Generalized Laplace Inference in Multiple Change-Points Models*

Alessandro Casini† Pierre Perron‡
University of Rome Tor Vergata Boston University

2nd March 2020

Abstract

Under the classical long-span asymptotic framework we develop a class of Generalized Laplace (GL) inference methods for the change-point dates in a linear time series regression model with multiple structural changes analyzed in, e.g., Bai and Perron (1998). The GL estimator is defined by an integration rather than optimization-based method and relies on the least-squares criterion function. It is interpreted as a classical (non-Bayesian) estimator and the inference methods proposed retain a frequentist interpretation. This approach provides a better approximation about the uncertainty in the data of the change-points relative to existing methods. On the theoretical side, depending on some input (smoothing) parameter, the class of GL estimators exhibits a dual limiting distribution; namely, the classical shrinkage asymptotic distribution, or a Bayes-type asymptotic distribution. We propose an inference method based on Highest Density Regions using the latter distribution. We show that it has attractive theoretical properties not shared by the other popular alternatives, i.e., it is bet-proof. Simulations confirm that these theoretical properties translate to better finite-sample performance.

JEL Classification: C12, C13, C22
Keywords: Asymptotic Distribution, Bet-Proof, Break Date, Change-point, Generalized Laplace Inference, Highest Density Region, Quasi-Bayes.

*We thank Zhongjun Qu for useful comments.

†Department of Economics and Finance, University of Rome Tor Vergata, Via Columbia 2, Rome, 00133, IT. Email: alessandro.casini@uniroma2.it.

‡Department of Economics, Boston University, 270 Bay State Road, Boston, MA 02215, US. Email: perron@bu.edu.

1 Introduction

In the context of the multiple change-points model analyzed in Bai and Perron (1998), we develop inference methods for the change-point dates for a class of Generalized Laplace (GL) estimators using a classical long-span asymptotic framework. They are defined by an integration rather than an optimization-based method, the latter typically characterizing classical extremum estimators. The idea traces back to Laplace (1774), who first suggested to interpret transformation of a least-squares criterion function as a statistical belief over a parameter of interest. Hence, a Laplace estimator is defined similarly to a Bayesian estimator although the former relies on a statistical criterion function rather than a parametric likelihood function. As a consequence, the GL estimator is interpreted as a classical (non-Bayesian) estimator and the inference methods proposed retain a frequentist interpretation such that the GL estimators are constructed as a function of integral transformations of the least-squares criterion. In a first step, we use the approach of Bai and Perron (1998) to evaluate the least-squares criterion function at all candidate break dates. We then apply a transformation to obtain a proper distribution over the parameters of interest, referred to as the Quasi-posterior. For a given choice of a loss function and (possibly) a prior density, the estimator is then defined either explicitly as, for example, the mean or median of the (weighted) Quasi-posterior or implicitly as the minimizer of a smooth convex optimization problem.

The underlying asymptotic framework considered is the long-span shrinkage asymptotics of Bai (1997), Bai and Perron (1998) and also Perron and Qu (2006) who considerably relaxed some conditions, where the magnitude of the parameter shift is sample-size dependent and approaches zero as the sample size increases. Early contributions to this approach are Hinkley (1971), Bhattacharya (1987), and Yao (1987) for estimating break points. For testing for structural breaks, see Hawkins (1977), Picard (1985), Kim and Siegmund (1989), Andrews (1993), Horváth (1993) and Andrews and Ploberger (1994). See also the reviews of Csörgő and Horváth (1997), Perron (2006), Casini and Perron (2019c) and references therein.

One of our goals is to develop GL estimates with better small-sample properties compared to least-squares estimates, namely lower Mean Absolute and Root-Mean Squared Errors, and confidence sets with accurate coverage probabilities and relatively short lengths for a wide range of break sizes, whether small or large; existing methods which work well for either small or large breaks, but not for both. A second goal is to establish theoretical results that support the reported finite-sample properties about inference.

The asymptotic distribution of the GL estimator is derived via a local parameter related to a normalized deviation from the true fractional break date. The normalization factor corresponds to the rate of convergence of the original (extremum) least-squares estimator as established by Bai and Perron (1998). The asymptotic distribution of the GL estimator then depends on a sample-size dependent smoothing parameter sequence applied to the least-squares criterion function. We derive two distinct limiting distributions corresponding to different smoothing sequences of the criterion function [cf. Jun, Pinkse, and Wan (2015) for a related application in the context of the cube-root asymptotics of Kim and Pollard (1990)]. In one case, the estimator displays the same limit law as the asymptotic distribution of the least-squares estimator derived in Bai and Perron (1998) [see also Hinkley (1971), Picard (1985) and Yao (1987)]. In a second case, the limiting distribution is characterized by a ratio of integrals over functions of Gaussian processes and resembles the limiting distribution of Bayesian change-point estimators. The latter is exploited for the purpose of constructing confidence sets for the break dates. We use the concept of highest density regions (HDR) introduced by Casini and Perron (2019a) for structural change problems, which best summarizes the properties of the probability distribution of interest. This procedure yields confidence sets for the break date which, in finite samples, better account for the uncertainty over the parameter space in finite-samples because it effectively incorporates a statistical measure of the uncertainty in the least-squares criterion function. As noted in the literature on likelihood-based inference in some classes of nongranular problems [see e.g., Chernozhukov and Hong (2004), Ghosal, Ghosh, and Samanta (1995), Hirano and Porter (2003) and Ibragimov and Has'minskiĭ (1981)], the Maximum Likelihood Estimator (MLE) is generally not an asymptotically sufficient statistic in these models and so the likelihood contains more information asymptotically than the MLE. Hence, likelihood-based procedures are generally not functions of the MLE even asymptotically. This incompleteness property motivated the study of the entire likelihood rather than just the MLE. Likewise, our method exploits the entire behavior of the objective function.

Laplace's seminal insight has been applied successfully in many disciplines. In econometrics, Chernozhukov and Hong (2003) introduced Laplace-type estimators as an alternative to classical (regular) extremum estimators in several problems such as censored median regression and nonlinear instrumental variable; see also Forneron and Ng (2018) for a review and comparisons. Their main motivation was to solve the curse of dimensionality inherent to the computation of such estimators. In contrast, the class of GL estimators in structural change models serves distinct multiple purposes. First, inference about the break dates presents several challenges, in particular to provide methods with a satisfactory performance uniformly over different data-generating mechanisms and break magnitudes. The GL inference proves to be reliable and accurate in finite-samples. Second, it leads to inference methods that have both frequentist and credibility properties which is not shared by the other popular methods.

Turning to the problem of constructing confidence sets for a single break date, the standard asymptotic method for the linear regression model was proposed in Bai (1997), while Elliott and Müller (2007) proposed to invert the locally best invariant test of Nyblom (1989), and Eo and Morley (2015) suggested to invert the likelihood-ratio statistic of Qu and Perron (2007). The latter were mainly motivated by finite-sample results indicating that the exact coverage rates of the confidence intervals obtained from Bai's (1997) method are often below the nominal level when the magnitude of the break is small. It has been shown that the method of Elliott and Müller (2007) delivers the most accurate coverage rates but the average length of the confidence sets is significantly larger than with other methods. The confidence sets for the break dates constructed from the GL inference that we develop result in exact coverage rates close to the nominal level and short length of the confidence sets. This holds true whether the magnitude of the break is small or large. In fact, we show that GL inference is bet-proof, a measure of "reasonableness" of frequentist inference in non-regular problems [see, e.g., Buehler (1959)].

The GL inference developed in this paper has been applied by Casini and Perron (2019b) to achieve finite-sample improvements under the continuous record asymptotic framework of Casini and Perron (2019a). The latter proposed an alternative asymptotic framework to explain the non-standard features of the finite-sample distribution of the least-squares estimator.

The paper is organized as follows. We first focus on the single change-point case. Section 2 presents the statistical setting. We develop the asymptotic theory in Section 3 and the inference methods in Section 4. Results for multiple change-points models are given in Section 5 while Section 6 discusses some theoretical properties of GL inference. Section 7 presents simulation results about the finite-sample performance. Section 8 concludes. All proofs are included in a supplement [Casini and Perron (2020)].

2 The Model and the Assumptions

This section introduces the structural change model with a single break, reviews the least-squares estimation method for the break date, and presents the relevant assumptions. We start with introducing the formal setup for our analysis. The following notation is used throughout. We denote the transpose of a matrix A by A'. We use $\|\cdot\|$ to denote the Euclidean norm of a linear space, i.e., $\|x\| = \left(\sum_{i=1}^{p} x_i^2\right)^{1/2}$ for $x \in \mathbb{R}^p$. For a matrix A, we use the vector-induced norm, i.e., $\|A\| = \sup_{x\neq 0} \|Ax\| / \|x\|$. All vectors are column vectors. For two vectors a and b, we write $a \leq b$ if the inequality holds component-wise. We use $\lfloor \cdot \rfloor$ to denote the largest smaller integer function. Boldface is used for sets. We use $\overset{\mathbb{P}}{\to}$ and $\overset{d}{\to}$ to denote convergence in probability and convergence in distribution, respectively. $\mathbb{C}_b(\mathbf{E})$ $[\mathbb{D}_b(\mathbf{E})]$ is the collection of bounded continuous [càdlàg] functions from some specfied set $\mathbf{E}$ to $\mathbb{R}$. Weak convergence on either $\mathbb{C}_b(\mathbf{E})$ or $\mathbb{D}_b(\mathbf{E})$ is denoted by $\Rightarrow$. The symbol "$\triangleq$" stands for definitional equivalence.

