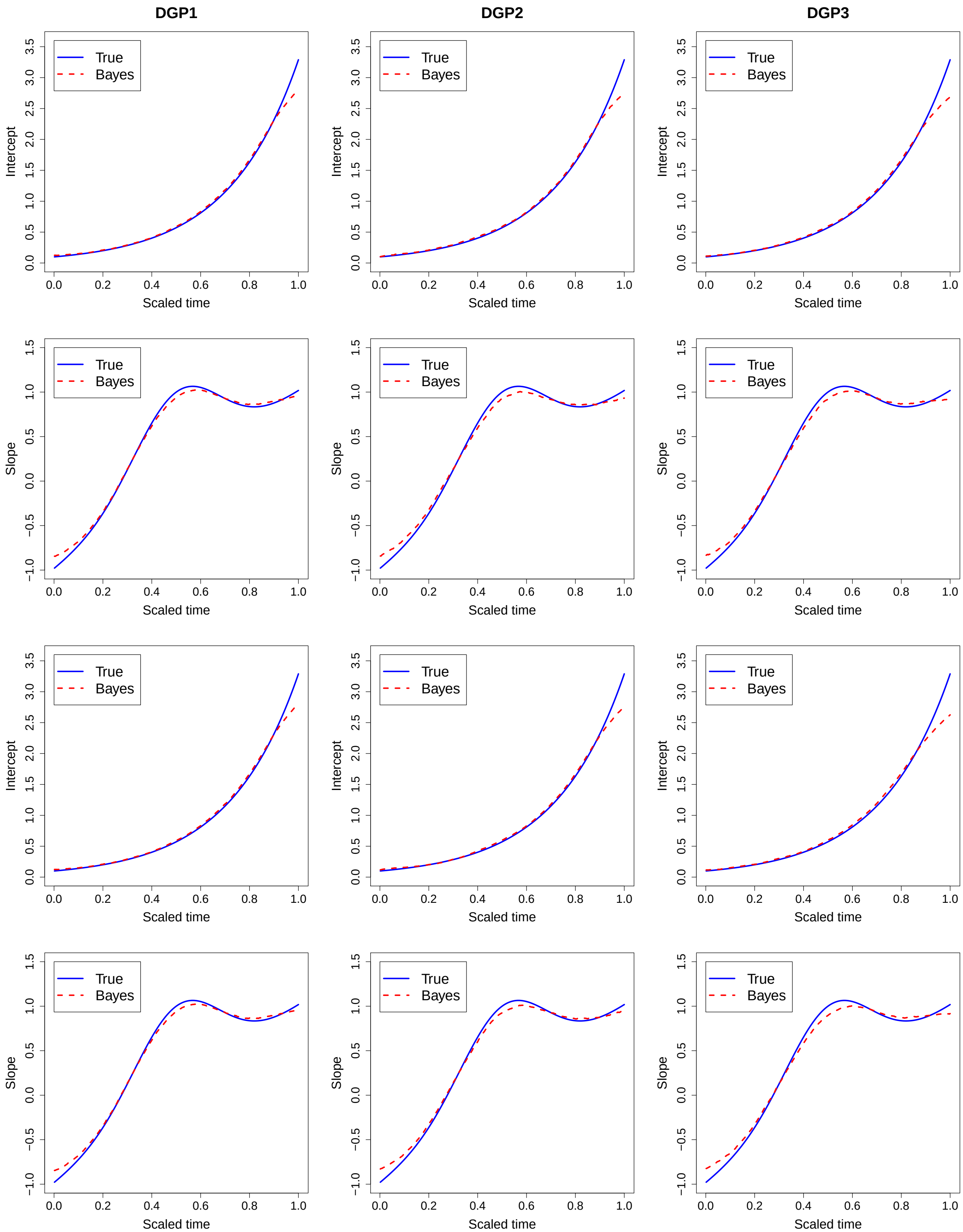


Table 2 presents the estimation results for coefficient functions in different data generating processes in both Case 1 and Case 2. The columns of Bayes_1 and Bayes_2 in Table 2 denote the mean square error of local constant estimator of coefficient functions with bandwidth estimated by Bayesian approach in Case 1 and Case 2, respectively.

Table 2 reveals the following results.

- Under each DGP, with each bandwidth selection method, when the sample size increases, the mean squared error decreases.
- Under each DGP, the mean squared error with bandwidth selected by our Bayesian method is smaller than that with bandwidth selected by the cross–validation method and normal reference rule.
- An interesting fact is that the mean squared error with the bandwidth selected by the cross–validation is larger than that with normal reference rule. This is because that our simulated data are dependent and cross–validation method does not work well if data under consideration are dependent.
- For DGP1, the mean squared errors in Case 1 and Case 2 are similar. This is because that normal distribution is a special case of mixture normal distribution.
- For DGP2 and DGP3, the mean squared error in Case 2 is smaller than that in Case 1. This is reasonable because for DGP2, the mixture of two normal distributions is the correct assumption. It is expected to have a smaller mean squared error than its competitor. However, due to the small distance between the two normal components, the difference is not very obvious. For DGP3, as the centralized χ^2 distribution is asymmetric, a single normal distribution cannot capture this feature. In contrary, the mixture of two normal distributions can.

4 Application

4.1 Okun's law

Since the seminal work of Okun (1962), there have been many studies that provide evidence of the correlation between variations in unemployment rate and real output over a business cycle.

However, many studies in the current literature on Okun's law tend to focus on the lack of robustness of Okun's coefficient without questioning the constant linear relationship (see for example, Lee, 2000; Silvapulle, Moosa and Silvapulle, 2004). In this section, we employ the time–varying coefficient model to demonstrate the time–varying feature of Okun's coefficient.

The sample contains yearly paired observations of unemployment rate and GDP of the U.S. over the period from 1948 to 2010. The data were collected from the OECD database. Let y_t denote the unemployment rate and x_t denote the logarithm of GDP. Define $\Delta y_t = y_t - y_{t-1}$ and $\Delta x_t = x_t - x_{t-1}$.

Okun's law is usually investigated through the model given by

$$\Delta y_t = \alpha_1 + \alpha_2 \Delta x_t + u_t,$$

where Δy_t is the yearly change of the unemployment rate, Δx_t is the yearly change of the log GDP and u_t is assumed to follow a stationary AR(1) process. The time series plots of unemployment rate, GDP, change of unemployment rate and GDP growth rate are presented in Figure 2.

The ordinary least squares (OLS) estimates of α_1 and α_2 are, respectively, 0.0135(0.0000) and −0.3989(0.0000) with corresponding p values given in the parentheses. We fitted the time–varying coefficient model:

$$\Delta y_t = \beta_1(\tau_t) + \beta_2(\tau_t)\Delta x_t + u_t$$

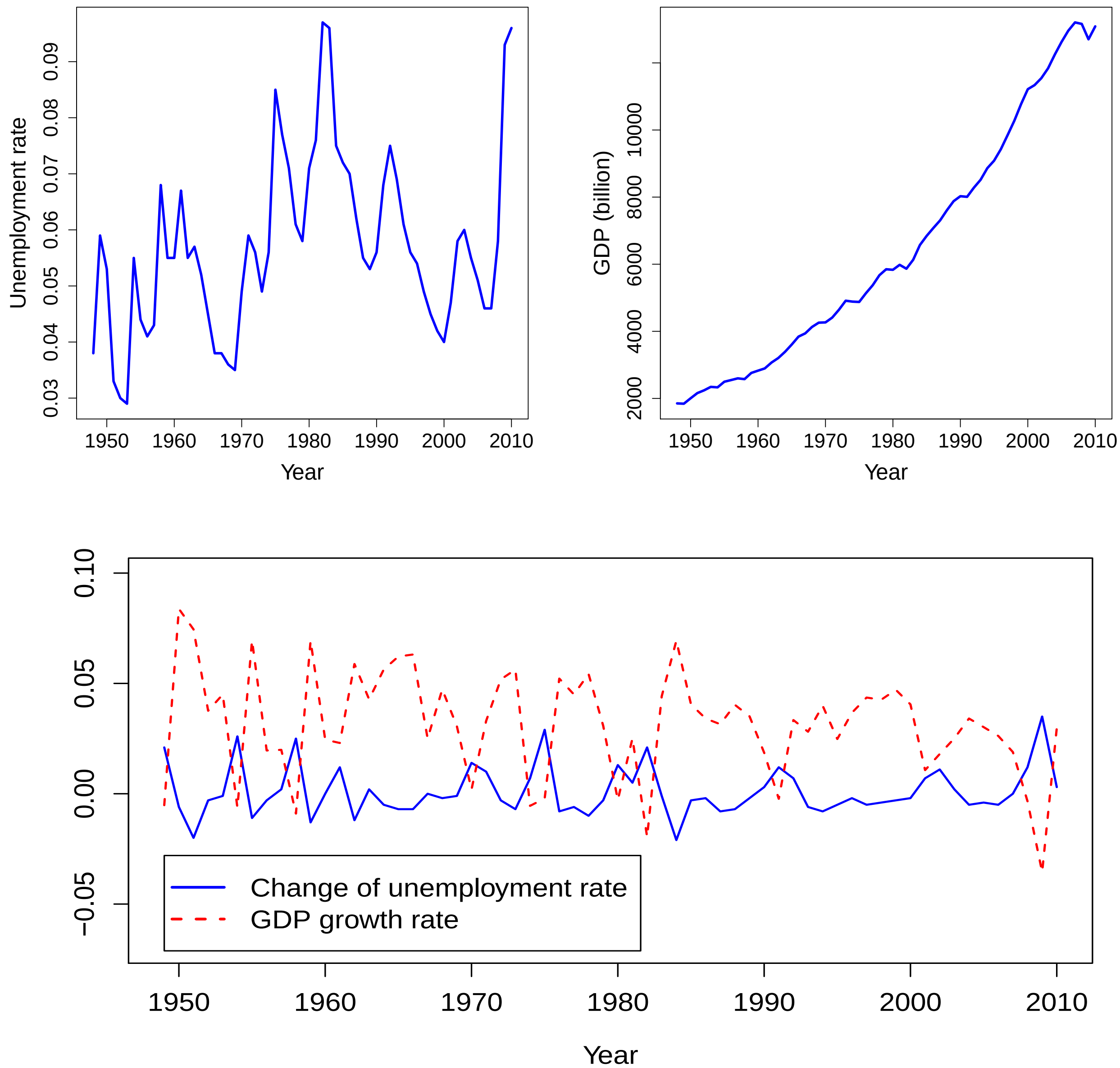
to the sample data, where the errors are assumed to follow an AR(1) process $u_t = \alpha u_{t-1} + v_t$ with v_t being assumed to follow a mixture of two normal distributions. Local constant estimates of $\beta_1(\tau)$ and $\beta_2(\tau)$ were computed with the bandwidth chosen through NRR, CV and Bayesian sampling. The resulting estimates $\widehat{\beta}_1(\tau)$ and $\widehat{\beta}_2(\tau)$ are presented in Figure 3.

From Figure 3, we found that the local constant estimates of $\beta_1(\tau)$ and $\beta_2(\tau)$ with the bandwidth estimated through Bayesian sampling are clearly different with those with bandwidth selected through CV and NRR. The bandwidth selected through CV tend to over–smooth the unknown functions while the bandwidth by NRR tend to under–smooth the unknown functions.

The MSE values of the constant coefficient model and time–varying coefficient model are respectively, 0.2894×10^{-4} and 0.2273×10^{-4}. The introduction of time–varying coefficient improves model fitting by 21.46% in comparison to the constant coefficient model.

In order to examine whether the two coefficients are time–varying, we employ the method proposed by Cai, Fan and Yao (2000) to test the null hypothesis of constant coefficients against the

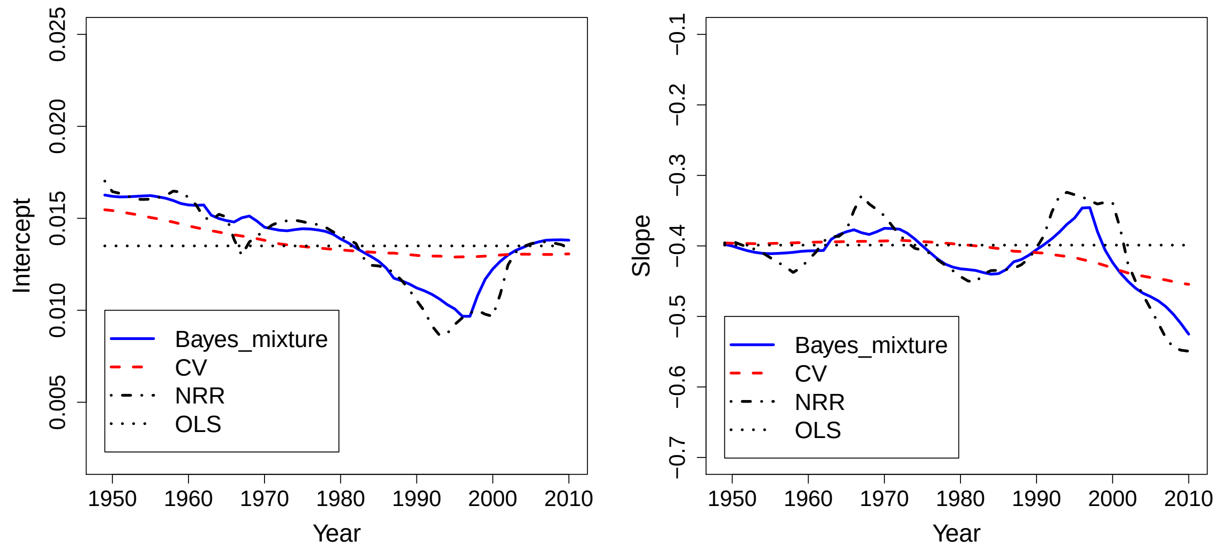
Figure 2: Time series plots of unemployment rate, GDP, change of unemployment rate and GDP growth rate



alternative of time–varying coefficients. The proposed bootstrap test (see Appendix B for details) with 1000 repetitions produced a p value that is approximately 0. Therefore, we rejected the null hypothesis of constant coefficients in the linear regression model at the 1% significance level.

We also derived the 95% point–wise confidence intervals of the two time–varying coefficients based on the asymptotic results obtained in Cai (2007) and we plotted the confidence intervals in Figure 4.

Figure 3: Local constant estimates of $\beta_1(\tau)$ and $\beta_2(\tau)$ with the bandwidth derived through NRR, CV and Bayesian sampling



According to the patterns of the estimated time–varying coefficients shown in Figure 3, a piecewise polynomial function of time might be appropriate to approximate each coefficient function. Consider the time–varying coefficient model given by

$$\Delta y_t = \beta_1^*(\tau_t) + \beta_2^*(\tau_t)\Delta x_t + u_t, \tag{7}$$

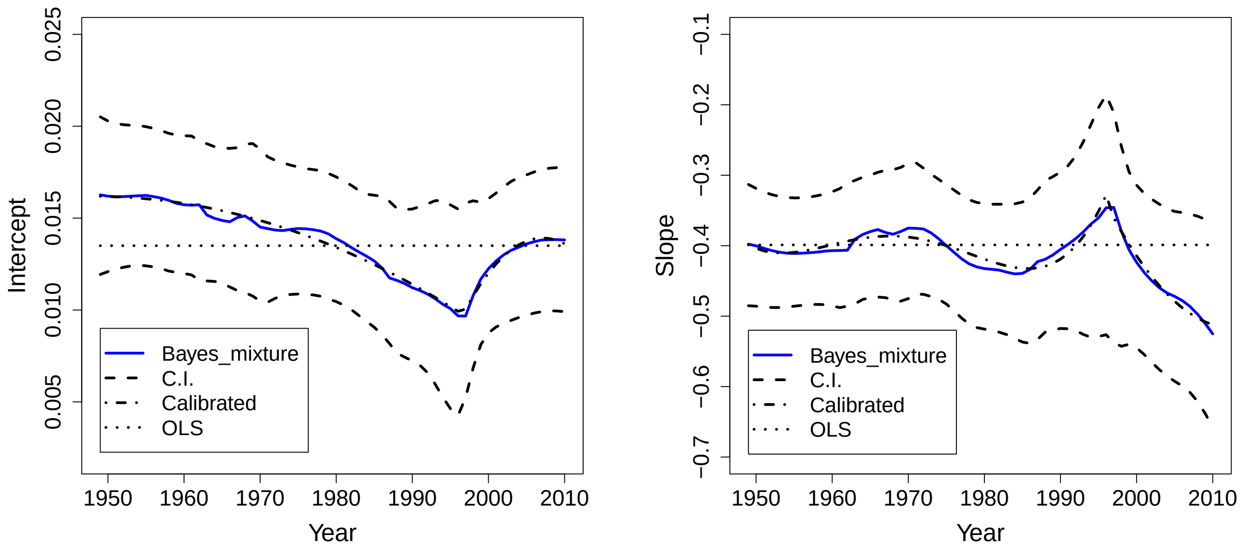
where $\beta_j^*(\tau_t)$, for $j = 1$ and 2, are known–form functions of $\tau_t = \frac{t}{n}$ given by

$$\begin{aligned}
\beta_1^*(\tau_t) &= (0.0162 - 0.0104\tau_t)\, I(1 \le t \le 48) \\
&\quad + \left(-0.1179 + 0.2764\tau_t - 0.1449\tau_t^2\right) I(49 \le t \le 62), \\
\beta_2^*(\tau_t) &= \left(-0.3913 - 0.5054\tau_t + 4.0887\tau_t^2 - 10.0764\tau_t^3 + 7.4589\tau_t^4\right) I(1 \le t \le 48) \\
&\quad + \left(2.1162 - 5.0395\tau_t + 2.4110\tau_t^2\right) I(49 \le t \le 62).
\end{aligned}$$

The p values of all the above estimated coefficients are all zeros, and this indicates that all coefficients are significant at the 1% significance level.