We consider a sample of observations $\{(y_t,\, w_t,\, z_t) :\, t = 1, \ldots, T\}$, defined on a filtered probability space $(\Omega,\, \mathscr{F},\, \mathbb{P})$, on which all of the random elements introduced in what follows are defined. The model is

$$y_t = w_t'\phi^0 + z_t'\delta_1^0 + e_t, \quad \left(t = 1, \ldots, T_b^0\right) \qquad y_t = w_t'\phi^0 + z_t'\delta_2^0 + e_t, \quad \left(t = T_b^0 + 1, \ldots, T\right) \tag{2.1}$$

where y_t is an scalar dependent variable, w_t and z_t are regressors of dimensions, p and q, respectively, and e_t is an unobserved error term. The true parameter vectors ϕ^0, δ_1^0 and δ_2^0 are unknown and we define $\delta^0 \triangleq \delta_2^0 - \delta_1^0$, with $\delta^0 \neq 0$ so that a structural change at date T_b^0. It is useful to re-parametrize the model. Letting $x_t \triangleq (w_t', z_t')'$ and $\beta^0 \triangleq \left((\phi^0)', (\delta_1^0)'\right)'$, we can rewrite the model as

$$y_t = x_t'\beta^0 + e_t, \quad \left(t = 1, \ldots, T_b^0\right) \qquad y_t = x_t'\beta^0 + z_t'\delta^0 + e_t, \quad \left(t = T_b^0 + 1, \ldots, T\right). \tag{2.2}$$

More generally, we can define $z_t \triangleq D'x_t$, where D is a $(p+q) \times q$ matrix with full column rank. A pure structural change model in which all regression parameters are subject to change corresponds to $D = I_{(p+q)\times(p+q)}$, whereas a partial structural change model arises when $D = (0_{q\times p}, I_{q\times q})'$. In order to facilitate the derivations, we reformulate model (2.2) in matrix format. Let $Y = (y_1, \ldots, y_T)'$, $X = (x_1, \ldots, x_T)'$, $e = (e_1, \ldots, e_T)'$, $X_1 = (x_1, \ldots, x_{T_b}, 0, \ldots, 0)'$, $X_2 = (0, \ldots, 0, x_{T_b+1}, \ldots, x_T)'$ and $X_0 = (0, \ldots, 0, x_{T_b^0+1}, \ldots, x_T)'$. Further, define Z_1, Z_2 and Z_0 in a similar way: $Z_1 = X_1D$, $Z_2 = X_2D$ and $Z_0 = X_0D$. We omit the dependence of the matrices X_i and Z_i $(i = 1, 2)$ on T_b. Then, (2.2) is equivalent to

$$Y = X\beta + Z_0\delta + e. \tag{2.3}$$

Let $\theta^0 \triangleq \left((\phi^0)', (\delta_1^0)', (\delta^0)'\right)'$ denote the true value of the parameter vector $\theta \triangleq (\phi, \delta_1, \delta)$. The break date least-squares (LS) estimator $\widehat{T}_b^{\mathrm{LS}}$ is the minimizer of the sum of squared residuals [denoted $S_T(\theta, T_b)$] from (2.3). The parameter θ can be concentrated out resulting in a criterion function depending only on T_b, i.e., $\widehat{T}_b^{\mathrm{LS}} = \arg\min_{1\leq T_b \leq T} S_T(\widehat{\theta}^{\mathrm{LS}}(T_b), T_b)$ where $\widehat{\theta}^{\mathrm{LS}}(T_b) = \arg\min_\theta S_T(\theta, T_b)$. We also have

$$\begin{aligned} \arg\min_{1\leq T_b\leq T} S_T(\widehat{\theta}^{\mathrm{LS}}(T_b), T_b) &= \arg\max_{T_b} \widehat{\delta}^{\mathrm{LS}\prime}(T_b)(Z_2'MZ_2)\widehat{\delta}^{\mathrm{LS}}(T_b) \\ &\triangleq \arg\max_{T_b} Q_T(\widehat{\delta}^{\mathrm{LS}}(T_b), T_b), \end{aligned} \tag{2.4}$$

where $M_X \triangleq I - X(X'X)^{-1}X'$, $\widehat{\delta}^{\mathrm{LS}}(T_b)$ is the least-squares estimator of δ^0 obtained by regressing Y on X and Z_2 and the statistic $Q_T\left(\widehat{\delta}^{\mathrm{LS}}(T_b), T_b\right)$ is the numerator of the sup-Wald statistic. The Laplace-type inference builds on the least-squares criterion function $Q_T(\delta(T_b), T_b)$, where $\delta(T_b)$ stands for $\widehat{\delta}^{\mathrm{LS}}(T_b)$ to minimize notational burden.

Assumption 2.1. $T_b^0 = \lfloor T\lambda_b^0 \rfloor$, *where* $\lambda_b^0 \in \Gamma^0 \subset (0, 1)$.

Assumption 2.2. *With* $\{\mathscr{F}_t, t = 1, 2, \ldots\}$ *a sequence of increasing* σ*-fields,* $\{z_te_t, \mathscr{F}_t\}$ *forms a* L^r*-mixingale sequence with* $r = 2 + \nu$ *for some* $\nu > 0$. *That is, there exist nonnegative constants* $\{\varrho_{1,t}\}_{t\geq 1}$ *and* $\{\varrho_{2,j}\}_{j\geq 0}$ *such that* $\varrho_{2,j} \to 0$ *as* $j \to \infty$, *for all* $t \geq 1$ *and* $j \geq 0$, *and we have for* $r \geq 1$: *(i)* $\|\mathbb{E}(z_te_t|\mathscr{F}_{t-j})\|_r \leq \varrho_{1,t}\varrho_{2,j}$, *(ii)* $\|z_te_t - \mathbb{E}(z_te_t|\mathscr{F}_{t-j})\|_r \leq \varrho_{1,t}\varrho_{2,j+1}$. *In addition, (iii)*

$\max_t \varrho_{1,t} < C_1 < \infty$ *and (iv)* $\sum_{j=0}^{\infty} j^{1+\nu} \varrho_{2,j} < \infty$ *for some* $\nu > 0$*, (v)* $\|z_t\|_{2r} < C_2 < \infty$ *and* $\|e_t\|_{2r} < C_3 < \infty$ *for some* $C_1,\, C_2,\, C_3 > 0$.

Assumption 2.3. *There exists an* $l_0 > 0$ *such that for all* $l > l_0$*, the minimum eigenvalues of* $H_l^* = (1/l) \sum_{T_b^0 - l + 1}^{T_b^0} x_t x_t'$ *and* $H_l^{**} = (1/l) \sum_{T_b^0 + 1}^{T_b^0 + l} x_t x_t'$ *are bounded away from zero. These matrices are invertible when* $l \geq p + q$ *and have stochastically bounded norms uniformly in* l.

Assumption 2.4. $T^{-1} X' X \xrightarrow{\mathbb{P}} \Sigma_{XX}$*, where* Σ_{XX}*, a positive definite matrix.*

These assumptions are standard and similar to those in Perron and Qu (2006). It is well-known that only the fractional break date λ_b^0 (not T_b^0) can be consistently estimated, with $\widehat{\lambda}_b^{\mathrm{LS}}$ having a T-rate of convergence. The corresponding result for the break date estimator $\widehat{T}_b^{\mathrm{LS}}$ states that, as T increases, $\widehat{T}_b^{\mathrm{LS}}$ remains within a bounded distance from T_b^0. However, this does not affect the estimation problem of the regression coefficients θ^0, for which $\widehat{\theta}^{\mathrm{LS}}$ is a regular estimator; i.e., $\sqrt{T}$-consistent and asymptotically normally distributed, since the estimation of the regression parameters is asymptotically independent from the estimation of the change-point. Hence, given the fast rate of convergence of $\widehat{\lambda}_b^{\mathrm{LS}}$, the regression parameters are essentially estimated as if the change-point was known. More complex is the derivation of the asymptotic distribution of $\widehat{\lambda}_b^{\mathrm{LS}}$; e.g., Hinkley (1971) for an i.i.d. Gaussian process with a change in the mean. Therefore, to make progress it is necessary to consider a shrinkage asymptotic setting in which the size of the shift converges to zero as $T \to \infty$; see Picard (1985) and Yao (1987) and extended by Bai (1997) to general linear models.

3 Generalized Laplace Estimation

We define the GL estimator in Section 3.1 and discuss its usefulness in Section 3.2. Section 3.3 describes the asymptotic framework under which we derive the limiting distribution with the results presented in Section 3.4.

3.1 The Class of Laplace Estimators

The class of GL estimators relies on the original least-squares criterion function $Q_T(\delta(T_b), T_b)$, with the parameter of interest being $\lambda_b^0 = T_b^0 / T$. With the criterion function $Q_T(\delta(T_b), T_b)$, or equivalently $Q_T(\delta(\lambda_b), \lambda_b)$, the Quasi-posterior $p_T(\lambda_b)$ is defined by the exponential transformation,

$$p_T(\lambda_b) \triangleq \frac{\exp\left(Q_T(\delta(\lambda_b), \lambda_b)\right) \pi(\lambda_b)}{\int_{\Gamma^0} \exp\left(Q_T(\delta(\lambda_b), \lambda_b)\right) \pi(\lambda_b)\, d\lambda_b}, \tag{3.1}$$

where $\pi(\cdot)$ is a weighting function. Note that $p_T(\lambda_b)$ defines a proper distribution over the parameter space Γ^0. The $\mathscr{L}(\theta, T_b)$-class of estimators are the solutions of smooth convex optimization problems for a given loss function, restricting attention to convex loss functions $l_T(\cdot)$. Examples include (a) $l_T(r) = a_T^m |r|^m$, the polynomial loss function (the squared loss function is obtained when $m = 2$ and the absolute deviation loss function when $m = 1$); (b) $l_T(r) = a_T(\tau - \mathbf{1}(r \leq 0))r$, the check loss function; where a_T is a divergent sequence. We define the Expected Risk function, under the density $p_T(\cdot)$ and the loss $l_T(\cdot)$ as $\mathcal{R}_{l,T}(s) \triangleq \mathbb{E}_{p_T}\left[l_T\left(s - \widetilde{\lambda}_b\right)\right]$, where $\widetilde{\lambda}_b$ is a random variable with distribution p_T and $\mathbb{E}_{p_T}$ denotes expectation taken under p_T. Using (3.1) we have,

$$\mathcal{R}_{l,T}(s) \triangleq \int_{\Gamma^0} l_T(s - \lambda_b) p_T(\lambda_b) d\lambda_b. \tag{3.2}$$