The graphs of the calibrated coefficient functions are presented in Figure 4, where we could find that each piecewise polynomial function coefficient is very close to the corresponding local constant estimate. Figure 4 also shows that the 95% point–wise confidence intervals covered the calibrated time–varying coefficients. Thus, the calibrated coefficient functions are appropriate to capture the time–varying dynamics of Okun's coefficient.

Figure 4: Local constant estimates of $\beta_1(\tau)$ and $\beta_2(\tau)$ and the calibrated parametric functions with 95% point–wise confidence intervals



In order to validate the calibrated parametric time trends for the time–varying coefficients, we further test the null hypothesis that the two time–varying coefficients are piecewise polynomial functions of time against the alternative of local constant estimates. The bootstrap test procedure is based on 1000 repetitions. The proposed bootstrap test produced a p value of 0.9980. Therefore, we could not reject the null hypothesis at the 1% significance level. Our finding indicates that in different time period, Okun's coefficient is varying.

4.2 Consumption growth

The relationship between consumption growth and income growth has been extensively investigated in empirical macroeconomics during the past several decades. Such a relationship is usually investigated through the linear model given by

$$y_t = \alpha_1 + \alpha_2 x_t + \varepsilon_t, \tag{8}$$

where y_t is the consumption growth, which is the first difference of logarithm of the consumption expenditure, and x_t is the income growth, which is the first difference of logarithm of the disposable income. Carrol and Summers (1991) estimated the coefficient of income growth rate α_2, which is 0.601 based on the data for 15 OECD countries during the period from 1960 to 1985. Campbell and Mankiw (1989) estimated α_2 using the U.S. quarterly data during the period from 1953 to 1986.

They found that the estimate of α_2 is 0.316 through OLS, and this estimate becomes larger when the lagged consumption growth rates are used as instrumental variables.

However, the relationship between consumption growth and income growth may not necessarily be a constant over time. For example, if a consumer receives additional information on his/her present or future income, he/she would adjust his/her level of consumption discontinuously to be consistent with his/her new inter–temporal budget constraint. If the consumer knows that the income growth rate will increase, he/she will adjust his/her level and rate of consumption. However, the model given by (8) indicates that the response only depends on the level of income growth rate.

In this section, we treat the coefficients of (8) are time–varying, and investigate the time–varying relationship between consumers' consumption expenditure growth and disposable income growth during the period from 1960 to 2009 in the U.S. The data are quarterly and are available at the website of the Bureau of Economic Analysis. The time series plots of y_t and x_t are presented in Figure 5.

If the two coefficients are assumed to be constants, the OLS estimates of α_1 and α_2 are respectively, 0.0055 and 0.3399, with their p values being both zero. Assuming the two coefficients are time–varying, we fitted the model:

$$y_t = \beta_1(\tau_t) + \beta_2(\tau_t)x_t + u_t$$

to the above sample, where the errors are assumed to follow an AR(1) process $u_t = \alpha u_{t-1} + v_t$ with v_t being assumed to follow a mixture of two normal distributions. Local constant estimates of $\beta_1(\tau)$ and $\beta_2(\tau)$ were computed with the bandwidth estimated through the proposed Bayesian sampling. For comparison purpose, we also used NRR and CV to choose the bandwidths. The graphs of $\widehat{\beta}_1(\tau)$ and $\widehat{\beta}_2(\tau)$ with their bandwidths estimated/chosen through Bayesian sampling, NRR and CV, are presented in Figure 6. Note that the local constant estimators of the time–varying coefficients are determined by the bandwidth. The time–varying coefficients with their bandwidth estimated through Bayesian sampling are clearly different from those with their bandwidth chosen through NRR and CV.

We find that with the bandwidth estimated through Bayesian sampling, the long–term trend of the slope coefficient is decreasing. This reflects a overall weakening relationship between consumption growth and disposable income growth during the sample period, although for each

Figure 5: Time series plots of consumption growth and income growth

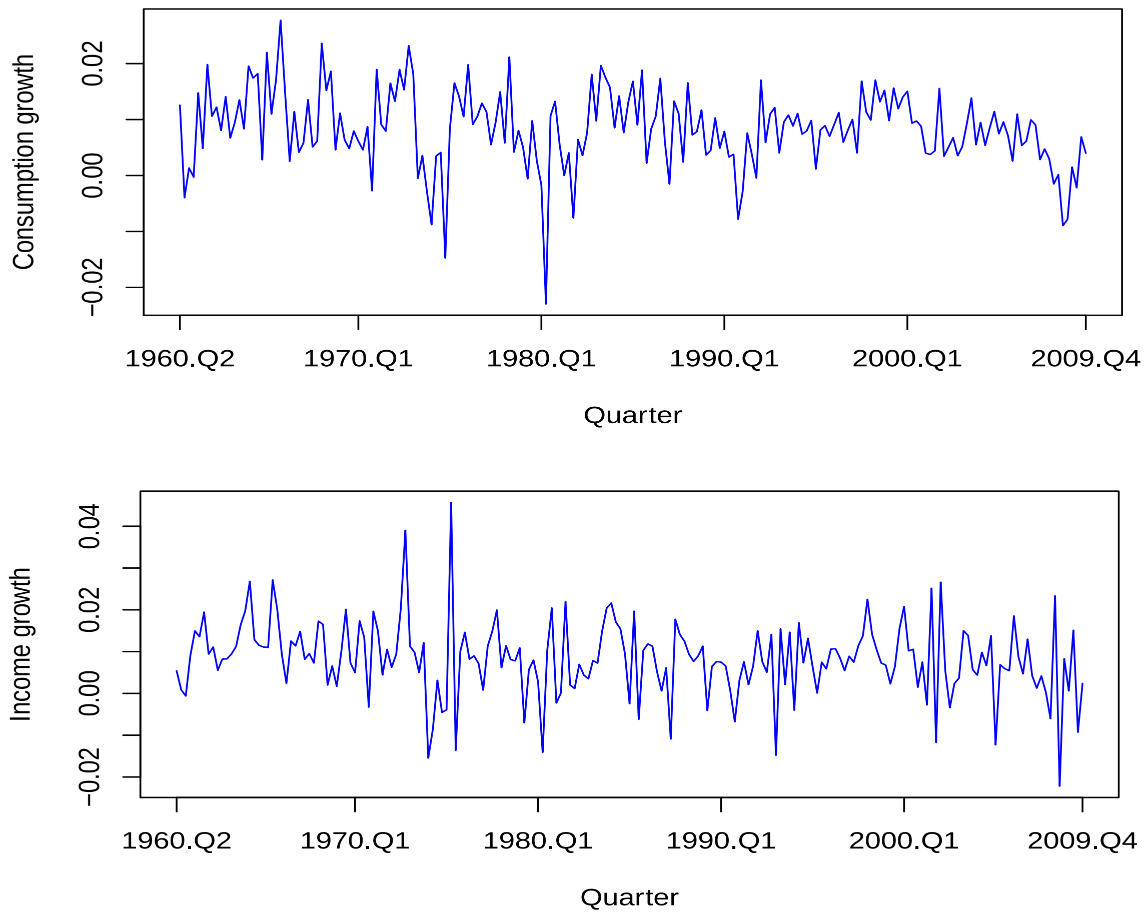
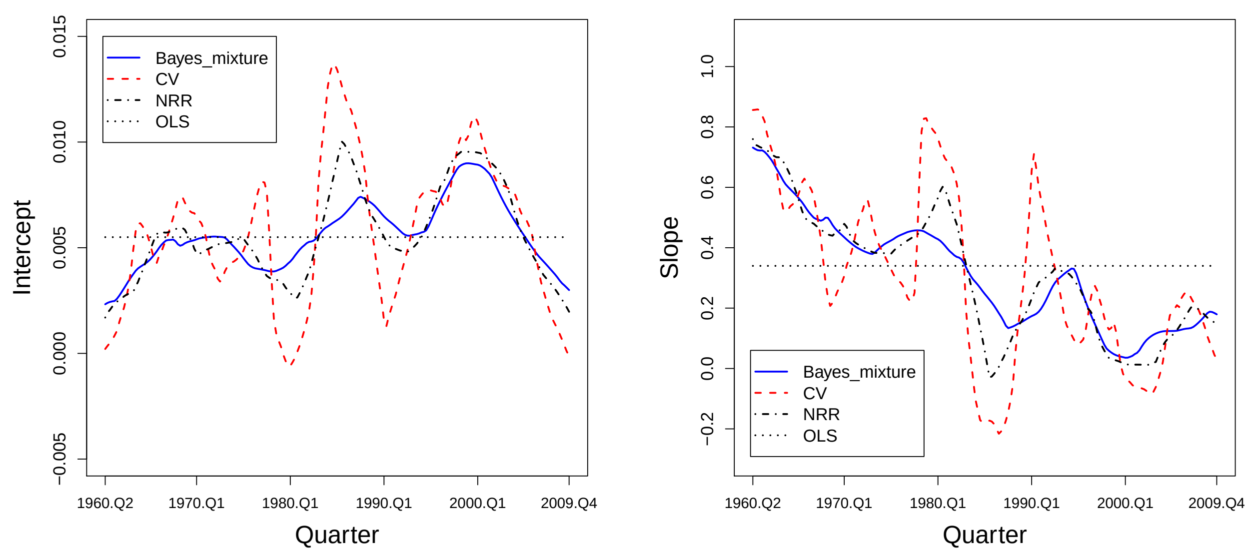


Figure 6: Local constant estimates of $\beta_1(\tau)$ and $\beta_2(\tau)$ with bandwidth derived through NRR, CV and Bayesian sampling.



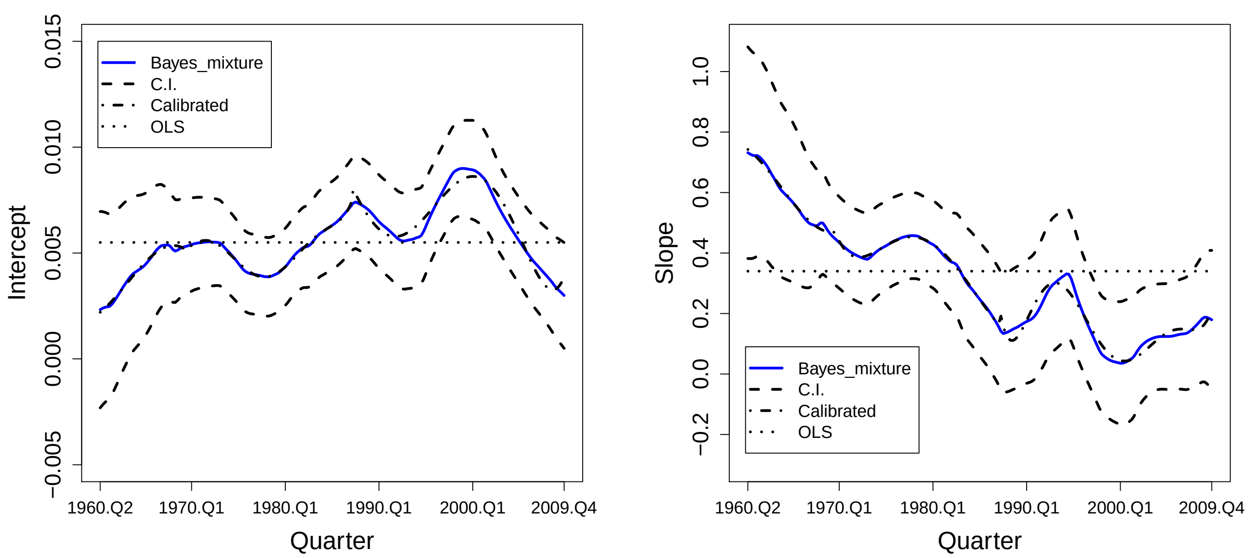
specific segment during the sample period, the slope coefficient exhibited up and down due to different rates of economy growth.

The slope coefficient can be interpreted as the marginal propensity to consume (MPC). A possible explanation is that in aggregation, households increase their spending as their disposable income rises and households consume a smaller and smaller proportion of disposable income as disposable income increases. The MPC is the ratio of a change in consumption over a change in income that caused such a consumption change. From the data, we found that disposable income demonstrated an increasing trend, therefore the MPC or the slope coefficient had a long–term decreasing trend.

The mean squared error (MSE) of residuals is 0.3393×10^{-4} for the time–varying coefficient model, and is 0.3872×10^{-4} for the constant–coefficient model. The introduction of time–varying coefficients improves model fitting by 12.37% in comparison to the constant coefficient model.

We derived the asymptotic 95% point–wise confidence intervals of the time–varying coefficients, which are plotted in Figure 7. The construction of the confidence interval is same as the one for the time–varying coefficients in Okun's law.

Figure 7: Local constant estimates of $\beta_1(\tau)$ and $\beta_2(\tau)$ and the calibrated parametric functions with 95% point–wise confidence intervals



In order to examine whether the coefficients are time–varying, we tested the null hypothesis of constant coefficients against the alternative of time–varying coefficients. The test procedure is the same as the one used in Section 4.1 and is described in Appendix B. The proposed bootstrap test with 1000 repetitions produced a p value that is approximately 0. Therefore, we rejected the null hypothesis of constant coefficients in the linear regression model at the 1% significance level.

According to the patterns of the estimated time–varying coefficients shown in Figure 6, a piecewise polynomial function of time might be appropriate to approximate each coefficient function. Consider the time–varying coefficient model given by

$$y_t = \beta_1^*(\tau_t) + \beta_2^*(\tau_t)x_t + u_t, \tag{9}$$

where $\beta_j^*(\tau_t)$, for $j = 1$ and 2, are known–form functions of τ_t. Fitting this model to the sample, we derived the coefficient estimates of the time–varying function. The two estimated time–varying functions are

$$\begin{aligned}
\beta_1^*(\tau_t) &= \left(0.0021 + 0.0163\tau_t + 0.1682\tau_t^2 - 0.8959\tau_t^3\right) I(1 \le t \le 38) \\
&+ \left(-0.0443 + 0.6412\tau_t - 2.9231\tau_t^2 + 5.5466\tau_t^3 - 3.7168\tau_t^4\right) I(39 \le t \le 108) \\
&+ \left(1.1570 - 6.0412\tau_t + 11.6796\tau_t^2 - 9.8480\tau_t^3 + 3.0556\tau_t^4\right) I(109 \le t \le 199), \\
\beta_2^*(\tau_t) &= \left(0.7481 - 1.0483\tau_t - 13.0197\tau_t^2 + 56.8208\tau_t^3\right) I(1 \le t \le 38) \\
&+ \left(0.7540 - 4.2510\tau_t + 16.4953\tau_t^2 - 19.8881\tau_t^3\right) I(39 \le t \le 108) \\
&+ \left(-11.5819 + 48.3087\tau_t - 64.7049\tau_t^2 + 28.2402\tau_t^3\right) I(109 \le t \le 199),
\end{aligned}$$

where $I(\cdot)$ is an indicator function. The p values of all estimated coefficients are zeros, indicating that all coefficients are significant at the 1% significance level.

The graphs of the above–calibrated coefficient functions are presented in Figure 7, where we could find that each calibrated coefficient function is very close to its corresponding local constant estimate. Figure 7 also shows that the 95% point–wise confidence intervals covered the calibrated time–varying coefficients, respectively. Therefore, the calibrated coefficients are appropriate to capture the time–varying nature of the coefficients.

In order to validate the calibrated piecewise functions for the time–varying coefficients, we tested the null hypothesis that the time–varying coefficients are piecewise polynomial functions of time against the alternative of local constant estimates. The test procedure is the same as the one used in Section 4.1. The proposed bootstrap test with 1000 repetitions produced a p value of 0.9930. Therefore, we could not reject the null hypothesis at the 1% significance level.

5 Conclusions and discussion

In this paper, we have proposed a Bayesian approach to bandwidth estimation for local constant estimation of time–varying coefficient time series models. We have established a completely new large sample theory for the proposed Bayesian bandwidth estimator as well as Bayesian estimators of the unknown parameters involved in the error density. From Monte Carlo simulation studies, we have found that our proposed Bayesian bandwidth estimator has good finite sample performance and it can achieve better performance than NRR and CV in estimating the bandwidths for local constant estimators in the regression function. Applying the proposed Bayesian method to the estimation of bandwidths for the time–varying coefficient regression models that reflect Okun's law and the relationship between consumption growth and income growth, we have found that the proposed Bayesian method works very well for the time–varying coefficient time series model. Furthermore, based on the corresponding nonparametric estimates, we have proposed parametric functions as approximations to time–varying Okun's coefficient and time–varying slope coefficient of consumption growth model.

In this study, we assume an AR(1) error process to take into account serial correlation. It would be straightforward to extend this error structure to more general cases. In addition, we will take one step further to accommodate heteroscedasticity in our future work. In terms of nonparametric estimation method, it would be natural to extend the local constant estimation to local polynomial estimation, such as the local linear estimation method. Since a local polynomial kernel estimation method involves much more technicalities than the local constant kernel method, and the proofs of the main results are already very technical, we wish to leave such extensions to future research. One of the advantages of using a local polynomial kernel method is that certain derivatives of the $\beta(\cdot)$ functions may be estimated naturally. As a result, it is ready to propose a plug–in method.

Another important issue is the assumption on $f(v;\eta)$. In this paper, we assume the density $f(v;\eta)$ is parametrically unknown possibly indexed by a vector of unknown parameters, and we then employ the proposed Bayesian method to estimate the unknown parameters. In future research, we will consider the case when $f(v)$ is nonparametrically unknown. In this case, we may estimate

$f(v)$ by a kernel density estimator of the form:

$$\widehat{f}(v;b) = \frac{1}{n}\sum_{t=1}^{n}\frac{1}{b}L\left(\frac{v_t - v}{b}\right),$$

and we will see whether b may be estimated in a similar way to what has been done for h. In addition, it is possible to examine the out–of–sample performance of the time–varying coefficient models through some forecasting scheme. In empirical studies, it would then be interesting to investigate forecasting performance based on our calibrated parametric forms.

References

Cai, Z. (2002), 'A two–stage approach to additive time series models', *Statistica Neerlandica* **56**, 415–433.

Cai, Z. (2007), 'Trending time–varying coefficient time series models with serially correlated errors', *Journal of Econometrics* **136**, 163–188.

Cai, Z., Fan, J. and Yao, Q. (2000), 'Functional–coefficient regression models for nonlinear time series', *Journal of the American Statistical Association* **95**, 941–956.

Cai, Z. and Tiwari, R. C. (2000), 'Application of a local linear autoregressive model to BOD time series', *Environmetrics* **11**, 341–350.

Campbell, J. Y. and Mankiw, N. G. (1989), Consumption, income, and interest rates: Reinterpreting the time series evidence, *in* O. J. Blanchard and S. Fischer, eds, 'NBER Macroeconomics Annual 1989', MIT Press, pp. 185–246.

Carrol, C. D. and Summers, L. H. (1991), Consumption growth parallels income growth: Some new evidence, *in* B. D. Bernheim and J. B. Shoven, eds, 'National Saving and Economic Performan', University of Chicago Press, pp. 305–348.

Chen, C.-F. (1985), 'On asymptotic normality of limiting density functions with Bayesian implications', *Journal of the Royal Statistical Society. Series B (Methodological)* **47**, 540–546.

Chen, J., Gao, J. and Li, D. (2012), 'Semiparametric trending panel data models with cross–sectional dependence', *Journal of Econometrics* **171**, 71–85.

Fan, J., Heckman, N. E. and Wand, M. P. (1995), 'Local polynomial kernel regression for generalized linear models and quasi–likelihood functions', *Journal of the American Statistical Association* **90**, 141–150.

Gao, J. (2007), *Nonlinear Time Series: Semi– and Non–Parametric Methods*, Chapman & Hall/CRC, London.

Gao, J. and Hawthorne, K. (2006), 'Semiparametric estimation and testing of the trend of temperature series', *The Econometrics Journal* **9**, 332–355.

Ghysels, E. (1998), 'On stable factor structures in the pricing of risk: Do time–varying betas help or hurt?', *The Journal of Finance* **53**, 549–573.

Jagannathan, R. and Wang, Z. (1996), 'The conditional CAPM and the cross–section of expected returns', *The Journal of Finance* **51**, 3–53.

Kim, J.-Y. (1998), 'Large sample properties of posterior densities, Bayesian information criterion and the likelihood principle in nonstationary time series models', *Econometrica* **66**, 359–380.

Lee, J. (2000), 'The robustness of Okun's law: Evidence from OECD countries', *Journal of Macroeconomics* **22**, 331–356.

Li, D., Chen, J. and Gao, J. (2011), 'Non–parametric time–varying coefficient panel data models with fixed effects', *The Econometrics Journal* **14**, 387–408.

Okun, A. (1962), 'Potential GNP: Its measurement and significance', *Proceedings of the Business and Economic Statistics Section of the American Statistical Association* pp. 89–104.

Phillips, P. C. B. and Ploberger, W. (1996), 'An asymtotic theory of Bayesian inference for time series', *Econometrica* **64**, 381–412.

Robinson, P. M. (1989), Nonparametric estimation of time–varying parameters, *in* P. Hackl, ed., 'Statistical Analysis and Forecasting of Economis Structural Change', Springer, Berlin, pp. 253–264.

Silvapulle, P., Moosa, I. A. and Silvapulle, M. J. (2004), 'Asymmetry in Okun's law', *Canadian Journal of Economics* **37**, 353–374.

Walker, A. (1969), 'On the asymptotic behaviour of posterior distributions', *Journal of the Royal Statistical Society. Series B (Methodological)* **31**, 80–88.

Wang, K. Q. (2003), 'Asset pricing with conditioning information: A new test', *The Journal of Finance* **58**, 161–196.

Zhang, X., King, M. L. and Hyndman, R. J. (2006), 'A Bayesian approach to bandwidth selection for multivariate kernel density estimation', *Computational Statistics and Data Analysis* **50**, 3009–3031.

Appendix A1: Assumptions

In this appendix, we describe the conditions that are required to establish the asymptotic results, although some of them might not be the weakest possible.

Assumption 1. (i) Let $\{x_t\}$ be strictly stationary and α–mixing with mixing coefficient satisfying $\rho(s) \le c s^{-c_0}$ for $0 < c < \infty$, $2 < c_0 < 4$ and s large enough. Let $\Sigma_x = \mathbb{E}[x_1 x_1^\top]$ be positive definite and $\mathbb{E}[\|x_1 x_1^\top\|^4] < \infty$.

(ii) Consider the case where $h = a_n \lambda$. Let λ be a continuous random variable with $\pi_\lambda(\cdot)$ being its density such that $\int \lambda^{-1}\pi_\lambda(\lambda)d\lambda < \infty$, $\int \lambda^{1-c_0/2}\pi_\lambda(\lambda)d\lambda < \infty$ and $\int \lambda^{2-c_0/2}\pi_\lambda(\lambda)d\lambda < \infty$ for $2 < c_0 < 4$, where c_0 is same as that in (i).

(iii) As $n \to \infty$, $a_n \to 0$ and $na_n \to \infty$.

Assumption 2. (i) $K(u)$ is a symmetric probability kernel function with $K(0) > 0$, $\int K(u)du = 1$, $\int u^2 K(u)du < \infty$, $\int K^2(u)du < \infty$, $\int |u|^{1-\frac{c_0}{2}} K(u)du < \infty$ and $\int |u|^{2-\frac{c_0}{2}} K(u)du < \infty$ for $2 < c_0 < 4$, where c_0 is same as that in Assumption 1(i).

(ii) $\beta_j(\cdot)$, for $j = 1, 2, \cdots, k$, are twice differentiable and each of the second derivative of $\beta_j(\cdot)$ is continuous in [0,1].

Assumption 3. (i) Let $\{v_t\}$ be an i.i.d. continuous random variables with $\mathbb{E}_\theta[v_1] = 0$ and $0 < \mathbb{E}_\theta[v_1^2] < \infty$. In addition, v_t and x_t are independent of each other.

(ii) Let $f(v;\theta)$ be the conditional density of v_1 given θ. Suppose that $f(v;\theta)$ is differentiable with respect to v. Let $\gamma(v;\theta) = \frac{f^{(1)}(v;\theta)}{f(v;\theta)}$ with $\mathbb{E}_\theta[\gamma(v_1;\theta)] = 0$, $\mathbb{E}_\theta[v_1\gamma(v_1;\theta)] \neq 0$ and $\mathbb{E}_\theta[\gamma^4(v_1;\theta)] < \infty$.

(iii) As $n \to \infty$, $\mathbb{E}[v_1\gamma_n(v_1)] \to C_\gamma$ for $|C_\gamma| < \infty$ and $\sqrt{n}\,\sigma_n^{-1} a_n \to \infty$, where $\sigma_n^2 = \mathbb{E}[w_1^2] \cdot \mathbb{E}\left[\left(v_1\gamma_n(v_1) - \mathbb{E}[v_1\gamma_n(v_1)]\right)^2\right]$, in which $\gamma_n(v_t) = \int (\mathbb{E}[\theta] - \theta)\gamma(v_t;\theta) A_n(\theta)\pi(\theta)d\theta$ with $A_n(\theta) = \frac{e^{n\mathbb{E}_\theta[\log f(v_1;\theta)]}}{\int e^{n\mathbb{E}_\theta[\log f(v_1;\theta)]}\pi(\theta)d\theta}$, $\mathbb{E}[w_1^2] = \mathbb{E}[\left(x_1^\top \Sigma_x^{-1} x_1\right)^2]$ and $\mathbb{E}_\theta[\log f(v_1;\theta)] = \int \log f(v;\theta) f(v;\theta)dv$.

Assumption 4. As $n \to \infty$, $\mathbb{E}[v_1\phi_n(v_1)] \to C_\phi$ for $|C_\phi| < \infty$, $a_n\delta_n^2 \to 0$ and $\sqrt{n}\,\delta_n^{-1} a_n \to \infty$, where $\phi_n(v_t) = \int \gamma(v_t;\theta) A_n(\theta)\pi(\theta)d\theta$ and $\delta_n^2 = \mathbb{E}[w_1^2] \cdot \mathbb{E}\left[\left(v_1\phi_n(v_1) - \mathbb{E}[v_1\phi_n(v_1)]\right)^2\right]$.

Assumption 5. (i) $g_n(\theta)$ is twice differentiable with respect to θ in some neighborhood of $\widehat{\theta}_n$.

(ii) The density $\pi_\theta(\theta)$ is twice differentiable and the second derivative is continuous.

(iii) $\int_{\theta \in D_n} \theta f_n(\theta|X_n, Y_n)d\theta = o_P(\Delta_n^{-1/2}(\widehat{\theta}_n))$ and $\frac{\int_{D_n} e^{g_n(\theta)}\pi_\theta(\theta)d\theta}{\int e^{g_n(\theta)}\pi_\theta(\theta)d\theta} = o_P(\Delta_n^{-1/2}(\widehat{\theta}_n))$, where $D_n = \left\{\theta : \Delta_n^{1/2}(\widehat{\theta}_n)\|\widehat{\theta}_n - \theta\| \geq c\right\}$ for some positive c.

(iv) As $n \to \infty$, $\Delta_n^{1/2}(\widehat{\theta}_n)\left(\widehat{\theta}_n - \mathbb{E}[\theta]\right) \to_D \mathcal{N}(0, \Sigma_0)$, where Σ_0 is a positive definite matrix.

Remark: Assumption 1(i) assumes x_t is stationary and α–mixing, which can be satisfied by many linear and nonlinear time series. Assumption 1(ii) imposes some moment conditions on λ. Assumption 1(iii) is required for the consistency of local constant estimators. If we further assume that $\{v_s\}$ and $\{x_t\}$ are independent for all (s, t) with $\mathbb{E}[x_1] = 0$, we may relax c_0 to just $c_0 = 2$ in Assumption 1(i). This point will be made in the proofs of Theorems 1 and 2.

Assumption 2(i) is a standard assumption for kernel functions. The commonly used Gaussian and uniform kernel functions satisfy Assumption 2(i). Assumption 2(ii) imposes smoothness constraints on the coefficient functions $\beta_j(\cdot)$, which is commonly used in literature; see Cai (2007) and Robinson (1989).

Assumption 3(i)(ii) ensures that $f(v;\theta)$ is differentiable and the fourth moment of $\gamma(v;\theta)$ exists. The conditions of $\mathbb{E}_\theta[\gamma(v_1;\theta)] = 0$ and $\mathbb{E}_\theta[v_1\gamma(v_1;\theta)] \neq 0$ are automatically satisfied when $f(v;\theta)$ is the density function of a Normal random variable. As pointed out in the proofs of Theorems 1 and 2, we need not assume $\mathbb{E}_\theta[\gamma(v_1;\theta)] = 0$ when $\mathbb{E}[x_1] = 0$. Assumptions 3 and 4 are automatically satisfied when $f(v;\theta)$ is the density function of a Normal random variable of the form $f(v;\theta) = f(v;\eta) = \frac{1}{\sqrt{2\pi\,\eta}}\exp\left(-\frac{v^2}{2\eta}\right)$ and η has a density function of the form $\pi(\eta) = \exp\left(-\eta\right) I[\eta > 0]$.

Assumption 5 is similar to what has been used by Kim (1998). Assumption 5 (i)(ii) is used to make sure that we can do Taylor expansion for $g_n(\theta)$ and $\pi_\theta(\theta)$. Assumption 5(iii) is another commonly used one. For example, if the posterior density is a Normal density, then $\int_{\theta\in D_n}\theta f_n(\theta|X_n)d\theta = o_P(n^{-1/2})$ and this assumption is automatically satisfied. We illustrate Assumption 5 (iv) using the following example. Suppose we have $X_i = \theta + e_i$, $e_i \sim \mathcal{N}(0,1)$. Note that θ is a random variable in Bayesian analysis. If we estimate $\theta_0 = \mathbb{E}[\theta]$ by the sample mean, denoted as $\widehat{\theta}_n$, then this assumption is actually a quite standard result for a conventional maximum likelihood estimation method.

Appendix A2: Proofs of Theorems 1 and 2

In this paper, our objective is to estimate $(\lambda,\theta^\top)^\top$ by the conditional mean, which is given by

$$\mathbb{E}\left[(\lambda,\theta^\top)^\top|X_n,Y_n\right] = \frac{\iint(\lambda,\theta^\top)^\top e^{l_n(\lambda,\theta)}\pi_\lambda(\lambda)\pi_\theta(\theta)d\lambda d\theta}{\iint e^{l_n(\lambda,\theta)}\pi_\lambda(\lambda)\pi_\theta(\theta)d\lambda d\theta} = \begin{pmatrix}\mathbb{E}_\star[\lambda|X_n,Y_n]\\ \mathbb{E}_\star[\theta|X_n,Y_n]\end{pmatrix}.$$

Equivalently, we can write $\mathbb{E}_\star[\lambda|X_n,Y_n]$ and $\mathbb{E}_\star[\theta|X_n,Y_n]$ as follows:

$$\mathbb{E}_\star[\lambda|X_n,Y_n] = \frac{\iint \lambda e^{l_n(\lambda,\theta)}\pi_\lambda(\lambda)\pi_\theta(\theta)d\lambda d\theta}{\iint e^{l_n(\lambda,\theta)}\pi_\lambda(\lambda)\pi_\theta(\theta)d\lambda d\theta},$$
$$\mathbb{E}_\star[\theta|X_n,Y_n] = \frac{\iint \theta e^{l_n(\lambda,\theta)}\pi_\lambda(\lambda)\pi_\theta(\theta)d\lambda d\theta}{\iint e^{l_n(\lambda,\theta)}\pi_\lambda(\lambda)\pi_\theta(\theta)d\lambda d\theta}.$$

Since the Bayesian estimates $\left(\mathbb{E}_\star[\lambda|X_n,Y_n],\mathbb{E}_\star[\theta^\top|X_n,Y_n]\right)^\top$ do not have closed–form expressions, we therefore approximate them by $\widehat{\lambda}_{mn}$ and $\widehat{\theta}_{mn}$, respectively.

We will show the large sample properties of $\widehat{\lambda}_{mn}$ and $\mathbb{E}_\star[\lambda|X_n,Y_n]$ in the proof of Theorem 1 in Appendix C of the supplementary material.

The large sample properties of $\widehat{\theta}_{mn}$ and $\mathbb{E}_{\star}[\theta|X_n, Y_n]$ involved in Theorem 2 will also be given in Appendix C of the supplementary material.

Appendix A3: Proof of Theorem 3

Let $h_0 = \mathbb{E}[h]$ and $\lambda_0 = \mathbb{E}[\lambda]$. It is easy to find that $h_0 = a_n\lambda_0$. From Theorem 1, we have that $\widehat{h}_{mn} - h_0 = O_P(n^{-1/2}\delta_n)$. Therefore, by Assumption 2 (ii), we can get that $\widehat{\beta}(\tau;\widehat{h}_{mn}) - \widehat{\beta}(\tau;h_0) = O_P(n^{-1/2}\delta_n)$.

It follows from Robinson (1989) that $\sqrt{nh_0}\left(\widehat{\beta}(\tau;h_0) - \beta(\tau)\right) \to_D \mathcal{N}(0, \Sigma_\beta)$, where $\Sigma_\beta = \sigma_u^2 \cdot \int K^2(w)dw \cdot \Sigma_x^{-1}$, in which $\sigma_u^2 = \mathbb{E}[u_1^2]$.

Therefore, under Assumptions 1–4, as $n \to \infty$, we have

$$
\begin{aligned}
&\sqrt{nh_0}\left(\widehat{\beta}(\tau;\widehat{h}_{mn}) - \beta(\tau)\right) = \sqrt{nh_0}\left(\widehat{\beta}(\tau;\widehat{h}_{mn}) - \widehat{\beta}(\tau;h_0)\right) + \sqrt{nh_0}\left(\widehat{\beta}(\tau;h_0) - \beta(\tau)\right) \\
&= O_P\left(\sqrt{a_n\delta_n^2}\right) + \sqrt{nh_0}\left(\widehat{\beta}(\tau;h_0) - \beta(\tau)\right) = o_P(1) + \sqrt{nh_0}\left(\widehat{\beta}(\tau;h_0) - \beta(\tau)\right) \to_D \mathcal{N}(0, \Sigma_\beta), \qquad (10)
\end{aligned}
$$

which completes the proof of Theorem 3.

Appendix B: Nonparametric Specification Testing

To examine whether the coefficients are time–varying, we test the null hypothesis

$$\mathrm{H}_0 : \beta_1(\tau) = \alpha_1,\ \beta_2(\tau) = \alpha_2,$$

against the alternative hypothesis given by

$$\mathrm{H}_1 : \beta_1(\tau) \text{ and } \beta_2(\tau) \text{ have a nonparametric specification.}$$

The test statistic is defined as

$$TS = \frac{RSS_0 - RSS_1}{RSS_1}, \qquad (11)$$

where RSS_0 is the residual sum of square (RSS) under the null hypothesis, and RSS_1 is the RSS under the alternative hypothesis. The null hypothesis is rejected for a large value of TS. The p–value is computed by employing the following bootstrap procedure.

1) Estimate $\beta(\tau_t) = (\beta_1(\tau_t), \beta_2(\tau_t))^\top$ by local constant estimation method and get the resulting estimator $\widehat{\beta}(\tau_t)$, for $t = 1, 2, \cdots, n$.