The Laplace-type estimator $\widehat{\lambda}_b^{\mathrm{GL}}$ shall be interpreted as a decision rule that, given the information contained in the Quasi-posterior p_T, is least unfavorable according to the loss function l_T and the prior density π. Then $\widehat{\lambda}_b^{\mathrm{GL}}$ is the minimizer of the expected risk function (3.2), i.e., $\widehat{\lambda}_b^{\mathrm{GL}} \triangleq \arg\min_{s\in\Gamma^0}[\mathcal{R}_{l,T}(s)]$. Observe that the GL estimator $\widehat{\lambda}_b^{\mathrm{GL}}$ results in the mean (median) of the Quasi-posterior upon choosing the squared (absolute deviation) loss function. The choice of the loss and of the prior density functions hinges on the statistical problem addressed. In the structural change problem, a natural choice for the Quasi-prior π is the density of the asymptotic distribution of $\widehat{\lambda}_b^{\mathrm{LS}}$. This requires to replace the population quantities appearing in that distribution by consistent plug-in estimates—cf. Bai and Perron (1998)—and derive its density via simulations as in Casini and Perron (2019a). The attractiveness of the Quasi-posterior (3.1) is that it provides additional information about the parameter of interest λ_b^0 beyond what is already included in the point estimate $\widehat{\lambda}_b^{\mathrm{LS}}$ and its distribution (see Section 3.2). This approach will result in more accurate inference in finite-samples even in cases with high uncertainty in the data as we shall document in Section 7. This is supported in Section 6 showing that the GL inference is bet-proof which is a desirable theoretical property in non-regular problems.

Assumption 3.1. *Let $l_T(r) \triangleq l(a_T r)$, with a_T a positive divergent sequence. $\boldsymbol{L}$ denotes the set of functions $l : \mathbb{R} \to \mathbb{R}_+$ that satisfy (i) $l(r)$ is defined on $\mathbb{R}$, with $l(r) \geq 0$ and $l(r) = 0$ if and only if $r = 0$; (ii) $l(r)$ is continuous at $r = 0$; (iii) $l(\cdot)$ is convex and $l(r) \leq 1 + |r|^m$ for some $m > 0$.*

Assumption 3.2. *$\pi : \mathbb{R} \to \mathbb{R}_+$ is a continuous, uniformly positive density function satisfying $\pi^0 \triangleq \pi(\lambda_b^0) > 0$, and for some finite $C_\pi < \infty$, $\pi^0 < C_\pi$. Also, $\pi(\lambda_b) = 0$ for all $\lambda_b \notin \Gamma^0$, and π is twice continuously differentiable with respect to λ_b at λ_b^0.*

Assumption 3.1 is similar to those in Bickel and Yahav (1969), Ibragimov and Has'minskiĭ (1981) and Chernozhukov and Hong (2003). The convexity assumption on $l_T(\cdot)$ is guided by practical considerations. The dominant restriction in part (iii) is conventional and implicitly assumes that the loss function has been scaled by some constant. What is important is that the growth of the function $l_T(r)$ as $|r| \to \infty$ is slower than $\exp(\epsilon |r|)$ for any $\epsilon > 0$. Assumption 3.2 on the prior is satisfied for any reasonable choice. For priors that have a peak at λ_b^0 one can applies some smoothing techniques to make it differentiable locally. The large-sample properties of the $\mathscr{L}(\theta, T_b)$-class are studied under the shrinkage asymptotic setting of Bai (1997) and Bai and Perron (1998). Thus, we need the following assumption.

Assumption 3.3. *Let $\delta_T \triangleq \delta_T^0 \triangleq v_T \delta^0$ where $v_T > 0$ is a scalar satisfying $v_T \to 0$ as $T \to \infty$ and $T^{1/2-\vartheta} v_T \to \infty$ for some $\vartheta \in (0, 1/4)$.*

We omit the superscript 0 from δ_T^0 for notational convenience since it should not cause any confusion. Assumption 3.3 requires the magnitude of the break to shrink to zero at any slower rate than $T^{-1/2}$. The specific rates allowed differ from those in Bai (1997) and Bai and Perron (1998), since they require $\vartheta \in (0, 1/2)$. The reason is merely technical; the asymptotics of the Laplace-type estimator involve smoothing the criterion function, and thus one needs to guarantee that $\widehat{\lambda}_b$ approaches λ_b^0 at a sufficiently fast rate. Under the shrinkage asymptotics, Proposition 1 and Corollary 1 in Bai (1997) state that $T \|\delta_T\|^2 \left(\widehat{\lambda}_b^{\mathrm{LS}} - \lambda_b^0\right) = O_{\mathbb{P}}(1)$ and $\widehat{\delta}_T^{\mathrm{LS}} - \delta_T = o_{\mathbb{P}}(1)$, respectively.

3.2 Discussion about the GL Approach

We use Figure 1-2 to illustrate the main idea behind the usefulness of the GL method. They present plots of the density of the distribution of $\widehat{T}_b^{\mathrm{LS}}$ for the simple model $y_t = \phi^0 + z_t (\delta_1^0 + \delta^0 \mathbf{1}\{t > T_b^0\}) + e_t$ where $\{z_t\}$ follows an ARMA(1,1) process and $e_t \sim i.i.d.\, \mathscr{N}(0, 1)$. The distributions presented are the exact finite-sample distribution and Bai's (1997) classical large-N limit distribution—the span N here coincides with the sample size T. Noteworthy is the non-standard features of the finite-sample distribution when the break magnitude is small, which include multi-modality, fat tails and asymmetry—the latter if T_b^0 is not at mid-sample. The central mode is near $\widehat{T}_b^{\mathrm{LS}}$ while the other two modes are in the tails near the start and end of the sample period; when the break magnitude is small $\widehat{T}_b^{\mathrm{LS}}$ tends to locate the break in the tails since the evidence of a break is weak. It is evident that the classical large-N asymptotic distribution provides a poor approximation especially for small break sizes.

The GL method is useful because it weights the information from the least-squares criterion function with the information from the prior density—which, here, is the density of the asymptotic distribution of $\widehat{T}_b^{\mathrm{LS}}$. Note that the least-squares objective function is quite flat when the magnitude of the break is small and so $\widehat{T}_b^{\mathrm{LS}}$ is imprecise, while the resulting Quasi-posterior, or, e.g., the median of the Quasi-posterior, may lead to better estimates in finite-samples, because it takes into account the overall shape of the objective function which weighted by the prior becomes more informative about the uncertainty of the break date.

3.3 Normalized Version of $\mathcal{R}_{l,T}(s)$

In order to develop the asymptotic results, we introduce an input parameter sequence $\{\gamma_T\}$ whose properties are specified below and work with a normalized version of $\mathcal{R}_{l,T}(s)$ in order to be able to derive the relevant limit results. We assume that $\lambda_b^0 \in \Gamma^0 \subset (0,\,1)$ is the unknown extremum of $\widetilde{Q}(\theta^0,\,\lambda_b) = \mathbb{E}[Q_T(\theta^0,\,\lambda_b)]$ and that $\theta^0 \triangleq \left((\phi^0)',\,(\delta_1^0)',\,(\delta^0)'\right)' \in \mathbf{S} \subset \mathbb{R}^p \times \mathbb{R}^q \times \mathbb{R}^q$. Our analysis is within a vanishing neighborhood of θ^0. For any $\theta \in \mathbf{S}$, let $\lambda_b^0(\theta)$ be an arbitrary element of $\Gamma^0(\theta) \triangleq \left\{\lambda_b \in \Gamma^0 :\, \widetilde{Q}(\theta,\,\lambda_b) = \sup_{\widetilde{\lambda}_b \in \Gamma^0} \widetilde{Q}\left(\theta,\,\widetilde{\lambda}_b\right)\right\}$. Provided a uniqueness condition is assumed (see Assumption 3.6), $\Gamma^0(\theta)$ contains a single element, λ_b^0. Further, let $\overline{Q}_T(\theta,\,\lambda_b) \triangleq Q_T(\theta,\,\lambda_b) - Q_T(\theta,\,\lambda_b^0)$, $Q_T^0(\theta,\,\lambda_b) \triangleq \mathbb{E}[Q_T(\theta,\,\lambda_b) - Q_T(\theta,\,\lambda_b^0)\,|\,X]$, and $G_T(\theta,\,\lambda_b) \triangleq \overline{Q}_T(\theta,\,\lambda_b) - Q_T^0(\theta,\,\lambda_b)$. These expressions are given by $G_T(\theta,\,\lambda_b) = g_e(\theta,\,\lambda_b)$, $Q_T^0 = g_d(\theta,\,\lambda_b)$ and $\overline{Q}_T = g_d(\theta,\,\lambda_b) + g_e(\theta,\,\lambda_b)$, where

$$g_d(\theta,\,\lambda_b) = \delta_T' \left\{(Z_0' M Z_2)(Z_2' M Z_2)^{-1}(Z_2' M Z_0) - Z_0' M Z_0\right\} \delta_T, \tag{3.3}$$

and

$$\begin{aligned} g_e(\theta,\,\lambda_b) &= 2\delta_T'(Z_0' M Z_2)(Z_2' M Z_2)^{-1} Z_2 M e - 2\delta_T'(Z_0' M e) \\ &\quad + e' M Z_2 (Z_2' M Z_2)^{-1} Z_2 M e - e' M Z_0 (Z_0' M Z_0)^{-1} Z_0' M e. \end{aligned} \tag{3.4}$$