2) For each $t = 1, 2, \cdots, n$, generate

$$y_t^* = x_t^\top \widehat{\beta}(\tau_t) + \varepsilon_t^*,$$

where $\varepsilon_t^* = \widehat{\sigma} u_t^*$, $\widehat{\sigma}$ is the standard deviation of $\{y_t - x_t^\top \widehat{\beta}(\tau_t) : t = 1, 2, \cdots, n\}$ and $\{u_t^*\}$ is a sequence of independent and identically distributed random variables drawn from a prespecified distribution with mean zero and unit variance, such as $\mathrm{N}(0, 1)$.

3) Use the data set $\{(y_t^*, x_t) : t = 1, 2, \cdots, n\}$ to estimate $\alpha = (\alpha_1, \alpha_2)^\top$ and $\beta(\cdot)$. Calculate the corresponding RSS_0^* and RSS_1^*, and then compute TS^*.

4) Repeat Steps 2) and 3) for B times to obtain the empirical distribution for TS^*. Then, the p–value of the test is computed by $\frac{1}{B}\sum_{i=1}^{B} I(TS_i^* \geq TS)$, where $I(\cdot)$ is an indicator function.

Online Supplementary Document

Appendix C: Proofs of Theorems 1 and 2

Observe that

$$\begin{aligned}\widehat{v}_t &= \widehat{u}_t - \alpha \widehat{u}_{t-1} = y_t - x_t^\top \widehat{\beta}(\tau_t;h) - \alpha\left(y_{t-1} - x_{t-1}^\top \widehat{\beta}(\tau_{t-1};h)\right)\\ &= u_t + x_t^\top(\beta(\tau_t) - \widehat{\beta}(\tau_t;h)) - \alpha\left(u_{t-1} + x_{t-1}^\top(\beta(\tau_{t-1}) - \widehat{\beta}(\tau_{t-1};h))\right)\\ &= u_t - \alpha u_{t-1} + \Delta(\tau_t;h) - \alpha\Delta(\tau_{t-1};h)\\ &= v_t + \Delta(\tau_t;h) - \alpha\Delta(\tau_{t-1};h) = v_t + \Gamma(\tau_t;\alpha,h),\end{aligned}$$

where $\Delta(\tau_t;h) = x_t^\top(\beta(\tau_t) - \widehat{\beta}(\tau_t;h))$ and $\Gamma(\tau_t;\alpha,h) = \Delta(\tau_t;h) - \alpha\Delta(\tau_{t-1};h)$.

Therefore, we can further show

$$\begin{aligned}l_n(\lambda,\theta) = \log L_n(\lambda,\theta) &= \sum_{t=1}^n \log f(\widehat{v}_t;\theta)\\ &= \sum_{t=1}^n \log f(v_t + \Gamma(\tau_t;\alpha,h);\theta)\\ &= \sum_{t=1}^n \log\left[\left(f(v_t;\theta) + f^{(1)}(v_t;\theta)\Gamma(\tau_t;\alpha,h)\right)(1+o_P(1))\right]\\ &= \sum_{t=1}^n \log\left[f(v_t;\theta)\left(1 + \frac{f^{(1)}(v_t;\theta)}{f(v_t;\theta)}\Gamma(\tau_t;\alpha,h)\right)(1+o_P(1))\right]\\ &= \sum_{t=1}^n \log f(v_t;\theta) + \sum_{t=1}^n \log\left(1 + \gamma(v_t;\theta)\Gamma(\tau_t;\alpha,h)\right) + o_P(1)\\ &= \sum_{t=1}^n \log f(v_t;\theta) + \sum_{t=1}^n \gamma(v_t;\theta)\Gamma(\tau_t;\alpha,h) + o_P(1),\end{aligned} \tag{12}$$

where $\gamma(v_t;\theta) = \frac{f^{(1)}(v_t;\theta)}{f(v_t;\theta)}$.

Thus, we have

$$\begin{aligned}L_n(\lambda,\theta) &= e^{\sum_{t=1}^n \log f(v_t;\theta)} e^{\sum_{t=1}^n \gamma(v_t;\theta)\Gamma(\tau_t;\alpha,h)} e^{o_P(1)}\\ &= (1+o_P(1)) \prod_{t=1}^n f(v_t;\theta)\left(\sum_{t=1}^n \gamma(v_t;\theta)\Gamma(\tau_t;\alpha,h) + 1\right)\\ &= (1+o_P(1))\, G_n(\theta)\left(\sum_{t=1}^n \gamma(v_t;\theta)\Gamma(\tau_t;\alpha,h) + 1\right),\end{aligned} \tag{13}$$

where $G_n(\theta) = \prod_{t=1}^n f(v_t;\theta)$.

It is easy to see that

$$L_n(\lambda,\theta) - G_n(\theta) = G_n(\theta)\sum_{t=1}^n \gamma(v_t;\theta)\Gamma(\tau_t;\alpha,h).$$

As $\Gamma(\tau_t;\alpha,h) = \Delta(\tau_t;h) - \alpha\Delta(\tau_{t-1};h)$, we have

$$\sum_{t=1}^n \gamma(v_t;\theta)\Gamma(\tau_t;\alpha,h) = \sum_{t=1}^n \gamma(v_t;\theta)\Delta(\tau_t;h) - \alpha\sum_{t=1}^n \gamma(v_t;\theta)\Delta(\tau_{t-1};h).$$

We investigate $\Delta(\tau_t;h)$ as follows:

$$\begin{aligned}
\Delta(\tau_t;h) &= x_t^\top(\beta(\tau_t)-\widehat{\beta}(\tau_t;h)) = x_t^\top\left(\beta(\tau_t)-[p_n(\tau_t;h)]^{-1}q_n(\tau_t;h)\right)\\
&= x_t^\top\left\{[p_n(\tau_t;h)]^{-1}[p_n(\tau_t;h)]\beta(\tau_t)-[p_n(\tau_t;h)]^{-1}\frac{1}{nh}\sum_{s=1}^n x_s(x_s^\top\beta(\tau_s)+u_s)K\left(\frac{\tau_s-\tau_t}{h}\right)\right\}\\
&= -x_t^\top[p_n(\tau_t;h)]^{-1}\frac{1}{nh}\sum_{s=1}^n x_s u_s K\left(\frac{\tau_s-\tau_t}{h}\right)\\
&+ x_t^\top[p_n(\tau_t;h)]^{-1}\frac{1}{nh}\sum_{s=1}^n x_s x_s^\top(\beta(\tau_t)-\beta(\tau_s))K\left(\frac{\tau_s-\tau_t}{h}\right) = -\Delta_1(\tau_t;h)+\Delta_2(\tau_t;h),
\end{aligned}$$

in which $p_n(\tau_t;h)=\frac{1}{nh}\sum_{s=1}^n x_s x_s^\top K\left(\frac{\tau_s-\tau_t}{h}\right)$, $\quad q_n(\tau_t;h)=\frac{1}{nh}\sum_{s=1}^n x_s y_s K\left(\frac{\tau_s-\tau_t}{h}\right)$,
$\Delta_1(\tau_t;h)=x_t^\top[p_n(\tau_t;h)]^{-1}\frac{1}{nh}\sum_{s=1}^n x_s u_s K\left(\frac{\tau_s-\tau_t}{h}\right)$, and
$\Delta_2(\tau_t;h)=x_t^\top[p_n(\tau_t;h)]^{-1}\frac{1}{nh}\sum_{s=1}^n x_s x_s^\top(\beta(\tau_t)-\beta(\tau_s))K\left(\frac{\tau_s-\tau_t}{h}\right)$.

Therefore, we have

$$\sum_{t=1}^n \gamma(v_t;\theta)\Delta(\tau_t;h) = -\sum_{t=1}^n \gamma(v_t;\theta)\Delta_1(\tau_t;h)+\sum_{t=1}^n \gamma(v_t;\theta)\Delta_2(\tau_t;h). \tag{14}$$

Note that $p_n(\tau;h)$ can be expressed as

$$\begin{aligned}
p_n(\tau;h) &= \frac{1}{nh}\sum_{s=1}^n x_s x_s^\top K\left(\frac{\tau_s-\tau}{h}\right) = (1+o(1))\mathbb{E}[x_1x_1^\top]\frac{1}{nh}\sum_{s=1}^n K\left(\frac{\tau_s-\tau}{h}\right)\\
&= (1+o(1))\mathbb{E}[x_1x_1^\top]\int\frac{1}{h}K\left(\frac{u-\tau}{h}\right)du = (1+o(1))\mathbb{E}[x_1x_1^\top]\int_{-\frac{\tau}{h}}^{\frac{1-\tau}{h}} k(v)dv.
\end{aligned}$$

As we assume that $h=a_n\lambda$ with $a_n\to 0$ as $n\to\infty$ and λ being a continuous random variable, so $h\to 0$ as $n\to\infty$. Therefore, $\int_{-\frac{\tau}{h}}^{\frac{1-\tau}{h}} k(v)dv = 1+o(1)$ and

$$p_n(\tau;h) = (1+o(1))\mathbb{E}[x_1x_1^\top] = \Sigma_x(1+o(1)).$$

Therefore, it follows that

$$\begin{aligned}
\sum_{t=1}^n \gamma(v_t;\theta)\Delta_1(\tau_t;h) &= \sum_{t=1}^n \gamma(v_t;\theta)x_t^\top[p_n(\tau_t;h)]^{-1}\frac{1}{nh}\sum_{s=1}^n x_s u_s K\left(\frac{\tau_s-\tau_t}{h}\right)\\
&= \sum_{t=1}^n \gamma(v_t;\theta)x_t^\top\Sigma_x^{-1}\frac{1}{nh}\sum_{s=1}^n x_s u_s K\left(\frac{\tau_s-\tau_t}{h}\right)\\
&= \frac{1}{nh}\sum_{t=1}^n \gamma(v_t;\theta)x_t^\top\Sigma_x^{-1}\sum_{s=1}^n x_s u_s K\left(\frac{\tau_s-\tau_t}{h}\right)\\
&= \frac{K(0)}{nh}\sum_{t=1}^n \gamma(v_t;\theta)x_t^\top\Sigma_x^{-1}x_t u_t + \frac{1}{nh}\sum_{s=1}^n\sum_{t=1,t\neq s}^n \gamma(v_t;\theta)x_t^\top\Sigma_x^{-1}x_s u_s K\left(\frac{\tau_s-\tau_t}{h}\right)\\
&= I_{nt}(h)+J_{nt}(h),
\end{aligned}$$

where $I_{nt}(h)=\frac{K(0)}{nh}\sum_{t=1}^n\gamma(v_t;\theta)x_t^\top\Sigma_x^{-1}x_t u_t$ and $J_{nt}(h)=\frac{1}{nh}\sum_{s=1}^n\sum_{t=1,t\neq s}^n\gamma(v_t;\theta)x_t^\top\Sigma_x^{-1}x_s u_s K\left(\frac{\tau_s-\tau_t}{h}\right)$.

Since the error term u_t follows an AR(1) process, we can further express $I_{nt}(h)$ and $J_{nt}(h)$ as follows:

$$
\begin{aligned}
I_{nt}(h) &= \frac{K(0)}{nh}\sum_{t=1}^{n}\gamma(v_t;\theta)x_t^\top\Sigma_x^{-1}x_t u_t = \frac{K(0)}{nh}\sum_{t=1}^{n}\gamma(v_t;\theta)x_t^\top\Sigma_x^{-1}x_t(\alpha u_{t-1}+v_t)\\
&= \alpha\frac{K(0)}{nh}\sum_{t=1}^{n}\gamma(v_t;\theta)x_t^\top\Sigma_x^{-1}x_t u_{t-1} + \frac{K(0)}{nh}\sum_{t=1}^{n}\gamma(v_t;\theta)x_t^\top\Sigma_x^{-1}x_t v_t,\\
J_{nt}(h) &= \frac{1}{nh}\sum_{s=1}^{n}\sum_{t=1,t\neq s}^{n}\gamma(v_t;\theta)x_t^\top\Sigma_x^{-1}x_s u_s K\Big(\frac{\tau_s-\tau_t}{h}\Big)\\
&= \frac{1}{nh}\sum_{s=1}^{n}\sum_{t=1,t\neq s}^{n}\gamma(v_t;\theta)x_t^\top\Sigma_x^{-1}x_s(\alpha u_{s-1}+v_s)K\Big(\frac{\tau_s-\tau_t}{h}\Big)\\
&= \alpha\frac{1}{nh}\sum_{s=1}^{n}\sum_{t=1,t\neq s}^{n}\gamma(v_t;\theta)x_t^\top\Sigma_x^{-1}x_s u_{s-1}K\Big(\frac{\tau_s-\tau_t}{h}\Big)\\
&+ \frac{1}{nh}\sum_{s=1}^{n}\sum_{t=1,t\neq s}^{n}\gamma(v_t;\theta)x_t^\top\Sigma_x^{-1}x_s v_s K\Big(\frac{\tau_s-\tau_t}{h}\Big).
\end{aligned}
$$

Similarly, we can get

$$
\alpha\sum_{t=1}^{n}\gamma(v_t;\theta)\Delta(\tau_{t-1};h) = -\alpha\sum_{t=1}^{n}\gamma(v_t;\theta)\Delta_1(\tau_{t-1};h) + \alpha\sum_{t=1}^{n}\gamma(v_t;\theta)\Delta_2(\tau_{t-1};h), \tag{15}
$$

where $\sum_{t=1}^{n}\gamma(v_t;\theta)\Delta_1(\tau_{t-1};h) = I_{n(t-1)}(h) + J_{n(t-1)}(h)$, $I_{n(t-1)}(h) = \frac{K(0)}{nh}\sum_{t=1}^{n}\gamma(v_t;\theta)x_{t-1}^\top\Sigma_x^{-1}x_{t-1}u_{t-1}$ and $J_{n(t-1)}(h) = \frac{1}{nh}\sum_{s=1}^{n}\sum_{t-1\neq s}^{n}\gamma(v_t;\theta)x_{t-1}^\top\Sigma_x^{-1}x_s u_s K\big(\frac{\tau_s-\tau_{t-1}}{h}\big)$.

Observe that

$$
\begin{aligned}
&\sum_{t=1}^{n}\gamma(v_t;\theta)\Delta_1(\tau_t;h) - \alpha\sum_{t=1}^{n}\gamma(v_t;\theta)\Delta_1(\tau_{t-1};h)\\
&= \frac{K(0)}{nh}\sum_{t=1}^{n}v_t\gamma(v_t;\theta)x_t^\top\Sigma_x^{-1}x_t + \frac{1}{nh}\sum_{s=1}^{n}\sum_{t=1,t\neq s}^{n}\gamma(v_t;\theta)x_t^\top\Sigma_x^{-1}x_s v_s K\Big(\frac{\tau_s-\tau_t}{h}\Big) + o_P(1)\\
&= Q_{1n} + Q_{2n} + o_P(1),
\end{aligned}
$$

where $Q_{1n} = \frac{K(0)}{nh}\sum_{t=1}^{n}v_t\gamma(v_t;\theta)x_t^\top\Sigma_x^{-1}x_t$ and $Q_{2n} = \frac{1}{nh}\sum_{s=1}^{n}\sum_{t=1,t\neq s}^{n}\gamma(v_t;\theta)x_t^\top\Sigma_x^{-1}x_s v_s K\big(\frac{\tau_s-\tau_t}{h}\big)$.

As $n\to\infty$, the law of large numbers implies

$$
hQ_{1n} = K(0)\frac{1}{n}\sum_{t=1}^{n}v_t\gamma(v_t;\theta)x_t^\top\Sigma_x^{-1}x_t \to_P K(0)\cdot\mathbb{E}_\theta[\gamma(v_1;\theta)v_1]\cdot\mathbb{E}[w_1]\neq 0,
$$

where $w_t = x_t^\top\Sigma_x^{-1}x_t$. So $Q_{1n} = O_P(1/h)$.

We now show that Q_{2n} is a higher–order term than Q_{1n}. Because $\mathbb{E}_\theta[v_1] = \mathbb{E}_\theta[\gamma(v_1;\theta)] = 0$, we have

$$
\begin{aligned}
\mathbb{E}_{\theta,\lambda}\left[Q_{2n}^2\right] &= \frac{1}{n^2h^2}\mathbb{E}_{\theta,\lambda}\left[\sum_{t=1}^{n}\left(\sum_{s=1,\neq t}^{n}\gamma(v_t;\theta)x_t^\top\Sigma_x^{-1}x_s v_s K\left(\frac{\tau_s-\tau_t}{h}\right)\right)\right]^2 \\
&= \frac{1}{n^2h^2}\sum_{t=1}^{n}\left(\sum_{s=1,\neq t}^{n}\mathbb{E}_\theta\left[\gamma^2(v_t;\theta)\right]\mathbb{E}_\theta[v_s^2]\cdot\mathbb{E}[w_s w_t]K^2\left(\frac{\tau_s-\tau_t}{h}\right)\right) \\
&\leq \frac{C_{vw}}{n^2h^2}\sum_{t=1}^{n}\sum_{s=1,t\neq s}^{n}K^2\left(\frac{\tau_s-\tau_t}{h}\right) = \frac{C_{vw}}{nh^2}\sum_{t=1}^{n}\frac{1}{n}\sum_{s=1,t\neq s}^{n}K^2\left(\frac{\tau_s-\tau_t}{h}\right) \\
&= \frac{C_{vw}(1+o(1))}{nh^2}\sum_{t=1}^{n}\int_0^1 K^2\left(\frac{u-\tau_t}{h}\right)du = \frac{C_{vw}}{nh}\sum_{t=1}^{n}\int_{-\tau_t/h}^{1-\tau_t/h}K^2(v)dv \\
&= \frac{C_{vw}}{h}\frac{1}{n}\sum_{t=1}^{n}m(\tau_t;h) = \frac{C_{vw}(1+o(1))}{h}\int K^2(u)du,
\end{aligned}
$$

where $C_{vw} > 0$ is some constant and $m(\tau;h) = \int_{-\tau/h}^{1-\tau/h}K^2(u)du = (1+o(1))\int K^2(u)du$. So we have $Q_{2n} = O_P(1/\sqrt{h})$. This deduces that $Q_{1n}+Q_{2n} = O_P(1/h)\left(1+O_P(\sqrt{h})\right)$, which implies that Q_{1n} is the leading term.

Meanwhile, we have

$$
\begin{aligned}
&\sum_{t=1}^{n}\gamma(v_t;\theta)\Delta_2(\tau_t;h) = \sum_{t=1}^{n}\gamma(v_t;\theta)x_t^\top[p_n(\tau_t;h)]^{-1}\frac{1}{nh}\sum_{s=1}^{n}x_s x_s^\top(\beta(\tau_t)-\beta(\tau_s))K\left(\frac{\tau_s-\tau_t}{h}\right) \\
&= \frac{1}{nh}\sum_{t=1}^{n}\gamma(v_t;\theta)x_t^\top\Sigma_x^{-1}\sum_{s=1}^{n}x_s x_s^\top K\left(\frac{\tau_s-\tau_t}{h}\right)(\beta(\tau_t)-\beta(\tau_s)) \\
&= \frac{1+o(1)}{nh}\sum_{t=1}^{n}\sum_{s=1,s\neq t}^{n}\gamma(v_t;\theta)x_t^\top\Sigma_x^{-1}x_s x_s^\top K\left(\frac{\tau_s-\tau_t}{h}\right)\left(\beta^{(1)}(\tau_t)(\tau_t-\tau_s)+1/2\beta^{(2)}(\tau_t)(\tau_t-\tau_s)^2\right) \\
&= \frac{1+o(1)}{n}\sum_{t=1}^{n}\sum_{s=1,s\neq t}^{n}\gamma(v_t;\theta)x_t^\top\Sigma_x^{-1}x_s x_s^\top K_1\left(\frac{\tau_s-\tau_t}{h}\right)\beta^{(1)}(\tau_t) \\
&+ \frac{(1+o(1))h}{2n}\sum_{t=1}^{n}\sum_{s=1,s\neq t}^{n}\gamma(v_t;\theta)x_t^\top\Sigma_x^{-1}x_s x_s^\top K_2\left(\frac{\tau_s-\tau_t}{h}\right)\beta^{(2)}(\tau_t) \\
&\equiv (1+o(1))\ Q_{3n} + (1+o(1))\ Q_{4n}. \qquad (16)
\end{aligned}
$$

We then show that $\mathbb{E}_{\theta,\lambda}[Q_{jn}] \to 0$ and $\mathbb{E}_{\theta,\lambda}\left(Q_{jn}-\mathbb{E}_{\theta,\lambda}[Q_{jn}]\right)^2 \to 0$ for $j = 3,4$. Obviously, we have

$$
\begin{aligned}
\mathbb{E}_{\theta,\lambda}[Q_{3n}] &= \frac{1}{n}\sum_{t=1}^{n}\sum_{s=1,s\neq t}^{n}\mathbb{E}_\theta\left[\gamma(v_t;\theta)x_t^\top\Sigma_x^{-1}x_s x_s^\top\right]K_1\left(\frac{\tau_s-\tau_t}{h}\right)\beta^{(1)}(\tau_t), \\
\mathbb{E}_{\theta,\lambda}[Q_{4n}] &= \frac{h}{2n}\sum_{t=1}^{n}\sum_{s=1,s\neq t}^{n}\mathbb{E}_\theta\left[\gamma(v_t;\theta)x_t^\top\Sigma_x^{-1}x_s x_s^\top\right]K_2\left(\frac{\tau_s-\tau_t}{h}\right)\beta^{(2)}(\tau_t),
\end{aligned}
$$

where $K_1(u) = uK(u)$ and $K_2(u) = u^2K(u)$.

Under Assumption 3(ii), by Lemma A.1 of Gao (2007), we obtain[3]

$$
\begin{aligned}
&\left\|\mathbb{E}_\theta\left[\gamma(v_t;\theta)x_t^\top\Sigma_x^{-1}x_sx_s^\top\right]-\mathbb{E}_\theta\left[\gamma(v_t;\theta)x_t^\top\Sigma_x^{-1}\right]\mathbb{E}\left[x_sx_s^\top\right]\right\|\\
&=\left\|\mathbb{E}_\theta\left[\gamma(v_t;\theta)x_t^\top\Sigma_x^{-1}x_sx_s^\top\right]-\mathbb{E}_\theta\left[\gamma(v_t;\theta)\right]\mathbb{E}\left[x_t^\top\Sigma_x^{-1}\right]\mathbb{E}\left[x_sx_s^\top\right]\right\|\\
&=\left\|\mathbb{E}_\theta\left[\gamma(v_t;\theta)x_t^\top\Sigma_x^{-1}x_sx_s^\top\right]\right\|\\
&\le\left(\mathbb{E}_\theta\left[\|\gamma(v_t;\theta)x_t^\top\Sigma_x^{-1}\|^p\right]\right)^{1/p}\left(\mathbb{E}\left[\|x_sx_s^\top\|^q\right]\right)^{1/q}\rho^{1-Q}(|s-t|),
\end{aligned}
$$

where $Q=\frac{1}{p}+\frac{1}{q}<1$. Without loss of generosity, we choose $p=q=4$. Then under Assumptions 1(i) and 3(ii), we have

$$
\begin{aligned}
&\left\|\mathbb{E}_\theta\left[\gamma(v_t;\theta)x_t^\top\Sigma_x^{-1}x_sx_s^\top\right]\right\|\le\left(\mathbb{E}_\theta\left[\|\gamma(v_t;\theta)x_t^\top\Sigma_x^{-1}\|^4\right]\right)^{1/4}\left(\mathbb{E}\left[\|x_sx_s^\top\|^4\right]\right)^{1/4}\rho^{1/2}(|s-t|)\\
&=\left(\mathbb{E}_\theta\left[\gamma^4(v_t;\theta)\right]\right)^{1/4}\left(\mathbb{E}\left[\|x_t^\top\Sigma_x^{-1}\|^4\right]\right)^{1/4}\left(\mathbb{E}\left[\|x_sx_s^\top\|^4\right]\right)^{1/4}\rho^{1/2}(|s-t|)\equiv A_0\,\rho^{1/2}(|s-t|),
\end{aligned}
$$

where $A_0=\left(\mathbb{E}_\theta\left[\gamma^4(v_t;\theta)\right]\right)^{1/4}\left(\mathbb{E}\left[\|x_t^\top\Sigma_x^{-1}\|^4\right]\right)^{1/4}\left(\mathbb{E}\left[\|x_sx_s^\top\|^4\right]\right)^{1/4}$.

Therefore, we obtain

$$
\begin{aligned}
\left|\sum_{t=1}^n\gamma(v_t;\theta)\Delta_2(\tau_t;h)\right|&\le\frac{M}{n}\sum_{t=1}^n\sum_{s=1,s\neq t}^n\left\|\mathbb{E}_\theta\left[\gamma(v_t;\theta)x_t^\top\Sigma_x^{-1}x_sx_s^\top\right]\right\|\left|K_1\left(\frac{\tau_s-\tau_t}{h}\right)\right|\left\|\beta^{(1)}(\tau_t)\right\|\\
&+\frac{Mh}{2n}\sum_{t=1}^n\sum_{s=1,s\neq t}^n\left\|\mathbb{E}_\theta\left[\gamma(v_t;\theta)x_t^\top\Sigma_x^{-1}x_sx_s^\top\right]\right\|K_2\left(\frac{\tau_s-\tau_t}{h}\right)\left\|\beta^{(2)}(\tau_t)\right\|\\
&\le\frac{M}{n}\sum_{t=1}^n\sum_{s=1,s\neq t}^n\rho^{1/2}(|s-t|)\left|K_1\left(\frac{\tau_s-\tau_t}{h}\right)\right|\left\|\beta^{(1)}(\tau_t)\right\|\\
&+\frac{Mh}{2n}\sum_{t=1}^n\sum_{s=1,s\neq t}^n\rho^{1/2}(|s-t|)K_2\left(\frac{\tau_s-\tau_t}{h}\right)\left\|\beta^{(2)}(\tau_t)\right\|\\
&=\frac{M}{n}\sum_{t=1}^n\sum_{s=1,s\neq t}^n\rho^{1/2}(n|\tau_s-\tau_t|)\left|K_1\left(\frac{\tau_s-\tau_t}{h}\right)\right|\left\|\beta^{(1)}(\tau_t)\right\|\\
&+\frac{Mh}{2n}\sum_{t=1}^n\sum_{s=1,s\neq t}^n\rho^{1/2}(n|\tau_s-\tau_t|)K_2\left(\frac{\tau_s-\tau_t}{h}\right)\left\|\beta^{(2)}(\tau_t)\right\|\\
&=Mn\,(1+o(1))\int_0^1\int_0^1\rho^{1/2}(n|v-u|)\left|K_1\left(\frac{v-u}{h}\right)\right|\left\|\beta^{(1)}(u)\right\|du\,dv\\
&+\frac{Mnh}{2}\,(1+o(1))\int_0^1\int_0^1\rho^{1/2}(n|v-u|)K_2\left(\frac{v-u}{h}\right)\left\|\beta^{(2)}(u)\right\|du\,dv\\
&=nhM\cdot A_{1n}\int_0^1\left\|\beta^{(1)}(u)\right\|du+\frac{Mnh^2}{2}\cdot A_{2n}\int_0^1\left\|\beta^{(2)}(u)\right\|du,
\end{aligned}
$$

where $A_{1n}=\int_{-\infty}^{+\infty}\rho^{1/2}(nh|w|)\,|K_1(w)|\,dw$, $A_{2n}=\int_{-\infty}^{+\infty}\rho^{1/2}(nh|w|)K_2(w)dw$ and M is a positive constant.

[3]As discussed in the remark about Assumption 1 in Appendix A1, we need only to require $\mathbb{E}_\theta\left[\gamma(v_t;\theta)x_t\right]=\mathbb{E}_\theta\left[\gamma(v_t;\theta)\right]\,\mathbb{E}[x_t]=0$ in an application of Lemma A.1 of Gao (2007).

Meanwhile, we will have $\mathbb{E}_\theta\left[\gamma(v_t;\theta)x_t^\top\Sigma_x^{-1}x_sx_s^\top\right]=\mathbb{E}_\theta\left[\gamma(v_t;\theta)\right]\mathbb{E}\left[x_t^\top\Sigma_x^{-1}x_sx_s^\top\right]=0$ if we assume that $\{v_t\}$ and $\{x_s\}$ are independent for all (s,t). In this case, there is no need to apply Lemma A.1 of Gao (2007) with the choice of (p,q). As a matter of fact, $\mathbb{E}_{\theta,\lambda}[Q_{jn}]=0$ and the derivation of $\mathbb{E}_{\theta,\lambda}\left[Q_{jn}\right]^2\to 0$ for $j=3,4$ follows similarly from that of $\mathbb{E}_{\theta,\lambda}[Q_{2n}]^2\to 0$.

Under Assumption 1(i), we have

$$
\begin{aligned}
|A_{1n}| &= \left|\int_{-\infty}^{+\infty} \rho^{1/2}(nh|w|)\,|K_1(w)|\,dw\right| \\
&\le \int_{-\infty}^{+\infty} \rho^{1/2}(nh|w|)|w|K(w)dw \\
&\le c(nh)^{-c_0/2}\int_{-\infty}^{+\infty} |w|^{-c_0/2}|w|K(w)dw \\
&\le c(nh)^{-c_0/2}\int_{-\infty}^{+\infty} |w|^{1-c_0/2}K(w)dw.
\end{aligned}
$$

Thus, under Assumptions 1(i)(iii) and 2(i), as $n\to\infty$, we have

$$
nh|A_{1n}| \le c(nh)^{1-c_0/2}\int_{-\infty}^{+\infty} |w|^{1-c_0/2}K(w)dw = O\big((nh)^{1-c_0/2}\big) \to 0.
$$

Similarly, we have

$$
\begin{aligned}
|A_{2n}| = \left|\int_{-\infty}^{+\infty} \rho^{1/2}(nh|w|)K_2(w)dw\right| &\le \int_{-\infty}^{+\infty} \rho^{1/2}(nh|w|)|w|^2K(w)dw \\
\le c(nh)^{-c_0/2}\int_{-\infty}^{+\infty} |w|^{2-c_0/2}K(w)dw.
\end{aligned}
$$

Hence, under Assumptions 1(i)(iii) and 2(i), as $n\to\infty$, we have

$$
nh^2|A_{2n}| \le cn^{1-c_0/2}h^{2-c_0/2}\int_{-\infty}^{+\infty} |w|^{2-c_0/2}K(w)dw = O\big(n^{1-c_0/2}h^{2-c_0/2}\big) \to 0.
$$

Thus, under Assumptions 1(i)(iii), 2(i) and 3(i)(ii), it follows that

$$
\left|\sum_{t=1}^{n} \gamma(v_t;\theta)\Delta_2(\tau_t;h)\right| \le nhM\int_0^1 \left\|\beta^{(1)}(u)\right\| du A_{1n} + \frac{nh^2}{2}M\int_0^1 \left\|\beta^{(2)}(u)\right\| du A_{2n} = o_P(1).
$$

Similarly, we can show that $\sum_{t=1}^n \gamma(v_t;\theta)\Delta_2(\tau_{t-1};h) = o_P(1)$.

Under Assumptions 1(ii)(iii), we therefore have

$$
\begin{aligned}
&\int \left|\sum_{t=1}^{n} \gamma(v_t;\theta)\Delta_2(\tau_t;h)\right| \pi_\lambda(\lambda)\pi_\theta(\theta)d\lambda d\theta \\
&\le \int \tilde{c}_1(nh)^{1-c_0/2}\pi_\lambda(\lambda)\pi_\theta(\theta)d\lambda d\theta + \int \tilde{c}_2 n^{1-c_0/2}h^{2-c_0/2}\pi_\lambda(\lambda)\pi_\theta(\theta)d\lambda d\theta \\
&= \tilde{c}_1 n^{1-c_0/2}a_n^{1-c_0/2}\int \lambda^{1-c_0/2}\pi_\lambda(\lambda)d\lambda + \tilde{c}_2 n^{1-c_0/2}a_n^{2-c_0/2}\int \lambda^{2-c_0/2}\pi_\lambda(\lambda)d\lambda \to 0,
\end{aligned}
$$

where $|\tilde{c}_1| < \infty$ and $|\tilde{c}_2| < \infty$.

In order to show that $\mathbb{E}_{\theta,\lambda}\left[\left(Q_{jn} - \mathbb{E}_{\theta,\lambda}[Q_{jn}^2]\right)^2\right] \to 0$ for $j = 3,4$, we need only to deal with a leading term of the form

$$
\frac{1}{n^2}\mathbb{E}_{\theta,\lambda}\left[\left(\sum_{t=1}^{n}\sum_{s=1,s\neq t}^{n} \left(\gamma(v_t;\theta)x_t^\top\Sigma_x^{-1}\right)\left(x_sx_s^\top - \mathbb{E}[x_sx_s^\top]\right)K_1\left(\frac{\tau_s-\tau_t}{h}\right)\beta^{(1)}(\tau_t)\right)^2\right] \to 0,
$$

which follows similarly from the derivation of $\mathbb{E}[Q_{2n}^2] \to 0$.

To sum up, we have

$$\begin{aligned}\sum_{t=1}^{n}\gamma(v_t;\theta)\Gamma(\tau_t;\alpha,h) &= \sum_{t=1}^{n}\gamma(v_t;\theta)\Delta(\tau_t;h)-\alpha\sum_{t=1}^{n}\gamma(v_t;\theta)\Delta(\tau_{t-1};h)\\ &= -\frac{K(0)}{nh}\sum_{t=1}^{n}v_t\gamma(v_t;\theta)x_t^\top\Sigma_x^{-1}x_t\,(1+o_P(1))\\ &= -\frac{K(0)\cdot(1+o_P(1))}{nh}\sum_{t=1}^{n}v_t\gamma(v_t;\theta)w_t,\end{aligned}$$

where $w_t = x_t^\top\Sigma_x^{-1}x_t$.

We are now ready to complete the proofs of Theorems 1 and 2.

Proof of Theorem 1.

Observe that

$$\mathbb{E}_\star[\lambda|X_n,Y_n] = \frac{\iint \lambda e^{l_n(\lambda,\theta)}\pi_\lambda(\lambda)\pi_\theta(\theta)d\lambda d\theta}{\iint e^{l_n(\lambda,\theta)}\pi_\lambda(\lambda)\pi_\theta(\theta)d\lambda d\theta} = \frac{\iint \lambda L_n(\lambda,\theta)\pi_\lambda(\lambda)\pi_\theta(\theta)d\lambda d\theta}{\iint L_n(\lambda,\theta)\pi_\lambda(\lambda)\pi_\theta(\theta)d\lambda d\theta} = \frac{q_1(X_n,Y_n)}{p_1(X_n,Y_n)}, \tag{17}$$

where $q_1(X_n,Y_n) = \iint \lambda L_n(\lambda,\theta)\pi_\lambda(\lambda)\pi_\theta(\theta)d\lambda d\theta$ and $p_1(X_n,Y_n) = \iint L_n(\lambda,\theta)\pi_\lambda(\lambda)\pi_\theta(\theta)d\lambda d\theta$.

$$\begin{aligned}\mathbb{E}[\lambda] &= \int\lambda\pi_\lambda(\lambda)d\lambda = \frac{\int\lambda\pi_\lambda(\lambda)d\lambda}{\int\pi_\lambda(\lambda)d\lambda} = \frac{\int\lambda\pi_\lambda(\lambda)d\lambda}{\int\pi_\lambda(\lambda)d\lambda}\frac{\int G_n(\theta)\pi_\theta(\theta)d\theta}{\int G_n(\theta)\pi_\theta(\theta)d\theta} = \frac{\iint\lambda G_n(\theta)\pi_\lambda(\lambda)\pi_\theta(\theta)d\lambda d\theta}{\iint G_n(\theta)\pi_\lambda(\lambda)\pi_\theta(\theta)d\lambda d\theta}\\ &= \frac{q(X_n,Y_n)}{p(X_n,Y_n)},\end{aligned} \tag{18}$$

where $q(X_n,Y_n) = \iint\lambda G_n(\theta)\pi_\lambda(\lambda)\pi_\theta(\theta)d\lambda d\theta$ and $p(X_n,Y_n) = \iint G_n(\theta)\pi_\lambda(\lambda)\pi_\theta(\theta)d\lambda d\theta$.