They are derived in Section A.2. The GL estimator $\widehat{\lambda}_b^{\mathrm{GL}}(\theta)$ is then the minimizer of a normalized version of $\mathcal{R}_{l,T}(s)$:

$$\begin{aligned} \Psi_{l,T}(s;\,\theta) &= \int_{\Gamma^0} l(s-\lambda_b) \frac{\exp\left(\left(\gamma_T / \left(T\|\delta_T\|^2\right)\right) \overline{Q}_T(\theta,\,\lambda_b)\right)\pi(\lambda_b)}{\int_{\Gamma^0} \exp\left(\left(\gamma_T / \left(T\|\delta_T\|^2\right)\right) \overline{Q}_T(\theta,\,\lambda_b)\right)\pi(\lambda_b)\,d\lambda_b} d\lambda_b \\ &= \int_{\Gamma^0} l(s-\lambda_b) \frac{\exp\left(\left(\gamma_T / \left(T\|\delta_T\|^2\right)\right)\left(G_T(\theta,\,\lambda_b) + Q_T^0(\theta,\,\lambda_b)\right)\right)\pi(\lambda_b)}{\int_{\Gamma^0} \exp\left(\left(\gamma_T / \left(T\|\delta_T\|^2\right)\right)\left(G_T(\theta,\,\lambda_b) + Q_T^0(\theta,\,\lambda_b)\right)\right)\pi(\lambda_b)\,d\lambda_b} d\lambda_b. \end{aligned} \tag{3.5}$$

Note that, under Condition 1 below, this is equivalent to the minimizer of $\mathcal{R}_{l,T}(s)$ since $\overline{Q}_T(\theta,\,\lambda_b)$ can always be normalized without affecting its maximization. Different choices of $\{\gamma_T\}$ give rise to GL estimators with different limiting distributions. Using δ_T or any consistent estimate (e.g., $\widehat{\delta}_T^{\mathrm{LS}}$) in the factor $\gamma_T / \left(T\|\delta_T\|^2\right)$ is irrelevant because they are asymptotically equivalent. Our analysis is local in nature and thus we write $\widehat{\lambda}_b^{\mathrm{GL}}(\widehat{\theta}) \triangleq \widehat{\lambda}_b^{\mathrm{GL},*}(r_T(\widehat{\theta}-\theta^0),\, r_T(\widehat{\theta}-\theta^0))$, where r_T is the convergence rate of $\widehat{\theta}-\theta^0$. Note that $G_T(\cdot,\,\cdot)$ and $Q_T^0(\cdot,\,\cdot)$ constitute the stochastic and the deterministic part of the objective function, respectively. Both depend on $r_T(\widehat{\theta}-\theta^0)$ and our proof proceeds in conditioning first on the effect of $r_T(\widehat{\theta}-\theta^0)$ on the deterministic part to obtain weak convergence of the stochastic part to a limit process that does not depend on this conditioning.

See below for more details. Hence, it is required to introduce two indices $\widetilde{v}$ and v, such that we define $\widehat{\lambda}_b^{\mathrm{GL}}(\widehat{\theta}) = \widehat{\lambda}_b^{\mathrm{GL},*}(\widetilde{v},\, v)$ as the minimizer of

$$\Psi_{l,T}(s;\,\widetilde{v},\,v) \triangleq \int_{\Gamma^0} l(s-\lambda_b)\times \tag{3.6}$$
$$\frac{\exp\left(\left(\gamma_T/\left(T\left\|\delta_T\right\|^2\right)\right)\left(G_T\left(\theta^0+\widetilde{v}/r_T,\,\lambda_b\right)+Q_T^0\left(\theta^0+v/r_T,\,\lambda_b\right)\right)\right)\pi\left(\lambda_b\right)}{\int_{\Gamma^0}\exp\left(\left(\gamma_T/\left(T\left\|\delta_T\right\|^2\right)\right)\left(G_T\left(\theta^0+\widetilde{v}/r_T,\,\lambda_b\right)+Q_T^0\left(\theta^0+v/r_T,\,\lambda_b\right)\right)\right)\pi\left(\lambda_b\right)d\lambda_b}d\lambda_b.$$

For each v, we show weak convergence as a function of $\widetilde{v}$ to a limit process that does not depend on v. In a second step, we use the monotonicity in v of Q_T^0 which, relying on the argument in Jurečková (1977), allows us to achieve weak convergence uniformly in v.

We first show the consistency and rate of convergence of $\widehat{\lambda}_b^{\mathrm{GL}}$. These results imply that θ^0 is estimated as if T_b^0 were known. Thus, $\widehat{\theta}$ is $\sqrt{T}$-consistent and asymptotically normal so that we set $r_T = \sqrt{T}$ hereafter. Since $\widehat{\lambda}_b^{\mathrm{GL}}$ is defined implicitly as an extremum estimator, its large-sample properties can be derived as follows. We first show, for each pair $(v,\,\widetilde{v})$ with $v,\,\widetilde{v}\in\mathbf{V}$, the convergence of the marginal distributions of the sample function $\Psi_{l,T}\left(s;\,v,\,\widetilde{v}\right)$ to the marginal distributions of the random function

$$\Psi_l^0\left(s\right) = \int_{\mathbb{R}} l\left(s-u\right)\left(\mathscr{V}\left(u\right)/\int_{\mathbb{R}}\mathscr{V}\left(v\right)dv\right)du,$$

where the limit process $\Psi_l^0\left(s\right)$ does not depend on v nor $\widetilde{v}$. Next, we show that the family of probability measures in $\mathbb{C}_b\left(\mathbf{K}\right)$, with $\mathbf{K}\triangleq\{s\in\mathbb{R}:\,|s|\leq K \text{ and } K<\infty\}$, generated by the contractions of $\Psi_{l,T}\left(s;\,\widetilde{v},\,v\right)$ on $\mathbf{K}$ is dense uniformly in $(v,\,\widetilde{v})$. Finally, we examine the oscillations of the minimizers of the sample criterion $\Psi_{l,T}\left(s;\,v,\,\widetilde{v}\right)$.

It is important to note that the results derived in this section are more general than what is required for the structural change model. The reason is that the change-point model is recovered as a special case corresponding to $\Psi_{l,T}\left(s\right) = \Psi_{l,T}\left(s;\,0,\,0\right)$. That is, defining the GL estimator in a $1/r_T$-neighborhood of the slope parameter vector θ^0 is not strictly necessary and one can essentially develop the same analysis with θ fixed at its true value θ^0. This relies on the properties of (orthogonal) least-squares projections and would not apply, for example, to the least absolute deviation (LAD) estimator of the break date [cf. Bai (1995)] for which $\Psi_{l,T}\left(s;\,\widetilde{v},\,v\right)$ should instead be considered. The same issue is present when estimating structural changes in the quantile regression model [cf. Oka and Qu (2010)] and in using instrumental variables models [cf. Hall, Han, and Boldea (2010) and Perron and Yamamoto (2014; 2015)]. We establish theoretical results under this more general setting since they may be useful for future work.

Let $\lambda_{b,T}^0\left(v\right) = \lambda_{b,T}^0\left(\theta^0+v/r_T\right)$, and $\{\psi_T\}$, $\{\gamma_T\}$ denote some sequences that increase to infinity with T and whose exact properties are specified below. Introduce the local parameter

$u=\psi_T\left(\lambda_b-\lambda_{b,T}^0(v)\right)$ and let $\pi_{T,v}(u)\triangleq\pi\left(\lambda_{b,T}^0(v)+u/\psi_T\right)$, $Q_{T,v}(u)\triangleq Q_T^0(\theta^0+v/r_T,\,\lambda_{b,T}^0(v)+u/\psi_T)$, and $\widetilde{G}_{T,v}(u,\,\widetilde{v})\triangleq G_T\left(\theta^0+\widetilde{v}/r_T,\,\lambda_{b,T}^0(v)+u/\psi_T\right)$. We characterize ψ_T with the results on consistency and rate of convergence of $\widehat{\lambda}_b^{\mathrm{GL}}$ in Proposition 3.1. Apply a simple substitution in (3.6) to yield,

$$\Psi_{l,T}(s;\,\widetilde{v},\,v)=\int_{\Gamma_T}l(s-u)\,\frac{\exp\left(\left(\gamma_T/T\left\|\delta_T\right\|^2\right)\left(\widetilde{G}_{T,v}(u,\,\widetilde{v})+Q_{T,v}(u)\right)\right)\pi_{T,v}(u)\,du}{\int_{\Gamma_T}\exp\left(\left(\gamma_T/T\left\|\delta_T\right\|^2\right)\left(\widetilde{G}_{T,v}(w,\,\widetilde{v})+Q_{T,v}(w)\right)\right)\pi_{T,v}(w)\,dw},\qquad(3.7)$$

where $\Gamma_T\triangleq\{u\in\mathbb{R}:\,\lambda_b^0+u/\psi_T\in\mathit{\Gamma}^0\}$.