Based on (17) and (18), we have that

$$\begin{aligned} q_1(X_n,Y_n)-q(X_n,Y_n) &= \iint\lambda\,(L_n(\lambda,\theta)-G_n(\theta))\,\pi_\lambda(\lambda)\pi_\theta(\theta)d\lambda d\theta,\\ p_1(X_n,Y_n)-p(X_n,Y_n) &= \iint (L_n(\lambda,\theta)-G_n(\theta))\,\pi_\lambda(\lambda)\pi_\theta(\theta)d\lambda d\theta,\end{aligned}$$

Therefore, by equation (13) we have

$$\begin{aligned}&\mathbb{E}_\star[\lambda|X_n,Y_n]-\mathbb{E}[\lambda] = \frac{q_1(X_n,Y_n)}{p_1(X_n,Y_n)}-\frac{q(X_n,Y_n)}{p(X_n,Y_n)}\\ &= \frac{1}{p_1(X_n,Y_n)p(X_n,Y_n)}\left[q_1(X_n,Y_n)p(X_n,Y_n)-p_1(X_n,Y_n)q(X_n,Y_n)\right]\\ &= \frac{1}{p_1(X_n,Y_n)p(X_n,Y_n)}\left[(q_1(X_n,Y_n)-q(X_n,Y_n))p(X_n,Y_n)-(p_1(X_n,Y_n)-p(X_n,Y_n))q(X_n,Y_n)\right]\\ &= \frac{1}{p_1(X_n,Y_n)}\left[(q_1(X_n,Y_n)-q(X_n,Y_n))-(p_1(X_n,Y_n)-p(X_n,Y_n))\mathbb{E}[\lambda]\right]\\ &= \frac{1}{p_1(X_n,Y_n)}\iint(\lambda-\mathbb{E}[\lambda])\,(L_n(\lambda,\theta)-G_n(\theta))\,\pi_\lambda(\lambda)\pi_\theta(\theta)d\lambda d\theta\\ &= \frac{1+o_P(1)}{p_1(X_n,Y_n)}\iint(\lambda-\mathbb{E}[\lambda])\,G_n(\theta)\sum_{t=1}^{n}\gamma(v_t;\theta)\Gamma(\tau_t;\alpha,h)\pi_\lambda(\lambda)\pi_\theta(\theta)d\lambda d\theta\\ &= (1+o_P(1))\int\frac{G_n(\theta)}{p_1(X_n,Y_n)}\left(\int(\lambda-\mathbb{E}[\lambda])\sum_{t=1}^{n}\gamma(v_t;\theta)\Gamma(\tau_t;\alpha,h)\pi_\lambda(\lambda)d\lambda\right)\pi_\theta(\theta)d\theta,\end{aligned}$$

where

$$
\begin{aligned}
p_1(X_n, Y_n) &= \iint L_n(\lambda,\theta)\pi_\lambda(\lambda)\pi_\theta(\theta)\,d\lambda d\theta \\
&= \iint G_n(\theta)\left(\sum_{t=1}^{n}\gamma(v_t;\theta)\Gamma(\tau_t;\alpha,h)+1\right)\pi_\lambda(\lambda)\pi_\theta(\theta)\,d\lambda d\theta \\
&= \int G_n(\theta)\left(1+\int\sum_{t=1}^{n}\gamma(v_t;\theta)\Gamma(\tau_t;\alpha,h)\pi_\lambda(\lambda)\,d\lambda\right)\pi_\theta(\theta)\,d\theta.
\end{aligned}
$$

Therefore, it follows that

$$
\begin{aligned}
\mathbb{E}_\star[\lambda|X_n,Y_n]-\mathbb{E}[\lambda] &= \frac{\int G_n(\theta)\left(\int(\lambda-\mathbb{E}[\lambda])\sum_{t=1}^{n}\gamma(v_t;\theta)\Gamma(\tau_t;\alpha,h)\pi_\lambda(\lambda)\,d\lambda\right)\pi_\theta(\theta)\,d\theta}{\int G_n(\theta)\left(1+\int\sum_{t=1}^{n}\gamma(v_t;\theta)\Gamma(\tau_t;\alpha,h)\pi_\lambda(\lambda)\,d\lambda\right)\pi_\theta(\theta)\,d\theta} \\
&= \frac{\int G_n(\theta)R_{2n}(\lambda,\theta)\pi_\theta(\theta)\,d\theta}{\int G_n(\theta)R_{1n}(\lambda,\theta)\pi_\theta(\theta)\,d\theta},
\end{aligned}
$$

where $R_{1n}(\lambda,\theta)=1+\int\sum_{t=1}^{n}\gamma(v_t;\theta)\Gamma(\tau_t;\alpha,h)\pi_\lambda(\lambda)\,d\lambda$ and

$$
R_{2n}(\lambda,\theta)=\int(\lambda-\mathbb{E}[\lambda])\sum_{t=1}^{n}\gamma(v_t;\theta)\Gamma(\tau_t;\alpha,h)\pi_\lambda(\lambda)\,d\lambda.
$$

Equivalently, we have

$$
\int G_n(\theta)R_{1n}(\lambda,\theta)\pi_\theta(\theta)\,d\theta\cdot(\mathbb{E}_\star[\lambda|X_n,Y_n]-\mathbb{E}[\lambda])=\int G_n(\theta)R_{2n}(\lambda,\theta)\pi_\theta(\theta)\,d\theta. \tag{19}
$$

Divide (19) by $\int G_n(\theta)\pi_\theta(\theta)\,d\theta$. Then we have

$$
\frac{\int G_n(\theta)R_{1n}(\lambda,\theta)\pi_\theta(\theta)\,d\theta}{\int G_n(\theta)\pi_\theta(\theta)\,d\theta}\cdot(\mathbb{E}_\star[\lambda|X_n,Y_n]-\mathbb{E}[\lambda])=\frac{\int G_n(\theta)R_{2n}(\lambda,\theta)\pi_\theta(\theta)\,d\theta}{\int G_n(\theta)\pi_\theta(\theta)\,d\theta}. \tag{20}
$$

We now express $R_{2n}(\lambda,\theta)$ as follows:

$$
\begin{aligned}
R_{2n}(\lambda,\theta) &= \int(\lambda-\mathbb{E}[\lambda])\sum_{t=1}^{n}\gamma(v_t;\theta)\Gamma(\tau_t;\alpha,h)\pi_\lambda(\lambda)\,d\lambda \\
&= (1+o_P(1))\int(\lambda-\mathbb{E}[\lambda])\left(-\frac{K(0)}{nh}\sum_{t=1}^{n}\gamma(v_t;\theta)x_t^\top\Sigma_x^{-1}x_tv_t\right)\pi_\lambda(\lambda)\,d\lambda \\
&= -(1+o_P(1))K(0)\frac{1}{na_n}\sum_{t=1}^{n}\gamma(v_t;\theta)x_t^\top\Sigma_x^{-1}x_tv_t\int(\lambda-\mathbb{E}[\lambda])\frac{1}{\lambda}\pi_\lambda(\lambda)\,d\lambda \\
&= -(1+o_P(1))K(0)\frac{1}{na_n}\sum_{t=1}^{n}\gamma(v_t;\theta)x_t^\top\Sigma_x^{-1}x_tv_t\left(1-\mathbb{E}[\lambda]\mathbb{E}[\lambda^{-1}]\right)+o_P(1) \\
&= -(1+o_P(1))K(0)\left(1-\mathbb{E}[\lambda]\mathbb{E}[\lambda^{-1}]\right)\frac{1}{na_n}\sum_{t=1}^{n}\gamma(v_t;\theta)x_t^\top\Sigma_x^{-1}x_tv_t+o_P(1) \\
&= \frac{d_1}{na_n}\sum_{t=1}^{n}v_t\gamma(v_t;\theta)x_t^\top\Sigma_x^{-1}x_t \\
&= \frac{d_1(1+o_P(1))}{na_n}\sum_{t=1}^{n}v_t\gamma(v_t;\theta)w_t,
\end{aligned}
$$

where $d_1=-K(0)\left(1-\mathbb{E}[\lambda]\mathbb{E}[\lambda^{-1}]\right)$.

Since $1+o_P(1) \to_P 1$, the replacement of $1+o_P(1)$ by 1 does not change anything in the derivation of the limit distribution below. By ignoring the high order $o_P(1)$ in the following derivations, we therefore have

$$\begin{aligned}\int G_n(\theta) R_{2n}(\lambda,\theta)\pi_\theta(\theta)d\theta &= \int G_n(\theta)\left[\frac{d_1}{na_n}\sum_{t=1}^n \gamma(v_t;\theta)w_t v_t\right]\pi_\theta(\theta)d\theta \\ &= \frac{d_1}{na_n}\sum_{t=1}^n w_t v_t \int \gamma(v_t;\theta)G_n(\theta)\pi_\theta(\theta)d\theta. \end{aligned} \tag{21}$$

Divide (21) by $\int G_n(\theta)\pi_\theta(\theta)d\theta$. It then follows that

$$\frac{\int G_n(\theta)R_{2n}(\lambda,\theta)\pi_\theta(\theta)d\theta}{\int G_n(\theta)\pi_\theta(\theta)d\theta} = \frac{d_1}{na_n}\sum_{t=1}^n w_t v_t \frac{\int \gamma(v_t;\theta)G_n(\theta)\pi_\theta(\theta)d\theta}{\int G_n(\theta)\pi_\theta(\theta)d\theta}.$$

We know that

$$G_n(\theta) = \prod_{t=1}^n f(v_t;\theta) = e^{\sum_{t=1}^n \log f(v_t;\theta)} = e^{n\mathbb{E}_\theta[\log f(v_1;\theta)]}(1+o_P(1)), \tag{22}$$

where $\mathbb{E}_\theta[\log f(v_1;\theta)] = \int \log f(v_1;\theta) f(v;\theta)dv$.

Let $A_n(\theta) = \frac{e^{n\mathbb{E}_\theta[\log f(v_1;\theta)]}}{\int e^{n\mathbb{E}_\theta[\log f(v_1;\theta)]}\pi_\theta(\theta)d\theta}$ and $\phi_n(v_t) = \int \gamma(v_t;\theta)A_n(\theta)\pi_\theta(\theta)d\theta$. Then

$$\frac{\int G_n(\theta)R_{2n}(\lambda,\theta)\pi_\theta(\theta)d\theta}{\int G_n(\theta)\pi_\theta(\theta)d\theta} = \frac{d_1}{na_n}\sum_{t=1}^n w_t v_t \phi_n(v_t). \tag{23}$$

Meanwhile, we have

$$\begin{aligned}&\int G_n(\theta)R_{1n}(\lambda,\theta)\pi_\theta(\theta)d\theta = \int G_n(\theta)\left(1+\int\sum_{t=1}^n \gamma(v_t;\theta)\Gamma(\tau_t;\alpha,h)\pi_\lambda(\lambda)d\lambda\right)\pi_\theta(\theta)d\theta \\ &= \int G_n(\theta)\pi_\theta(\theta)d\theta + \int G_n(\theta)\left(\int\sum_{t=1}^n \gamma(v_t;\theta)\Gamma(\tau_t;\alpha,h)\pi_\lambda(\lambda)d\lambda\right)\pi_\theta(\theta)d\theta \\ &= \int G_n(\theta)\pi_\theta(\theta)d\theta + \int G_n(\theta)\left(\int\left(-\frac{K(0)}{nh}\sum_{t=1}^n \gamma(v_t;\theta)x_t^\top \Sigma_x^{-1}x_t v_t\right)\pi_\lambda(\lambda)d\lambda\right)\pi_\theta(\theta)d\theta \\ &= \int G_n(\theta)\pi_\theta(\theta)d\theta - K(0)\mathbb{E}[\lambda^{-1}]\frac{1}{na_n}\sum_{t=1}^n w_t v_t \int \gamma(v_t;\theta)G_n(\theta)\pi_\theta(\theta)d\theta. \end{aligned} \tag{24}$$

Divide (24) by $\int G_n(\theta)\pi_\theta(\theta)d\theta$. Then we have

$$\begin{aligned}\frac{\int G_n(\theta)R_{1n}(\lambda,\theta)\pi_\theta(\theta)d\theta}{\int G_n(\theta)\pi_\theta(\theta)d\theta} &= 1-K(0)\mathbb{E}[\lambda^{-1}]\frac{1}{na_n}\sum_{t=1}^n w_t v_t \frac{\int\gamma(v_t;\theta)G_n(\theta)\pi_\theta(\theta)d\theta}{\int G_n(\theta)\pi_\theta(\theta)d\theta} \\ &= 1-K(0)\mathbb{E}[\lambda^{-1}]\frac{1}{na_n}\sum_{t=1}^n w_t v_t \phi_n(v_t). \end{aligned} \tag{25}$$

Based on (23) and (25), we have

$$\left(1-K(0)\mathbb{E}[\lambda^{-1}]\frac{1}{na_n}\sum_{t=1}^n w_t v_t\phi_n(v_t)\right)\cdot(\mathbb{E}_\star[\lambda|X_n,Y_n]-\mathbb{E}[\lambda]) = \frac{d_1}{na_n}\sum_{t=1}^n w_t v_t\phi_n(v_t). \tag{26}$$

Denote $Z_n = \frac{1}{\sqrt{n}}\sum_{t=1}^n w_t(v_t\phi_n(v_t) - \mathbb{E}[v_1\phi_n(v_1)])$. Since w_t and v_t are independent, we have $\mathbb{E}[Z_n]=0$ and

$$\begin{aligned}\mathbb{E}[Z_n^2] &= \frac{1}{n}\sum_{t=1}^n \mathbb{E}[w_t^2]\mathbb{E}\left[\left(v_t\phi_n(v_t)-\mathbb{E}[v_1\phi_n(v_1)]\right)^2\right] \\ &= \mathbb{E}[w_1^2]\mathbb{E}\left[\left(v_t\phi_n(v_t)-\mathbb{E}[v_1\phi_n(v_1)]\right)^2\right] \equiv \delta_n^2. \end{aligned}$$

By a standard central limit theorem for the i.i.d. random variable case, we have

$$\delta_n^{-1} Z_n \to_D \mathcal{N}(0,1). \tag{27}$$

We then have

$$\begin{aligned}
&\left(1 - K(0)\mathbb{E}[\lambda^{-1}]\frac{1}{na_n}\sum_{t=1}^{n} w_t v_t \phi_n(v_t)\right)(\mathbb{E}_\star[\lambda|X_n,Y_n] - \mathbb{E}[\lambda]) \\
&= \frac{d_1}{na_n}\sum_{t=1}^{n} w_t v_t \phi_n(v_t) = \frac{d_1}{\sqrt{n}a_n}\left(Z_n + \sqrt{n}\mathbb{E}[w_1]\mathbb{E}[v_1\phi_n(v_1)]\right).
\end{aligned}$$

Using $\frac{1}{n}\sum_{t=1}^{n}\left(w_t v_t \phi_n(v_t) - \mathbb{E}[w_1]\mathbb{E}[v_1\phi_n(v_1)]\right) \to 0$ as $n \to \infty$, and recalling the definitions of $b_{1n} = a_n^{-1}\mathbb{E}[w_1]\mathbb{E}[v_1\phi_n(v_1)]$ and $b_{0n} = 1 - K(0)\mathbb{E}[\lambda^{-1}]b_{1n}$ listed in Theorem 1, we obtain as $n \to \infty$

$$\sqrt{n}a_n\delta_n^{-1}\left(b_{0n}\ (\mathbb{E}_\star[\lambda|X_n,Y_n] - \mathbb{E}[\lambda]) - d_1 b_{1n}\right) = \frac{d_1}{\delta_n} Z_n \to_D \mathcal{N}(0, d_1^2) \tag{28}$$

by Assumption 4 and equation (27).

Therefore, under Assumptions 1–4, when $m \to \infty$ and $n \to \infty$, we have

$$\begin{aligned}
&\sqrt{n}\delta_n^{-1} a_n\left(b_{0n}\left(\widehat{\lambda}_{mn} - \mathbb{E}[\lambda]\right) - d_1 b_{1n}\right) \\
&= \sqrt{n}\delta_n^{-1} a_n\left(\frac{b_{0n}}{m}\sum_{j=1}^{m}\left(\lambda_{jn} - \mathbb{E}_\star[\lambda|X_n,Y_n]\right) + b_{0n}\left(\mathbb{E}_\star[\lambda|X_n,Y_n] - \mathbb{E}[\lambda]\right) - d_1 b_{1n}\right) \\
&= \frac{\sqrt{n}\delta_n^{-1} a_n b_{0n}}{\sqrt{m}}\frac{1}{\sqrt{m}}\sum_{j=1}^{m}\left(\lambda_{jn} - \mathbb{E}_\star[\lambda|X_n,Y_n]\right) + \sqrt{n}\delta_n^{-1} a_n\left(b_{0n}\left(\mathbb{E}_\star[\lambda|X_n,Y_n] - \mathbb{E}[\lambda]\right) - d_1 b_{1n}\right) \\
&= o_P(1) + \sqrt{n}\delta_n^{-1} a_n\left(b_{0n}\left(\mathbb{E}_\star[\lambda|X_n,Y_n] - \mathbb{E}[\lambda]\right) - d_1 b_{1n}\right) \longrightarrow_D \mathcal{N}(0, d_1^2).
\end{aligned} \tag{29}$$

Therefore, we have proved Theorem 1. In order to prove Theorem 2, we introduce a useful lemma in Lemma C below.

Lemma C. Under Assumption 5 listed in Appendix A1, we have as $n \to \infty$,

$$\Delta_n^{1/2}(\widehat{\theta}_n)\left(\mathbb{E}[\theta|X_n,Y_n] - \mathbb{E}[\theta]\right) \to_D \mathcal{N}(0, \Sigma_0),$$

where Σ_0 is same as that in Assumption 5 (iv) and it can be estimated by a traditional method, such as the maximum likelihood estimation method.

Proof of Lemma C. By definition, we have that

$$\mathbb{E}[\theta|X_n,Y_n] = \int \theta f_n(\theta|X_n,Y_n)d\theta = \frac{\int \theta e^{g_n(\theta)}\pi_\theta(\theta)d\theta}{\int e^{g_n(\theta)}\pi_\theta(\theta)d\theta} = \frac{q_n(X_n,Y_n)}{p_n(X_n,Y_n)},$$

where $q_n(X_n,Y_n) = \int \theta e^{g_n(\theta)}\pi_\theta(\theta)d\theta$ and $p_n(X_n,Y_n) = \int e^{g_n(\theta)}\pi_\theta(\theta)d\theta$.

Taking the first order Taylor expansion of $g_n(\theta)$ in the neighborhood of $\widehat{\theta}_n$, we have that

$$g_n(\theta) = g_n(\widehat{\theta}_n) + g_n^{(1)}(\widehat{\theta}_n)(\theta - \widehat{\theta}_n) + \frac{1}{2}g_n^{(2)}(\widehat{\theta}_n)(\theta - \widehat{\theta}_n)^2 = g_n(\widehat{\theta}_n) - \frac{1}{2}\Delta_n(\widehat{\theta}_n)(\theta - \widehat{\theta}_n)^2,$$

where $g_n^{(1)}(\widehat{\theta}_n)$ and $g_n^{(2)}(\widehat{\theta}_n)$ are the first order and second order derivatives of $g_n(\theta)$ evaluated at the point $\widehat{\theta}_n$, respectively.

So $q_n(X_n, Y_n)$ can be written as

$$q_n(X_n, Y_n) = \int_{C_n} \theta e^{g_n(\widehat{\theta}_n)} e^{-\frac{1}{2}\Delta_n(\widehat{\theta}_n)(\theta-\widehat{\theta}_n)^2} \pi_\theta(\theta) d\theta + \int_{D_n} \theta e^{g_n(\theta)} \pi_\theta(\theta) d\theta = q_{1n}(X_n, Y_n) + q_{2n}(X_n, Y_n),$$

where

$$q_{1n}(X_n, Y_n) = \int_{C_n} \theta e^{g_n(\widehat{\theta}_n)} e^{-\frac{1}{2}\Delta_n(\widehat{\theta}_n)(\theta-\widehat{\theta}_n)^2} \pi_\theta(\theta) d\theta \text{ and } q_{2n}(X_n, Y_n) = \int_{D_n} \theta e^{g_n(\theta)} \pi_\theta(\theta) d\theta.$$

Similarly, $p_n(X_n, Y_n)$ can be written as

$$p_n(X_n, Y_n) = \int_{C_n} e^{g_n(\widehat{\theta}_n)} e^{-\frac{1}{2}\Delta_n(\widehat{\theta}_n)(\theta-\widehat{\theta}_n)^2} \pi_\theta(\theta) d\theta + \int_{D_n} e^{g_n(\theta)} \pi_\theta(\theta) d\theta = p_{1n}(X_n, Y_n) + p_{2n}(X_n, Y_n),$$

where

$$p_{1n}(X_n, Y_n) = \int_{C_n} e^{g_n(\widehat{\theta}_n)} e^{-\frac{1}{2}\Delta_n(\widehat{\theta}_n)(\theta-\widehat{\theta}_n)^2} \pi_\theta(\theta) d\theta \text{ and } p_{2n}(X_n, Y_n) = \int_{D_n} e^{g_n(\theta)} \pi_\theta(\theta) d\theta.$$

We first express $q_{1n}(X_n, Y_n)$ as follows:

$$\begin{aligned}
&q_{1n}(X_n, Y_n) = \int_{C_n} \theta e^{g_n(\widehat{\theta}_n)} e^{-\frac{1}{2}\Delta_n(\widehat{\theta}_n)(\theta-\widehat{\theta}_n)^2} \pi_\theta(\theta) d\theta \\
&= \int_{|y|\le c} \left(\widehat{\theta}_n + \Delta_n^{-1/2}(\widehat{\theta}_n) y\right) \Delta_n^{-1/2}(\widehat{\theta}_n) e^{g_n(\widehat{\theta}_n)} \pi(\widehat{\theta}_n + \Delta_n^{-1/2}(\widehat{\theta}_n) y) e^{-\frac{1}{2}y^2} dy \\
&= \widehat{\theta}_n \Delta_n^{-1/2}(\widehat{\theta}_n) e^{g_n(\widehat{\theta}_n)} \int_{|y|\le c} \pi(\widehat{\theta}_n + \Delta_n^{-1/2}(\widehat{\theta}_n) y) e^{-\frac{1}{2}y^2} dy \\
&+ \Delta_n^{-1/2}(\widehat{\theta}_n) \Delta_n^{-1/2}(\widehat{\theta}_n) e^{g_n(\widehat{\theta}_n)} \int_{|y|\le c} \pi(\widehat{\theta}_n + \Delta_n^{-1/2}(\widehat{\theta}_n) y) y e^{-\frac{1}{2}y^2} dy \\
&= \widehat{\theta}_n \Delta_n^{-1/2}(\widehat{\theta}_n) e^{g_n(\widehat{\theta}_n)} \pi(\widehat{\theta}_n) \int_{|y|\le c} e^{-\frac{1}{2}y^2} dy + \widehat{\theta}_n \Delta_n^{-1/2}(\widehat{\theta}_n) e^{g_n(\widehat{\theta}_n)} \Delta_n^{-1/2}(\widehat{\theta}_n) \pi^{(1)}(\widehat{\theta}_n) \int_{|y|\le c} y e^{-\frac{1}{2}y^2} dy \\
&+ \Delta_n^{-1/2}(\widehat{\theta}_n) \Delta_n^{-1/2}(\widehat{\theta}_n) e^{g_n(\widehat{\theta}_n)} \pi(\widehat{\theta}_n) \int_{|y|\le c} y e^{-\frac{1}{2}y^2} dy + o_P(\Delta_n^{-1}(\widehat{\theta}_n)) \\
&= \widehat{\theta}_n \Delta_n^{-1/2}(\widehat{\theta}_n) e^{g_n(\widehat{\theta}_n)} \pi(\widehat{\theta}_n) \int_{|y|\le c} e^{-\frac{1}{2}y^2} dy + o_P(\Delta_n^{-1}(\widehat{\theta}_n)).
\end{aligned} \tag{30}$$

The last equality is true, because

$$\begin{aligned}
&\int_{|y|\le c} y e^{-\frac{1}{2}y^2} dy = \int_0^c y e^{-\frac{1}{2}y^2} dy + \int_{-c}^0 y e^{-\frac{1}{2}y^2} dy \\
&= \int_0^c y e^{-\frac{1}{2}y^2} dy + \int_c^0 (-y) e^{-\frac{1}{2}y^2} d(-y) = \int_0^c y e^{-\frac{1}{2}y^2} dy - \int_0^c y e^{-\frac{1}{2}y^2} dy = 0.
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
&p_{1n}(X_n, Y_n) = \int_{C_n} e^{g_n(\widehat{\theta}_n)} e^{-\frac{1}{2}\Delta_n(\widehat{\theta}_n)(\theta-\widehat{\theta}_n)^2} \pi_\theta(\theta) d\theta \\
&= \Delta_n^{-1/2}(\widehat{\theta}_n) e^{g_n(\widehat{\theta}_n)} \int_{|y|\le c} \pi(\widehat{\theta}_n + \Delta_n^{-1/2}(\widehat{\theta}_n) y) e^{-\frac{1}{2}y^2} dy \\
&= \Delta_n^{-1/2}(\widehat{\theta}_n) e^{g_n(\widehat{\theta}_n)} \pi(\widehat{\theta}_n) \int_{|y|\le c} e^{-\frac{1}{2}y^2} dy + \Delta_n^{-1/2}(\widehat{\theta}_n) e^{g_n(\widehat{\theta}_n)} \Delta_n^{-1/2}(\widehat{\theta}_n) \pi^{(1)}(\widehat{\theta}_n) \int_{|y|\le c} y e^{-\frac{1}{2}y^2} dy \\
&+ o_P(\Delta_n^{-1}(\widehat{\theta}_n)) = \Delta_n^{-1/2}(\widehat{\theta}_n) e^{g_n(\widehat{\theta}_n)} \pi(\widehat{\theta}_n) \int_{|y|\le c} e^{-\frac{1}{2}y^2} dy + o_P(\Delta_n^{-1}(\widehat{\theta}_n)).
\end{aligned} \tag{31}$$

Then, based on (30) and (31), we obtain

$$\frac{q_{1n}(X_n, Y_n)}{p_{1n}(X_n, Y_n)} = \widehat{\theta}_n + \frac{o_P(\Delta_n^{-1}(\widehat{\theta}_n))}{\Delta_n^{-1/2}(\widehat{\theta}_n)e^{g_n(\widehat{\theta}_n)}\pi(\widehat{\theta}_n)\int_{|y|\le c} e^{-\frac{1}{2}y^2}\,dy + o_P(\Delta_n^{-1}(\widehat{\theta}_n))} = \widehat{\theta}_n + o_P(\Delta_n^{-1/2}(\widehat{\theta}_n)). \tag{32}$$

Note that under Assumption 5 (iii), we have

$$\frac{q_{2n}(X_n, Y_n)}{p_n(X_n, Y_n)} = o_P(\Delta_n^{-1/2}(\widehat{\theta}_n)) \text{ and } \frac{p_{2n}(X_n, Y_n)}{p_n(X_n, Y_n)} = o_P(\Delta_n^{-1/2}(\widehat{\theta}_n)). \tag{33}$$

Therefore, based on (32) and (33), with Assumption 5 (iii), it is shown that

$$\begin{aligned}
|\mathbb{E}[\theta|X_n, Y_n] - \widehat{\theta}_n| &= \left|\frac{q_n(X_n, Y_n)}{p_n(X_n, Y_n)} - \widehat{\theta}_n\right| = \left|\frac{q_{1n}(X_n, Y_n)}{p_n(X_n, Y_n)} + \frac{q_{2n}(X_n, Y_n)}{p_n(X_n, Y_n)} - \widehat{\theta}_n\right| \\
&\le \left|\frac{q_{1n}(X_n, Y_n)}{p_n(X_n, Y_n)} - \widehat{\theta}_n\right| + \left|\frac{q_{2n}(X_n, Y_n)}{p_n(X_n, Y_n)}\right| \\
&= \left|\frac{q_{1n}(X_n, Y_n)}{p_n(X_n, Y_n)} - \frac{q_{1n}(X_n, Y_n)}{p_{1n}(X_n, Y_n)} + \frac{q_{1n}(X_n, Y_n)}{p_{1n}(X_n, Y_n)} - \widehat{\theta}_n\right| + \left|\frac{q_{2n}(X_n, Y_n)}{p_n(X_n, Y_n)}\right| \\
&\le \left|\frac{q_{1n}(X_n, Y_n)}{p_n(X_n, Y_n)} - \frac{q_{1n}(X_n, Y_n)}{p_{1n}(X_n, Y_n)}\right| + \left|\frac{q_{1n}(X_n, Y_n)}{p_{1n}(X_n, Y_n)} - \widehat{\theta}_n\right| + \left|\frac{q_{2n}(X_n, Y_n)}{p_n(X_n, Y_n)}\right| \\
&= \left|q_{1n}(X_n, Y_n)\right| \left|\frac{1}{p_n(X_n, Y_n)} - \frac{1}{p_{1n}(X_n, Y_n)}\right| + \left|\frac{q_{1n}(X_n, Y_n)}{p_{1n}(X_n, Y_n)} - \widehat{\theta}_n\right| + \left|\frac{q_{2n}(X_n, Y_n)}{p_n(X_n, Y_n)}\right| \\
&= \left|q_{1n}(X_n, Y_n)\right| \left|\frac{p_{2n}(X_n, Y_n)}{p_n(X_n, Y_n)p_{1n}(X_n, Y_n)}\right| + o_P(\Delta_n^{-1/2}(\widehat{\theta}_n)) \\
&= \left|\frac{q_{1n}(X_n, Y_n)}{p_{1n}(X_n, Y_n)}\right| \left|\frac{p_{2n}(X_n, Y_n)}{p_n(X_n, Y_n)}\right| + o_P(\Delta_n^{-1/2}(\widehat{\theta}_n)) = o_P(\Delta_n^{-1/2}(\widehat{\theta}_n)).
\end{aligned}$$

Thus, $\Delta_n^{1/2}(\widehat{\theta}_n)\left(\mathbb{E}[\theta|X_n, Y_n] - \widehat{\theta}_n\right) \to_P 0$.

Therefore, with Assumption 5(iv), we obtain

$$\Delta_n^{1/2}(\widehat{\theta}_n)\left(\mathbb{E}[\theta|X_n, Y_n] - \mathbb{E}[\theta]\right) = \Delta_n^{1/2}(\widehat{\theta}_n)\left(\mathbb{E}[\theta|X_n, Y_n] - \widehat{\theta}_n\right) + \Delta_n^{1/2}(\widehat{\theta}_n)\left(\widehat{\theta}_n - \mathbb{E}[\theta]\right) \to_D \mathcal{N}(0, \Sigma_0), \tag{34}$$

which completes the proof of Lemma C.

We next give the proof of Theorem 2 by making use of Lemma C.

Proof of Theorem 2(i).

From equation (12) in the proof of Theorem 1, we have obtained that

$$l_n(\lambda, \theta) = \log L_n(\lambda, \theta) = \sum_{t=1}^n \log f(\widehat{v}_t; \theta) = \sum_{t=1}^n \log f(v_t; \theta) + \sum_{t=1}^n \gamma(v_t; \theta)\Gamma(\tau_t; \alpha, h) + o_P(1),$$

where $\gamma(v_t;\theta) = \frac{f^{(1)}(v_t;\theta)}{f(v_t;\theta)}$. Therefore, we have

$$e^{l_n(\lambda,\theta)} = e^{\sum_{t=1}^n \log f(v_t;\theta)} e^{\sum_{t=1}^n \gamma(v_t;\theta)\Gamma(\tau_t;\alpha,h)} e^{o_P(1)} = e^{\sum_{t=1}^n \log f(v_t;\theta)} + R_n(\lambda, \theta),$$

where $R_n(\lambda,\theta) = \left(e^{\sum_{t=1}^n \gamma(v_t;\theta)\Gamma(\tau_t;\alpha,h) + o_P(1)} - 1\right) e^{\sum_{t=1}^n \log f(v_t;\theta)} = \left(\sum_{t=1}^n \gamma(v_t;\theta)\Gamma(\tau_t;\alpha,h)\right) G_n(\theta)$.