Assumption 3.4. *$\{(z_t,\,e_t)\}$ is second-order stationary within each regime such that $\mathbb{E}(z_tz_t')=V_1$ and $\mathbb{E}(e_t^2)=\sigma_1^2$ for $t\leq T_b^0$ and $\mathbb{E}(z_tz_t')=V_2$ and $\mathbb{E}(e_t^2)=\sigma_2^2$ for $t>T_b^0$.*

Assumption 3.5. *For $r\in[0,\,1]$, $(T_b^0)^{-1/2}\sum_{t=1}^{\lfloor rT_b^0\rfloor}z_te_t\Rightarrow\mathscr{G}_1(r)$ and $(T-T_b^0)^{-1/2}\sum_{t=T_b^0+1}^{T_b^0+\lfloor r(T-T_b^0)\rfloor}z_te_t\Rightarrow\mathscr{G}_2(r)$, where $\mathscr{G}_i(\cdot)$ is a multivariate Gaussian process on $[0,\,1]$ with zero mean and covariance $\mathbb{E}[\mathscr{G}_i(u),\,\mathscr{G}_i(s)]=\min\{u,\,s\}\,\Sigma_i$ $(i=1,\,2)$, and $\Sigma_1\triangleq\lim_{T\to\infty}\mathbb{E}\left[(T_b^0)^{-1/2}\sum_{t=1}^{T_b^0}z_te_t\right]^2$, $\Sigma_2\triangleq\lim_{T\to\infty}\mathbb{E}\left[(T-T_b^0)^{-1/2}\sum_{t=T_b^0+1}^{T}z_te_t\right]^2$. Furthermore, for any $0<r_0<1$ with $r_0<\lambda_0$, $T^{-1}\sum_{t=\lfloor r_0T\rfloor+1}^{\lfloor\lambda_0T\rfloor}z_tz_t'\overset{\mathbb{P}}{\to}(\lambda_0-r_0)V_1$, and with $\lambda_0<r_0$ $T^{-1}\sum_{t=\lfloor\lambda_0T\rfloor+1}^{\lfloor r_0T\rfloor}z_tz_t'\overset{\mathbb{P}}{\to}(r_0-\lambda_0)V_2$ so that λ_- and λ_+ (the minimum and maximum of the eigenvalues of the last two matrices) satisfy $0<\lambda_-\leq\lambda_+<\infty$.*

Assumption 3.4-3.5 are equivalent to A9 in Bai (1997) and A7 in Bai and Perron (1998). More specifically, Assumption 3.5 requires that, within each regime, an Invariance Principle holds for $\{z_te_t\}$. Let $\zeta_t\triangleq z_te_t$. For $u\leq0$ let $g(\zeta_t;\,u)\triangleq(\delta^0)'\sum_{t=T_b^0+\lfloor u/v_T^2\rfloor}^{T_b^0}\zeta_t$ and $\widetilde{g}(\zeta_t;\,u,\,\widetilde{v},\,v;\,\psi_T,\,r_T)\triangleq\sqrt{\psi_T}(\delta^0+\widetilde{v}/r_T)'\sum_{t=T\lambda_b^0(\theta^0+v/r_T)+\lfloor u/\psi_T\rfloor}^{T\lambda_b^0(\theta^0+v/r_T)}\zeta_t$. Define analogously $g(\zeta_t;\,u)$ and $\widetilde{g}(\zeta_t;\,u,\,\widetilde{v},\,v;\,\psi_T,\,r_T)$ for the case $T_b>T_b^0$. We now present some technical assumptions that are necessary for the derivation of the asymptotic results for the GL estimate.

Assumption 3.6. *For some neighborhood $\Theta^0\subset\mathbf{S}$ of θ^0, (i) for all $\lambda_b\neq\lambda_b^0$, $\widetilde{Q}(\theta^0,\,\lambda_b)<\widetilde{Q}(\theta^0,\,\lambda_b^0)$; (ii) for any $v,\,\widetilde{v}_1,\,\widetilde{v}_2\in\mathbf{V}$ and $u,\,s\in\mathbb{R}$, $\mathit{\Sigma}(u,\,s)\triangleq\lim_{T\to\infty}\mathbb{E}\left[\widetilde{g}(\zeta_t;\,u,\,\widetilde{v}_1,\,v;\,\psi_T,\,r_T)\,\widetilde{g}(\zeta_t;\,s,\,\widetilde{v}_2,\,v;\,\psi_T,\,r_T)'\right]$ does not depend on $v,\,\widetilde{v}_1,\,\widetilde{v}_2\in\mathbf{V}$.*

Part (i) of Assumption 3.6 is an identification condition. Assumption 3.6-(ii) holds whenever $\widehat{\lambda}_b$ is consistent. With Assumption 3.6-(ii) we fully characterize the Gaussian component of the limit process $\mathscr{V}(\cdot)$; it implies that $\mathit{\Sigma}(\cdot,\,\cdot)$ is strictly positive and that

$$\forall u,\,s\in\mathbb{R}:\begin{cases}\forall c>0:\;\mathit{\Sigma}(cu,\,cs)=c\mathit{\Sigma}(u,\,s)\,,\\ \mathit{\Sigma}(u,\,u)+\mathit{\Sigma}(s,\,s)-2\mathit{\Sigma}(u,\,s)=\mathit{\Sigma}(u-s,\,u-s)\,,\end{cases}\qquad(3.8)$$

where the second implication requires some simple but tedious manipulations. Finally, the following assumption is automatically satisfied if $l(\cdot)$ is a convex function with a unique minimum.

Assumption 3.7. *$\xi_l^0 \triangleq \xi(\lambda_b^0)$ is uniquely defined by*

$$\Psi_l\left(\xi_l^0\right) \triangleq \inf_s \Psi_l(s) = \inf_s \int_{\mathbb{R}} l(s-u)\left(\exp(\mathscr{V}(u)) / \left(\int_{\mathbb{R}} \exp(\mathscr{V}(w))\,dw\right)\right) du,$$

where

$$\mathscr{V}(s) \triangleq \mathscr{W}(s) - \Lambda^0(s) \triangleq \begin{cases} 2\left((\delta^0)' \Sigma_1 \delta^0\right)^{1/2} W_1(-s) - |s|\,(\delta^0)' V_1 \delta^0, & \text{if } s \leq 0 \\ 2\left((\delta^0)' \Sigma_2 \delta^0\right)^{1/2} W_2(s) - s\,(\delta^0)' V_2 \delta^0, & \text{if } s > 0, \end{cases} \tag{3.9}$$

and W_1, W_2 are independent standard Wiener processes defined on $[0, \infty)$.

3.4 Asymptotic Results for the GL Estimate

We first show the consistency and rate of convergence of the GL estimator. The latter allows us to characterize the rate of ψ_T and proceed with the asymptotic analysis in a neighborhood of λ_b^0. In practice, the squared loss function is often employed. Hence, it is useful to first present in Theorem 3.1, the theoretical results for this case for which the GL estimator is $\widehat{\lambda}_b^{\mathrm{GL}} = \int_{\Gamma^0} \lambda_b p_T(\lambda_b)\,d\lambda_b$, i.e., the Quasi-posterior mean. This allows us to keep the theoretical results tractable and provide the main intuition without the need of complex notation. This case is also instructive since we can compare our results with corresponding ones for the least-squares and Bayesian change-point estimators. Corresponding results for general loss functions are given in Theorem 3.2.

3.4.1 Consistency and Rate of Convergence

The rate of convergence is similar to that of the LS estimator; the difference being that $\psi_T = T^{1-2\vartheta}$ with $\vartheta \in (0, 1/2)$ for the LS estimator and $\vartheta \in (0, 1/4)$ for the GL estimator.

Proposition 3.1. *Under Assumption 2.1-2.4, 3.1-3.3 and 3.6-(i): (i) $\widehat{\lambda}_b^{\mathrm{GL}} = \lambda_b^0 + o_{\mathbb{P}}(1)$; (ii) $\widehat{\lambda}_b^{\mathrm{GL}} = \lambda_b^0 + O_{\mathbb{P}}\left(\left(T\|\delta_T\|^2\right)^{-1}\right)$.*

3.4.2 The Asymptotic Distribution of the Quasi-posterior Mean

For the squared loss function $\widehat{\lambda}_b^{\mathrm{GL}}\left(\widehat{\theta}\right) \triangleq \widehat{\lambda}_b^{\mathrm{GL},*}(\widetilde{v}, v)$, where

$$\widehat{\lambda}_b^{\mathrm{GL},*}(\widetilde{v}, v) \triangleq \frac{\int_{\Gamma^0} \lambda_b \exp\left(\left(\gamma_T / \left(T\|\delta_T\|^2\right)\right)\left(G_T(\theta^0 + \widetilde{v}/r_T, \lambda_b) + Q_T^0(\theta^0 + v/r_T, \lambda_b)\right)\right)\pi(\lambda_b)\,d\lambda_b}{\int_{\Gamma^0} \exp\left(\left(\gamma_T / \left(T\|\delta_T\|^2\right)\right)\left(G_T(\theta^0 + \widetilde{v}/r_T, \lambda_b) + Q_T^0(\theta^0 + v/r_T, \lambda_b)\right)\right)\pi(\lambda_b)\,d\lambda_b}, \tag{3.10}$$

and v, $\widetilde{v}$ each belong to some compact set $\mathbf{V} \subset \mathbb{R}^{p+2q}$. For each $v \in \mathbf{V}$, we consider $\widehat{\lambda}_b^{\mathrm{GL},*}(\cdot, v)$ as a random process with paths in $\mathbb{D}_b(\mathbf{V})$. We focus on the weak convergence of $\widehat{\lambda}_b^{\mathrm{GL},*}(\cdot, v)$ for fixed v since the limit process is independent of v and constant as a function of $\widetilde{v}$; we then exploit monotonicity in v. More precisely, we will show that for $\lambda_{b,T}^0(v) = \lambda_{b,T}^0(\theta^0 + v/r_T)$ and diverging sequences $\{\gamma_T\}$ and $\{r_T\}$, the sequence $a_T\left(\widehat{\lambda}_b^{\mathrm{GL},*}(\widetilde{v}, v) - \lambda_{b,T}^0(v)\right)$ converges in distribution in $\mathbb{D}_b(\mathbf{V})$ for each v to a limit process not depending on v nor $\widetilde{v}$. Since it is monotonic in v, we do not need to show uniform convergence directly. Introduce the local parameter $u = \psi_T\left(\lambda_b - \lambda_{b,T}^0(v)\right)$; a simple substitution in (3.10) yields,

$$\psi_T\left(\widehat{\lambda}_b^{\mathrm{GL},*}(\widetilde{v}, v) - \lambda_{b,T}^0(v)\right) = \frac{\int_{\mathbb{R}} u \exp\left(\left(\gamma_T/\left(T\|\delta_T\|^2\right)\right)\left(\widetilde{G}_{T,v}(u, \widetilde{v}) + Q_{T,v}(u)\right)\right)\pi_{T,v}(u)\,du}{\int_{\mathbb{R}} \exp\left(\left(\gamma_T/\left(T\|\delta_T\|^2\right)\right)\left(\widetilde{G}_{T,v}(u, \widetilde{v}) + Q_{T,v}(u)\right)\right)\pi_{T,v}(u)\,du}, \tag{3.11}$$

where again we have used the notation $\pi_{T,v}(u) = \pi\left(\lambda_{b,T}^0(v) + u/\psi_T\right)$, $Q_{T,v}(u) = Q_T^0(\theta^0 + v/r_T\ \lambda_{b,T}^0(v) + u/\psi_T\big)$ and $\widetilde{G}_{T,v}(u, \widetilde{v}) = G_T\left(\theta^0 + \widetilde{v}/r_T,\ \lambda_{b,T}^0(v) + u/\psi_T\right)$. The limit of the GL estimator depends on the limit of the process $\left(\gamma_T/\left(T\|\delta_T\|^2\right)\right)\left(\widetilde{G}_{T,v}(u, \widetilde{v}) + Q_{T,v}(u)\right)$. As part of the proof of Theorem 3.1, we show that the sequence of processes $\left\{\widetilde{G}_{T,v}(u, \widetilde{v}),\ T \geq 1\right\}$ converges weakly in $\mathbb{D}_b(\mathbb{R} \times \mathbf{V})$ to a Gaussian process $\mathscr{W}$ not varying with v, whereas $Q_{T,v}(\cdot)$ is approximated by a (deterministic) drift process taking negative values, and is monotonic in v and flat in $\widetilde{v}$. We show that this implies that $\widehat{\lambda}_b^{\mathrm{GL},*}(\widetilde{v}, v) - \lambda_{b,T}^0(v)$ is monotonic in v which then leads to uniform convergence in v following the argument of Jurečková (1977).

In anticipation of the results, we make a few comments about the notation for the weak convergence of processes on the space of bounded càdlàg functions $\mathbb{D}_b$. Let $\mathbf{V} \subset \mathbb{R}^{p+2q}$ be a compact set. Let $W_T(u, \widetilde{v}, v)$ denote an arbitrary sample process with bounded càdlàg paths evaluated at the local parameters $u \in \mathbb{R}$, and $v, \widetilde{v} \in \mathbf{V}$. For each fixed $v \in \mathbf{V}$, we shall write $W_T(u, \widetilde{v}, v) \Rightarrow \mathscr{W}(u, \widetilde{v}, v)$ in $\mathbb{D}_b(\mathbb{R} \times \mathbf{V})$ whenever the process $W_T(\cdot, \cdot, v)$ converges weakly to $\mathscr{W}(\cdot, \cdot, v)$, where $\mathscr{W}(\cdot, \cdot, v)$ also belongs to $\mathbb{D}_b(\mathbb{R} \times \mathbf{V})$. As a shorthand, we shall omit the argument u ($\widetilde{v}$) if the limit process does not depend on u ($\widetilde{v}$). The same notational conventions are used for the case when W_T is only a function of $(\widetilde{v}, v)$. In Theorem 3.1 the convergence holds for every $v \in \mathbf{V}$, stated as convergence in $\mathbb{D}_b$.

Condition 1. As $T \to \infty$ there exist a positive finite number κ_γ such that $\gamma_T/T\|\delta_T\|^2 \to \kappa_\gamma$.

Theorem 3.1. *Assume $l(\cdot)$ is the squared loss function. Under Assumption 2.1-2.4 and 3.1-3.7, and Condition 1, then in $\mathbb{D}_b$,*

$$T\|\delta_T\|^2\left(\widehat{\lambda}_b^{\mathrm{GL}} - \lambda_b^0\right) \Rightarrow \frac{\int u \exp\left(\mathscr{W}(u) - \Lambda^0(u)\right)du}{\int \exp\left(\mathscr{W}(u) - \Lambda^0(u)\right)du} \triangleq \int u p_0^*(u)\,du, \tag{3.12}$$

where $\mathscr{W}(\cdot)$ and $\Lambda^0(\cdot)$ are defined in (3.9).

Theorem 3.1 states that the asymptotic distribution of the GL estimate is a ratio of integrals of functions of tight Gaussian processes. We shall compare this result with the limiting distribution of the Bayesian change-point estimator of Ibragimov and Has'minskiĭ (1981). They considered a simple diffusion process with a change-point in the deterministic drift [see their eq. (2.17) on pp. 338]. The limiting distribution of the GL estimate from Theorem 3.1 for the case of a break in the mean for model (2.1) is essentially the same as theirs. Hence, while the GL estimator has a classical (frequentist) interpretation, it is first-order equivalent in law to a corresponding Bayes-type estimator.

We now present a result about the dual nature of the limiting distribution of the GL estimator. The following proposition shows that, under different conditions on the smoothing sequence parameter $\{\gamma_T\}$, the GL estimator achieves different limiting distributions.

Condition 2. As $T \to \infty$, $T \|\delta_T\|^2 / \gamma_T = o(1)$.

Proposition 3.2. *Assume $l(\cdot)$ is the squared loss function. Under Assumption 2.1-2.4 and 3.1-3.7, and Condition 2, $T\|\delta_T\|^2 \left(\widehat{\lambda}_b^{\mathrm{GL}} - \lambda_b^0\right) \Rightarrow \arg\max_{s\in\mathbb{R}} \mathscr{V}(s)$ in $\mathbb{D}_b$.*

Corollary 3.1. *Define $\Xi_e \triangleq (\delta^0)' \Sigma_2 \delta^0 / (\delta^0)' \Sigma_1 \delta^0$ and $\Xi_Z \triangleq (\delta^0)' V_2 \delta^0 / (\delta^0)' V_1 \delta^0$. Under Assumption 2.1-2.4 and 3.1-3.7, and Condition 2, $\left((\delta_T' V_1 \delta_T)^2 / \delta_T' \Sigma_1 \delta_T\right) \left(\widehat{T}_b^{\mathrm{GL}} - T_{b,T}^0\right) \xrightarrow{d} \arg\max_{s\in\mathbb{R}} \mathscr{V}^*(s)$ in $\mathbb{D}_b$ where*

$$\mathscr{V}^*(s) = W_1(-s) - |s|/2 \text{ if } s \leq 0; \;\; \mathscr{V}^*(s) = \Xi_e^{1/2} W_2(s) - \Xi_Z s/2 \text{ if } s > 0.$$

Corollary 3.1 and Proposition 3.2 show that with enough smoothing is applied, the GL estimator is (first-order) asymptotically equivalent to the least-squares or MLE [cf. Bai (1997) and Yao (1987), respectively]. The intuition is that when the criterion function is sufficiently smoothed, the Quasi-posterior probability density converges to the generalized dirac probability measure concentrated at the argmax of the limit criterion function. This is analogous to a well-known result [cf. Corollary 5.11 in Robert and Casella (2004)], stating that in a parametric statistical experiment indexed by a parameter $\theta \in \Theta$, the MLE $\widehat{\theta}_T^{\mathrm{ML}}$ is the limit of a Bayes estimator as the smoothing parameter $\gamma \to \infty$, i.e., using obvious notation:

$$\widehat{\theta}_T^{\mathrm{ML}} = \arg\max_{\theta\in\Theta} L_T(\theta) = \lim_{\gamma\to\infty} \frac{\int_\Theta \theta \exp(\gamma L_T(\theta)) \pi(\theta) \, d\theta}{\int_\Theta \exp(\gamma L_T(\theta)) \pi(\theta) \, d\theta}.$$

3.4.3 The Asymptotic Distribution for General Loss Functions

For general loss functions satisfying Assumption 3.1, Theorem 3.2 shows that $T\left\|\delta_{T}\right\|^{2}\left(\widehat{\lambda}_{b}^{\mathrm{GL}}-\lambda_{b}^{0}\right)$ is (first-order) asymptotically equivalent to ξ_{l}^{0} defined by

$$\Psi_{l}\left(\xi_{l}^{0}\right) \triangleq \inf_{r} \Psi_{l}\left(r\right)=\inf_{r\in\mathbb{R}}\left\{\int_{\mathbb{R}} l\left(r-u\right) p_{0}^{*}\left(u\right) du\right\}. \tag{3.13}$$

Theorem 3.2. *Under Assumption 2.1-2.4 and 3.1-3.7, and Condition 1, for $l \in \boldsymbol{L}$, $T\left\|\delta_{T}\right\|^{2}(\widehat{\lambda}_{b}^{\mathrm{GL}}-\lambda_{b}^{0}) \Rightarrow \xi_{l}^{0}$ as defined by (3.13).*

The existence and uniqueness of ξ_{l}^{0} follow from Assumption 3.7. If one interprets $p_{0}^{*}\left(u\right)$ as a true posterior density function, then ξ_{l}^{0} would naturally be viewed as a Bayesian estimator for the loss function $l_{T}\left(\cdot\right)$. In particular, in analogy to the above comparison with the Bayesian estimator of Ibragimov and Has'minskiĭ (1981), one can interpret the GL estimator as a Quasi-Bayesian estimator. While this is by itself a theoretically interesting result, we actually exploit it to construct more reliable inference methods about the date of a structural change. Under the least-absolute deviation loss, the GL estimator converges in distribution to the median of $p_{0}^{*}\left(u\right)$. We shall use the results in Theorem 3.1-3.2 but not Proposition 3.2 since the latter implies the same confidence intervals as in Bai (1997) and Bai and Perron (1998). GL inference based on the Bayes-type limiting distribution provides a more accurate description of the uncertainty over the parameter space than the inference based on the density of $\arg\max_{s\in\mathbb{R}} \mathscr{V}\left(s\right)$ which underestimates uncertainty as shown by confidence interval with empirical coverage rates below the nominal level particularly when the magnitude of the break is small (see Section 7). After some investigation, we found that both estimation and inference under the least-absolute loss works well and this is what will be used in our simulation study.

4 Confidence Sets Based on the GL Estimator

In this section, we discuss inference procedures for the break date based on the large-sample results of the previous section. Inference under general loss functions based on Theorem 3.2 is what we recommend to use in practice, in particular with an absolute loss function.

Since the limiting distribution from Theorem 3.2 involves certain unknown quantities, we begin by assuming that they can be replaced by consistent estimates. They are easy to construct [cf. Bai (1997) and Bai and Perron (1998); see also Section 7].