Since $\exp(o_P(1)) \to_P 1$, the replacement of $\exp(o_P(1))$ by 1 does not change anything in the derivation of the limit distribution below. By ignoring the high order $o_P(1)$ in the following derivations, we then have the following two equations:

$$\begin{aligned}\iint e^{l_n(\lambda,\theta)}\pi_\lambda(\lambda)\pi_\theta(\theta)d\lambda d\theta &= \iint e^{\sum_{t=1}^n \log f(v_t;\theta)}\pi_\lambda(\lambda)\pi_\theta(\theta)d\lambda d\theta + \iint R_n(\lambda,\theta)\pi_\lambda(\lambda)\pi_\theta(\theta)d\lambda d\theta\\ &= \int G_n(\theta)\pi_\theta(\theta)d\theta + \iint R_n(\lambda,\theta)\pi_\lambda(\lambda)\pi_\theta(\theta)d\lambda d\theta, \end{aligned}\tag{35}$$

$$\begin{aligned}\iint \theta e^{l_n(\lambda,\theta)}\pi_\lambda(\lambda)\pi_\theta(\theta)d\lambda d\theta &= \iint \theta e^{\sum_{t=1}^n \log f(v_t;\theta)}\pi_\lambda(\lambda)\pi_\theta(\theta)d\lambda d\theta + \iint \theta R_n(\lambda,\theta)\pi_\lambda(\lambda)\pi_\theta(\theta)d\lambda d\theta\\ &= \int \theta G_n(\theta)\pi_\theta(\theta)d\theta + \iint \theta R_n(\lambda,\theta)\pi_\lambda(\lambda)\pi_\theta(\theta)d\lambda d\theta. \end{aligned}\tag{36}$$

It follows from (35) and (36) that

$$\begin{aligned}\mathbb{E}_\star[\theta|X_n,Y_n] &= \frac{\iint \theta e^{l_n(\lambda,\theta)}\pi_\lambda(\lambda)\pi_\theta(\theta)d\lambda d\theta}{\iint e^{l_n(\lambda,\theta)}\pi_\lambda(\lambda)\pi_\theta(\theta)d\lambda d\theta}\\ &= \frac{\int \theta G_n(\theta)\pi_\theta(\theta)d\theta + \iint \theta R_n(\lambda,\theta)\pi_\lambda(\lambda)\pi_\theta(\theta)d\lambda d\theta}{\int G_n(\theta)\pi_\theta(\theta)d\theta + \iint R_n(\lambda,\theta)\pi_\lambda(\lambda)\pi_\theta(\theta)d\lambda d\theta}.\end{aligned}$$

As $\mathbb{E}[\theta|X_n,Y_n] = \frac{\int \theta G_n(\theta)\pi_\theta(\theta)d\theta}{\int G_n(\theta)\pi_\theta(\theta)d\theta}$, we have

$$\begin{aligned}&\mathbb{E}_\star[\theta|X_n,Y_n] - \mathbb{E}[\theta|X_n,Y_n]\\ &= \frac{\int \theta G_n(\theta)\pi_\theta(\theta)d\theta + \iint \theta R_n(\lambda,\theta)\pi_\lambda(\lambda)\pi_\theta(\theta)d\lambda d\theta}{\int G_n(\theta)\pi_\theta(\theta)d\theta + \iint R_n(\lambda,\theta)\pi_\lambda(\lambda)\pi_\theta(\theta)d\lambda d\theta} - \frac{\int \theta G_n(\theta)\pi_\theta(\theta)d\theta}{\int G_n(\theta)\pi_\theta(\theta)d\theta}\\ &= \frac{\iint \theta R_n(\lambda,\theta)\pi_\lambda(\lambda)\pi_\theta(\theta)d\lambda d\theta}{\int G_n(\theta)\pi_\theta(\theta)d\theta} - \mathbb{E}[\theta|X_n,Y_n]\frac{\iint R_n(\lambda,\theta)\pi_\lambda(\lambda)\pi_\theta(\theta)d\lambda d\theta}{\int G_n(\theta)\pi_\theta(\theta)d\theta}.\end{aligned}\tag{37}$$

From the proof of Theorem 1, we have obtained that

$$\sum_{t=1}^n \gamma(v_t;\theta)\Gamma(\tau_t;\alpha,h) = -\frac{K(0)}{nh}\sum_{t=1}^n v_t\gamma(v_t;\theta)w_t\,(1+o_P(1)).$$

Recall that $R_n(\lambda,\theta) = \left(\sum_{t=1}^n \gamma(v_t;\theta)\Gamma(\tau_t;\alpha,h)\right)G_n(\theta)$. Therefore, we obtain

$$\begin{aligned}&\frac{\iint R_n(\lambda,\theta)\pi_\lambda(\lambda)\pi_\theta(\theta)d\lambda d\theta}{\int G_n(\theta)\pi_\theta(\theta)d\theta}\\ &= -\frac{\iint \left(\frac{K(0)}{nh}\sum_{t=1}^n \gamma(v_t;\theta)w_t v_t\right)G_n(\theta)\pi_\lambda(\lambda)\pi_\theta(\theta)d\lambda d\theta}{\int G_n(\theta)\pi_\theta(\theta)d\theta}\\ &= -K(0)\mathbb{E}[\lambda^{-1}]\frac{\int \left(\frac{1}{na_n}\sum_{t=1}^n \gamma(v_t;\theta)w_t v_t\right)G_n(\theta)\pi_\theta(\theta)d\theta}{\int G_n(\theta)\pi_\theta(\theta)d\theta}\\ &= -K(0)\mathbb{E}[\lambda^{-1}]\frac{1}{na_n}\sum_{t=1}^n w_t v_t\frac{\int \gamma(v_t;\theta)G_n(\theta)\pi_\theta(\theta)d\theta}{\int G_n(\theta)\pi_\theta(\theta)d\theta}.\end{aligned}$$

Similarly, we can have

$$
\begin{aligned}
&\frac{\iint \theta R_n(\lambda,\theta)\pi_\lambda(\lambda)\pi_\theta(\theta)d\lambda d\theta}{\int G_n(\theta)\pi_\theta(\theta)d\theta} \\
&= -\frac{\iint \theta\left(\frac{K(0)}{nh}\sum_{t=1}^n \gamma(v_t;\theta)w_t v_t\right)G_n(\theta)\pi_\lambda(\lambda)\pi_\theta(\theta)d\lambda d\theta}{\int G_n(\theta)\pi_\theta(\theta)d\theta} \\
&= -K(0)\mathbb{E}[\lambda^{-1}]\frac{\int\left(\frac{1}{na_n}\sum_{t=1}^n \gamma(v_t;\theta)w_t v_t\right)\theta G_n(\theta)\pi_\theta(\theta)d\theta}{\int G_n(\theta)\pi_\theta(\theta)d\theta} \\
&= -K(0)\mathbb{E}[\lambda^{-1}]\frac{1}{na_n}\sum_{t=1}^n w_t v_t \frac{\int \theta\gamma(v_t;\theta)G_n(\theta)\pi_\theta(\theta)d\theta}{\int G_n(\theta)\pi_\theta(\theta)d\theta}.
\end{aligned}
$$

Recall that

$$
G_n(\theta) = \prod_{t=1}^n f(v_t;\theta) = e^{\sum_{t=1}^n \log f(v_t;\theta)} = e^{n\mathbb{E}_\theta[\log f(v_1;\theta)]}(1+o_P(1)), \tag{38}
$$

where $\mathbb{E}_\theta[\log f(v_1;\theta)] = \int \log f(v_1;\theta)f(v;\theta)dv$. Denote $A_n(\theta) = \frac{e^{n\mathbb{E}_\theta[\log f(v_1;\theta)]}}{\int e^{n\mathbb{E}_\theta[\log f(v_1;\theta)]}\pi(\theta)d\theta}$.

Simple decompositions give

$$
\begin{aligned}
&\mathbb{E}_\star[\theta|X_n,Y_n] - \mathbb{E}[\theta|X_n,Y_n] \\
&= \frac{\iint \theta R_n(\lambda,\theta)\pi_\lambda(\lambda)\pi_\theta(\theta)d\lambda d\theta}{\int G_n(\theta)\pi_\theta(\theta)d\theta} - \mathbb{E}[\theta|X_n,Y_n]\frac{\iint R_n(\lambda,\theta)\pi_\lambda(\lambda)\pi_\theta(\theta)d\lambda d\theta}{\int G_n(\theta)\pi_\theta(\theta)d\theta} \\
&= -K(0)\mathbb{E}[\lambda^{-1}]\frac{1}{a_n n}\sum_{t=1}^n w_t v_t \frac{\int \theta\gamma(v_t;\theta)G_n(\theta)\pi_\theta(\theta)d\theta}{\int G_n(\theta)\pi_\theta(\theta)d\theta} \\
&+ \mathbb{E}[\theta|X_n,Y_n]K(0)\mathbb{E}[\lambda^{-1}]\frac{1}{a_n n}\sum_{t=1}^n w_t v_t \frac{\int \gamma(v_t;\theta)G_n(\theta)\pi_\theta(\theta)d\theta}{\int G_n(\theta)\pi_\theta(\theta)d\theta} \\
&= -K(0)\mathbb{E}[\lambda^{-1}]\frac{1}{a_n n}\sum_{t=1}^n w_t v_t \left(\frac{\int \theta\gamma(v_t;\theta)G_n(\theta)\pi_\theta(\theta)d\theta}{\int G_n(\theta)\pi_\theta(\theta)d\theta} - \mathbb{E}[\theta|X_n,Y_n]\frac{\int \gamma(v_t;\theta)G_n(\theta)\pi_\theta(\theta)d\theta}{\int G_n(\theta)\pi_\theta(\theta)d\theta}\right) \\
&= -K(0)\mathbb{E}[\lambda^{-1}]\frac{1}{a_n n}\sum_{t=1}^n w_t v_t \frac{\int (\theta - \mathbb{E}[\theta|X_n,Y_n])\gamma(v_t;\theta)G_n(\theta)\pi(\theta)d\theta}{\int G_n(\theta)\pi(\theta)d\theta} \\
&= K(0)\mathbb{E}[\lambda^{-1}]\frac{1}{a_n n}\sum_{t=1}^n w_t v_t \left(\int (\mathbb{E}[\theta]-\theta)\gamma(v_t;\theta)A_n(\theta)\pi(\theta)d\theta\right) \\
&+ K(0)\mathbb{E}[\lambda^{-1}]\frac{1}{a_n n}\sum_{t=1}^n w_t v_t \left(\int \gamma(v_t;\theta)A_n(\theta)\pi(\theta)d\theta\right)\cdot(\mathbb{E}[\theta|X_n,Y_n]-\mathbb{E}[\theta]).
\end{aligned} \tag{39}
$$

By Lemma C, we have $\mathbb{E}[\theta|X_n,Y_n]-\mathbb{E}[\theta] = o_P(\Delta_n^{-1/2}(\widehat{\theta}_n))$. Therefore, it follows that

$$
\begin{aligned}
\mathbb{E}_\star[\theta|X_n,Y_n] - \mathbb{E}[\theta|X_n,Y_n] &= K(0)\mathbb{E}[\lambda^{-1}]\,(1+o_P(1)) \\
&\times \frac{1}{na_n}\sum_{t=1}^n w_t v_t \left(\int (\mathbb{E}[\theta]-\theta)\gamma(v_t;\theta)A_n(\theta)\pi(\theta)d\theta\right).
\end{aligned}
$$

Let $\gamma_n(v_t) = \int(\mathbb{E}[\theta]-\theta)\gamma(v_t;\theta)A_n(\theta)\pi(\theta)d\theta$. Denote

$$
S_n = \frac{1}{\sqrt{n}}\sum_{t=1}^n w_t(v_t\gamma_n(v_t) - \mathbb{E}[v_1\gamma_n(v_1)]).
$$

Since w_t and v_t are independent, we have $\mathbb{E}[S_n] = 0$ and

$$\begin{aligned}\mathbb{E}[S_n^2] &= \frac{1}{n}\sum_{t=1}^{n}\mathbb{E}[w_t^2]\mathbb{E}\left[\left(v_t\gamma_n(v_t) - \mathbb{E}[v_1\gamma_n(v_1)]\right)^2\right] \\ &= \mathbb{E}[w_1^2]\mathbb{E}\left[\left(v_t\gamma_n(v_t) - \mathbb{E}[v_1\gamma_n(v_1)]\right)^2\right] \equiv \sigma_n^2.\end{aligned}$$

A standard central limit theorem implies that as $n \to \infty$

$$\sigma_n^{-1} S_n \to_D \mathcal{N}(0, I_p), \tag{40}$$

where I_p denotes the $p \times p$ identity matrix and p denotes the number of parameters in the vector θ.

Let $T_n = \mathbb{E}_\star[\theta|X_n, Y_n] - \mathbb{E}[\theta|X_n, Y_n] = K(0)\mathbb{E}[\lambda^{-1}]\frac{1}{\sqrt{n}a_n}\left(S_n + \sqrt{n}\mathbb{E}[w_1]\mathbb{E}[v_1\gamma_n(v_1)]\right)$. Based on (40), under Assumption 3 (iv), we can get that

$$\sqrt{n}\sigma_n^{-1}a_n\left(\frac{1}{K(0)\mathbb{E}[\lambda^{-1}]}T_n - a_n^{-1}\mathbb{E}[w_1]\mathbb{E}[v_1\gamma_n(v_1)]\right) = \sigma_n^{-1}S_n \to_D \mathcal{N}(0, I_p). \tag{41}$$

Equivalently, we have

$$\begin{aligned}&\sqrt{n}\sigma_n^{-1}a_n\left(\mathbb{E}_\star[\theta|X_n, Y_n] - \mathbb{E}[\theta|X_n, Y_n] - a_n^{-1}K(0)\mathbb{E}[\lambda^{-1}]\mathbb{E}[w_1]\mathbb{E}[v_1\gamma_n(v_1)]\right) \\ &\to_D \mathcal{N}(0, K^2(0)\mathbb{E}^2[\lambda^{-1}]I_p).\end{aligned} \tag{42}$$

By a conditional central limit theorem for the i.i.d. case, we have as $m \to \infty$

$$\frac{1}{\sqrt{m}}\sum_{j=1}^{m}\left(\theta_{jn} - \mathbb{E}_\star[\theta|X_n, Y_n]\right) = O_P(1).$$

Let $b_{2n} = a_n^{-1}K(0)\mathbb{E}[\lambda^{-1}]\mathbb{E}[w_1]\mathbb{E}[v_1\gamma_n(v_1)]$.

Then, under Assumptions 1–3 and 5, we obtain as $m \to \infty$ and $n \to \infty$

$$\begin{aligned}&\sqrt{n}\sigma_n^{-1}a_n\left(\widehat{\theta}_{mn} - \mathbb{E}[\theta] - b_{2n}\right) \\ &= \sqrt{n}\sigma_n^{-1}a_n\left(\frac{1}{m}\sum_{j=1}^{m}\theta_{jn} - \mathbb{E}_\star[\theta|X_n, Y_n] + \mathbb{E}_\star[\theta|X_n, Y_n] - \mathbb{E}[\theta|X_n, Y_n] - b_{2n} + \mathbb{E}[\theta|X_n, Y_n] - \mathbb{E}[\theta]\right) \\ &= \sqrt{n}\sigma_n^{-1}a_n\left(\frac{1}{m}\sum_{j=1}^{m}\theta_{jn} - \mathbb{E}_\star[\theta|X_n, Y_n]\right) + \sqrt{n}\sigma_n^{-1}a_n\left(\mathbb{E}_\star[\theta|X_n, Y_n] - \mathbb{E}[\theta|X_n, Y_n] - b_{2n}\right) \\ &+ \frac{\sqrt{n}\sigma_n^{-1}a_n}{\Delta_n^{1/2}(\widehat{\theta}_n)}\Delta_n^{1/2}(\widehat{\theta}_n)\left(\mathbb{E}[\theta|X_n, Y_n] - \mathbb{E}[\theta]\right) \\ &= \frac{\sqrt{n}\sigma_n^{-1}a_n}{\sqrt{m}}\frac{1}{\sqrt{m}}\sum_{j=1}^{m}\left(\theta_{jn} - \mathbb{E}_\star[\theta|X_n, Y_n]\right) + \sqrt{n}\sigma_n^{-1}a_n\left(\mathbb{E}_\star[\theta|X_n, Y_n] - \mathbb{E}[\theta|X_n, Y_n] - b_{2n}\right) + o_P(1) \\ &\to_D \mathcal{N}(0, K^2(0)\mathbb{E}^2[\lambda^{-1}]I_p).\end{aligned}$$

To sum up, under Assumptions 1–3 and 5, as $m \to \infty$ and $n \to \infty$, we have

$$\sqrt{n}\sigma_n^{-1}a_n\left(\widehat{\theta}_{mn} - \mathbb{E}[\theta] - b_{2n}\right) \to_D \mathcal{N}(0, K^2(0)\mathbb{E}^2[\lambda^{-1}]I_p),$$

which completes the proof of Theorem 2(i).

Proof of Theorem 2(ii).

Equation (39) implies

$$\mathbb{E}_\star[\theta|X_n,Y_n]-\mathbb{E}[\theta|X_n,Y_n] = -K(0)\mathbb{E}[\lambda^{-1}]\frac{1}{a_n n}\sum_{t=1}^n w_t v_t \frac{\int(\theta-\mathbb{E}[\theta|X_n,Y_n])\,\gamma(v_t;\theta)G_n(\theta)\pi(\theta)d\theta}{\int G_n(\theta)\pi(\theta)d\theta}$$
$$= -\frac{K(0)\mathbb{E}[\lambda^{-1}]}{a_n n}\sum_{t=1}^n v_t\gamma(v_t)w_t\left(\frac{\int(\theta-\mathbb{E}[\theta|X_n,Y_n])\,G_n(\theta)\pi(\theta)d\theta}{\int G_n(\theta)\pi(\theta)d\theta}\right)$$
$$= -\frac{K(0)\mathbb{E}[\lambda^{-1}]}{a_n n}\sum_{t=1}^n v_t\gamma(v_t)w_t\cdot(\mathbb{E}[\theta|X_n,Y_n]-\mathbb{E}[\theta|X_n,Y_n]) = 0$$

when $\gamma(v;\theta)\equiv\gamma(v)$.

Therefore, we have as $m\to\infty$ and $n\to\infty$

$$\Delta_n^{1/2}(\widehat{\theta}_n)\left(\widehat{\theta}_{mn}-\mathbb{E}[\theta]\right)$$
$$= \Delta_n^{1/2}(\widehat{\theta}_n)\left(\frac{1}{m}\sum_{j=1}^m\theta_{jn}-\mathbb{E}_\star[\theta|X_n,Y_n]+\mathbb{E}_\star[\theta|X_n,Y_n]-\mathbb{E}[\theta|X_n,Y_n]+\mathbb{E}[\theta|X_n,Y_n]-\mathbb{E}[\theta]\right)$$
$$= \Delta_n^{1/2}(\widehat{\theta}_n)\left(\frac{1}{m}\sum_{j=1}^m\theta_{jn}-\mathbb{E}_\star[\theta|X_n,Y_n]\right)+\Sigma_n^{1/2}(\widehat{\theta}_n)\,(\mathbb{E}_\star[\theta|X_n,Y_n]-\mathbb{E}[\theta|X_n,Y_n])$$
$$+\Delta_n^{1/2}(\widehat{\theta}_n)\,(\mathbb{E}[\theta|X_n,Y_n]-\mathbb{E}[\theta])$$
$$= \frac{\Delta_n^{1/2}(\widehat{\theta}_n)}{\sqrt{m}}\frac{1}{\sqrt{m}}\sum_{j=1}^m\left(\theta_{jn}-\mathbb{E}_\star[\theta|X_n,Y_n]\right)+\Delta_n^{1/2}(\widehat{\theta}_n)\,(\mathbb{E}[\theta|X_n,Y_n]-\mathbb{E}[\theta])$$
$$= \Delta_n^{1/2}(\widehat{\theta}_n)\,(\mathbb{E}[\theta|X_n,Y_n]-\mathbb{E}[\theta])+o_P(1)\to_D\mathcal{N}(0,\Sigma_0),$$

which completes the proof of Theorem 2(ii).