Assumption 4.1. *There exist sequences of estimators $\widehat{\lambda}_{b,T}$, $\widehat{\delta}_{T}$, $\widehat{\Xi}_{Z,T}$, and $\widehat{\Xi}_{e,T}$ such that $\widehat{\lambda}_{b,T}=\lambda_{0}+o_{\mathbb{P}}\left(1\right)$, $\widehat{\delta}_{T}=\delta_{T}+o_{\mathbb{P}}\left(1\right)$, $\widehat{\Xi}_{Z,T}=\Xi_{Z}+o_{\mathbb{P}}\left(1\right)$ and $\widehat{\Xi}_{e,T}=\Xi_{e}+o_{\mathbb{P}}\left(1\right)$. Furthermore, for all $u,\, s\in\mathbb{R}$ and any $c>0$, there exist covariation processes $\widehat{\Sigma}_{i,T}\left(\cdot\right)$ $(i=1,\,2)$ that satisfy (i)*

$\widehat{\Sigma}_{i,T}(u,\,s)=\Sigma_i^0(u,\,s)+o_{\mathbb{P}}(1)$, *(ii)* $\widehat{\Sigma}_{i,T}(u-s,\,u-s)=\widehat{\Sigma}_{i,T}(u,\,u)+\widehat{\Sigma}_{i,T}(s,\,s)-2\widehat{\Sigma}_{i,T}(s,\,u)$, *(iii)* $\widehat{\Sigma}_{i,T}(cu,\,cu)=c\widehat{\Sigma}_{i,T}(u,\,u)$, *(iv)* $\mathbb{E}\left\{\sup_{\|u\|=1}\widehat{\Sigma}_{i,T}^2(u,\,u)\right\}=O(1)$.

The first part and (i) of the second part follow from consistency of $\widehat{\lambda}_{b,T}$ and from an Invariance Principle [cf. Assumption 3.5]. Part (ii)-(iii) are implied by Assumption 3.6-(ii) and consistency of $\widehat{\lambda}_{b,T}$. Let $\left\{\widehat{\mathscr{W}_T}\right\}$ be a (sample-size dependent) sequence of two-sided zero-mean Gaussian processes with covariance $\widehat{\Sigma}_T$. Construct the process $\widehat{\mathscr{V}_T}$ by replacing the population quantities in $\mathscr{V}$ by their corresponding estimates from the first part of Assumption 4.1 and further, replace $\mathscr{W}$ by $\widehat{\mathscr{W}_T}$. Assumption 4.1-(i) basically implies that the finite-dimensional limit law of $\left\{\widehat{\mathscr{W}_T}\right\}$ is the same as the finite-dimensional laws of $\mathscr{W}$ while parts (ii)-(iii) are needed for the integrability of the transform $\exp\left(\widehat{\mathscr{V}_T}(\cdot)\right)$. Part (iv) is needed for the proof of the asymptotic stochastic equicontinuity of $\left\{\widehat{\mathscr{W}_T}\right\}$. Let $\widehat{\xi}_T$ be defined as the sample analogue of ξ_l^0 that uses $\widehat{\mathscr{V}_T}(v)$ in place of $\mathscr{V}_T(v)$ in (3.13). The distribution of $\widehat{\xi}_T$ can be evaluated numerically.

Proposition 4.1. *Let $l\in\boldsymbol{L}$ be continuous. Under Assumption 4.1, $\widehat{\xi}_T$ converges in distribution to the limiting distribution in Theorem 3.2.*

The asymptotic distribution theory of the GL estimator may be exploited in several ways to conduct inference about the break date. As emphasized by Casini and Perron (2019a), the finite-sample distribution of the break date least-squares estimator displays significant non-standard features (cf. Figure 1). Hence, a conventional two-sided confidence interval may not result in a confidence set with reliable properties across all break magnitudes and break locations. Thus, as in Casini and Perron (2019a), we use the concept of Highest Quasi-posterior Density (HQPD) regions, defined analogously to the Highest Density Region (HDR); cf. Hyndman (1996). See also Samworth and Wand (2010) and Mason and Polonik (2008, 2009) for more recent developments.

Definition 4.1. Highest Density Region: Let the density function $f_Y(y)$ of a random variable Y defined on a probability space $(\Omega_Y,\,\mathscr{F}_Y,\,\mathbb{P}_Y)$ and taking values on the measurable space $(\mathcal{Y},\,\mathscr{Y})$ be continuous and bounded. The $(1-\alpha)\,100\%$ Highest Density Region is a subset $\mathbf{S}(\kappa_\alpha)$ of $\mathcal{Y}$ defined as $\mathbf{S}(\kappa_\alpha)=\{y:\,f_Y(y)\geq\kappa_\alpha\}$ where κ_α is the largest constant that satisfies $\mathbb{P}_Y(Y\in\mathbf{S}(\kappa_\alpha))\geq 1-\alpha$.

For $s=T\,\|\delta_T\|^2\left(\widehat{\lambda}_b^{\mathrm{LS}}-\lambda_b^0\right)$, the asymptotic distribution theory of Bai (1997) suggests a belief $\pi(s)$ over $s\in\mathbb{R}$. This belief function can be used as a Quasi-prior for λ_b in the definition of the Quasi-posterior $p_T(\lambda_b)$. Let $\mu(\lambda_b)$ denote some density function defined by the Radon-Nikodym equation $\mu(\lambda_b)=dp_T(\lambda_b)/d\lambda_{\mathrm{L}}$, where λ_{L} denotes the Lebesgue measure. The following algorithm describes how to construct a confidence set for T_b^0.

Algorithm 1. *GL HQDR-based Confidence Sets for T_b^0:*
(1) Estimate by least-squares the break date and the regression coefficients from model (2.3);

(2) Set the Quasi-prior $\pi(\lambda_b)$ equal to the probability density of the limiting distribution from Corollary 3.1;
(3) Construct the Quasi-posterior given in (3.1);
(4) Obtain numerically the density $\mu(\lambda_b)$ as explained above and label it by $\widehat{\mu}(\lambda_b)$;
(5) Compute the Highest Quasi-Posterior Density (HQPD) region of the probability distribution $\widehat{p}_T(\lambda_b)$ and include the point T_b in the level $(1-\alpha)\%$ confidence set $C_{\mathrm{HQPD}}(\mathrm{cv}_\alpha)$ if T_b satisfies Definition 4.1.

If a general Quasi-prior $\pi(\lambda_b)$ is used, one begins directly with step 3.

In principle, any Quasi-prior $\pi(\lambda_b)$ satisfying Assumption 3.2 can be used. Note that $C_{\mathrm{HQPD}}(\mathrm{cv}_\alpha)$ retains a frequentist interpretation, since no parametric likelihood function of the data is required.

5 Models with Multiple Change-Points

Following Bai and Perron (1998), the multiple linear regression model with m change-points is

$$y_t = w_t'\phi^0 + z_t'\delta_j^0 + e_t, \qquad \left(t = T_{j-1}^0 + 1, \ldots, T_j^0\right)$$

for $j = 1, \ldots, m+1$, where by convention $T_0^0 = 0$ and $T_{m+1}^0 = T$. There are m unknown break points $(T_1^0, \ldots, T_m^0)$ and consequently $m+1$ regimes each corresponding to a distinct parameter value δ_j^0. The purpose is to estimate the unknown regression coefficients together with the break points when T observations on (y_t, w_t, z_t) are available. Many of the theoretical results follow directly from the single break case; the break points are asymptotically distinct and thus, given the mixing conditions, our results for the single break date extend readily to multiple breaks. More complicated is the computation of the estimates of the break dates which has been addressed by Bai and Perron (2003) who proposed an efficient algorithm based on the principle of dynamic programming; see also Hawkins (1976).

Let $T_i \triangleq \lfloor T\lambda_i \rfloor$ and $\theta \triangleq (\phi', \delta_1', \Delta_1' \ldots, \Delta_m')'$ where $\Delta_i = \delta_{i+1} - \delta_i$, $i = 1, \ldots, m$. The class $\mathscr{L}(\theta, T_i;\, 1 \leq i \leq m)$ of GL estimators in multiple change-points models relies on the least-squares criterion function $Q_T(\delta(\boldsymbol{\lambda}_b), \boldsymbol{\lambda}_b) = \sum_{i=1}^{m+1} \sum_{t=T_i-1}^{T_i} (y_t - w_t'\phi - z_t'\delta_i)^2$, with $\boldsymbol{\lambda}_b \triangleq (\lambda_i;\, 1 \leq i \leq m)$. In order to state the large-sample properties, we need to introduce the shrinkage theoretical framework of Bai and Perron (1998).

Assumption 5.1. *(i) Let $x_t = (w_t', z_t')'$, $X = (x_1, \ldots x_T)'$ and $\overline{X}_0 = diag\left(X_1^0, \ldots, X_{m+1}^0\right)$ be the diagonal partition of X at $(T_1^0, \ldots, T_m^0)$. For each $i = 1, \ldots, m+1$ $(X_1^0)'X_1^0/\left(T_i^0 - T_{i-1}^0\right)$ converges to a non-random positive definite matrix not necessarily the same for all i. (ii) Assumption 2.3 holds. (iii) The matrix $\sum_{t=k}^{l} z_t z_t'$ is invertible for $l - k \geq q$. (iv) $T_i^0 = \lfloor T\lambda_i^0 \rfloor$, where $0 < \lambda_1^0 <$*

$\cdots < \lambda_m^0 < 1$. *(v) Let* $\Delta_{T,i} = v_T \Delta_i^0$ *where* $v_T > 0$ *is a scalar satisfying* $v_T \to 0$ *and* $T^{1/2-\vartheta} v_T \to \infty$ *for some* $\vartheta \in (0,\, 1/4)$, *and* $\mathbb{E}\left\|z_t\right\|^2 < C$, $\mathbb{E}\left\|e_t\right\|^{2/\vartheta} < C$ *for some* $C < \infty$ *and all* t.

Assumption 5.2. *Let* $\Delta T_i^0 = T_i^0 - T_{i-1}^0$. *For* $i = 1, \ldots, m+1$, *uniformly in* $s \in [0,\, 1]$, *(a)* $(\Delta T_i^0)^{-1} \sum_{t=T_{i-1}^0+1}^{T_{i-1}^0+\lfloor s\Delta T_i^0 \rfloor} z_t z_t' \xrightarrow{\mathbb{P}} sV_i$, $(\Delta T_i^0)^{-1} \sum_{t=T_{i-1}^0+1}^{T_{i-1}^0+\lfloor s\Delta T_i^0 \rfloor} e_t^2 \xrightarrow{\mathbb{P}} s\sigma_i^2$, *and*

$$\left(\Delta T_i^0\right)^{-1} \sum_{t=T_{i-1}^0+1}^{T_{i-1}^0+\lfloor s\Delta T_i^0 \rfloor} \sum_{r=T_{i-1}^0+1}^{T_{i-1}^0+\lfloor s\Delta T_i^0 \rfloor} \mathbb{E}\left(z_t z_r' u_t u_r\right) \xrightarrow{\mathbb{P}} s\Sigma_i;$$

(b) $(\Delta T_i^0)^{-1/2} \sum_{t=T_{i-1}^0+1}^{T_{i-1}^0+\lfloor s\Delta T_i^0 \rfloor} z_t u_t \xrightarrow{\mathbb{P}} \mathscr{G}_i(s)$ *where* $\mathscr{G}_i(s)$ *is a multivariate Gaussian process on* $[0,\, 1]$ *with mean zero and covariance* $\mathbb{E}\left[\mathscr{G}_i(s)\,\mathscr{G}_i(u)\right] = \min\{s,\, u\}\,\Sigma_i$.

Next, for $i = 1, \ldots, m$, define $\Xi_{Z,i} = (\Delta_i^0)' V_{i+1} \Delta_i^0 / (\Delta_i^0)' V_i \Delta_i^0$, $\Xi_{e,i}^2 = (\Delta_i^0)' \Sigma_{i+1} \Delta_i^0 / (\Delta_i^0)' \Sigma_i \Delta_i^0$, and let $W_1^{(i)}(s)$ and $W_2^{(i)}(s)$ be independent Wiener processes defined on $[0,\, \infty)$, starting at 0 when $s = 0$; $W_1^{(i)}(s)$ and $W_2^{(i)}(s)$ are also independent over i. Finally, define

$$\mathscr{V}^{(i)}(s) \triangleq \mathscr{W}^{(i)}(s) - \Lambda_i^0(s) \triangleq \begin{cases} 2\left((\Delta_i^0)' \Sigma_i \Delta_i\right)^{1/2} W_1^{(i)}(-s) - |s|\,(\Delta_i^0)' V_i \Delta_i, & \text{if } s \leq 0 \\ 2\left((\Delta_i^0)' \Sigma_{i+1} \delta^0\right)^{1/2} W_2^{(i)}(s) - s\,(\Delta_i^0)' V_{i+1} \Delta_i, & \text{if } s > 0. \end{cases} \tag{5.1}$$

We now extend the notation of Section 3 to the present context. By redefining the Quasi-posterior $p(\boldsymbol{\lambda}_b)$ in terms of $\boldsymbol{\lambda}_b$, the GL estimator as the minimizer of the associated risk function [recall (3.2)], $\widehat{\boldsymbol{\lambda}}_b^{\mathrm{GL}} = \arg\min_{s \in \Gamma^0} \left[\mathcal{R}_{l,T}(s)\right]$, where now $\Gamma^0 = \mathbf{B}_1 \times \ldots \times \mathbf{B}_m$, with $\mathbf{B}_i$ a compact subset of $(0,\, 1)$. The sets $\mathbf{B}_i$ are disjoint and satisfy $\sup_{\lambda \in \mathbf{B}_i} < \inf_{\lambda \in \mathbf{B}_{i+1}}$ for all i.

Assumption 5.3. *Assumption 3.1-3.2 hold with obvious modifications to allow for the multidimensional parameter* $\boldsymbol{\lambda}_b \in \Gamma^0$. *Furthermore, Assumption 3.6 holds where now in part (i)* $\boldsymbol{\lambda}_b$ *replaces* λ_b, *and in part (ii)* $\Sigma^{(i)}(\cdot,\, \cdot)$ *(*$1 \leq i \leq m+1$*) replaces* $\Sigma(\cdot,\, \cdot)$ *and is defined analogously for each regime.*

Assumption 3.7 implies that $\xi_{l,i}^0 \triangleq \xi(\lambda_i^0)$ is uniquely defined by $\Psi_l\left(\xi_{l,i}^0\right) \triangleq \inf_s \Psi_{l,i}(s) = \inf_s \int_{\mathbb{R}} l(s-u) \left(\exp\left(\mathscr{V}^{(i)}(u)\right) / \left(\int_{\mathbb{R}} \exp\left(\mathscr{V}^{(i)}(w)\right) dw\right)\right) du$. The GL estimator is defined as the minimizer of

$$\mathcal{R}_{l,T} \triangleq \int_{\Gamma^0} l(s - \boldsymbol{\lambda}_b) \frac{\exp\left(-Q_T\left(\delta(\boldsymbol{\lambda}_b),\, \boldsymbol{\lambda}_b\right)\right) \pi(\boldsymbol{\lambda}_b)}{\int_{\Gamma^0} \exp\left(-Q_T\left(\delta(\boldsymbol{\lambda}_b),\, \boldsymbol{\lambda}_b\right)\right) \pi(\boldsymbol{\lambda}_b)\, d\boldsymbol{\lambda}_b} d\boldsymbol{\lambda}_b.$$

The analysis is now in terms of the $m \times 1$ local parameter u with components $u_i = T\left\|\Delta_{T,i}\right\|^2 (\lambda_i - \lambda_{i,T}^0(v))$, with $\lambda_{i,T}^0(v) = \lambda_{i,T}^0(\theta^0 + v/r_T)$.

Theorem 5.1-5.2 extend corresponding results from Theorem 3.1-3.2, respectively, to multiple change-points. The fast rate of convergence implies that asymptotically the behavior of the GL estimator only matters in a small neighborhood of each T_i^0. Since each such neighborhood increases at rate $1/v_T$ while $T \to \infty$ at a faster rate, given the mixing conditions, these asymptotically distinct and the limiting distribution is then similar to that in the single break case. This is the same argument underlying the analysis of Bai and Perron (1998) and of Ibragimov and Has'minskiĭ (1981). The same comments as those in Section 3 apply.

Condition 3. For $1 \leq i \leq m$ there exist positive finite numbers $\kappa_{\gamma,i}$ such that $\gamma_T/T \left\|\Delta_{T,i}\right\|^2 \to \kappa_{\gamma,i}$.

Theorem 5.1. *Assume $l(\cdot)$ is the squared loss function. Under Assumption 5.1-5.3 and Condition 3, we have in $\mathbb{D}_b$,*

$$T \left\|\Delta_{T,i}\right\|^2 \left(\widehat{\lambda}_i^{\mathrm{GL}} - \lambda_i^0\right) \Rightarrow \frac{\int u \exp\left(\mathscr{W}^{(i)}(u) - \Lambda_i^0(u)\right) du}{\int \exp\left(\mathscr{W}^{(i)}(u) - \Lambda_i^0(u)\right) du}. \tag{5.2}$$

Turning to the general case of loss functions satisfying Assumption 3.1, Theorem 5.2 shows that the random quantity $T \left\|\delta_T\right\|^2 \left(\widehat{\lambda}_i^{\mathrm{GL}} - \lambda_i^0\right)$ is (first-order) asymptotically equivalent to the random variable $\xi_{l,i}^0$ determined by

$$\Psi_l\left(\xi_{l,i}^0\right) \triangleq \inf_r \Psi_{l,i}(r) = \inf_{r\in\mathbb{R}} \left\{ \int_{\mathbb{R}} l(r-u) \frac{\exp\left(\mathscr{W}^{(i)}(u) - \Lambda_i^0(u)\right)}{\int \exp\left(\mathscr{W}^{(i)}(u) - \Lambda_i^0(u)\right) du} du \right\}. \tag{5.3}$$

Theorem 5.2. *Under Assumptions 5.1-5.3 and Condition 3, for $l \in \boldsymbol{L}$, $T \left\|\Delta_{T,i}\right\|^2 \left(\widehat{\lambda}_i^{\mathrm{GL}} - \lambda_i^0\right) \Rightarrow \xi_{l,i}^0$, as defined by (5.3).*

A direct consequence of the results of this section is that statistical inference for the break dates T_i^0 $(i = 1, \ldots, m)$ can be carried out using the same methods for the single break case as described in Section 4.

6 Theoretical Properties of GL Inference

This section shows that the GL-HPDR confidence sets are bet-proof. The betting framework and the notion of bet-proofness are useful to study the properties of frequentist inference in non-regular problems. The literature concluded that frequentist confidence sets may exhibit undesirable properties in non-regular problems [e.g., Buehler (1959), Cornfield (1969), Cox (1958), Müller and Norest (2016) Pierce (1973), Robinson (1977) and Wallace (1959)]. For example, the confidence sets can be too short or empty with positive probability. This arises because frequentist procedures often have the property that, conditional on a sample point lying in some subset of the sample space, the conditional confidence level is less than the unconditional confidence level uniformly in the parameters.

We use the same betting framework as in Buehler (1959). Let $\mathbb{P}(\cdot|\,\lambda_b)$ denote the likelihood of the data $Y \in \mathcal{Y}$ conditional on $\lambda_b \in \Gamma^0$. Assume $\mathbb{P}(\cdot|\,\lambda_b)$ has density $p(\cdot|\,\lambda_b)$ with respect to a finite measure ζ. We define a $1-\alpha$ confidence set by a rejection probability rule $\varphi : \Gamma^0 \times \mathcal{Y} \mapsto [0,\,1]$ satisfying $\int [1-\varphi(\lambda_b,\,y)]\, p(y|\,\lambda_b)\, d\zeta(y) \geq 1-\alpha$, with $\varphi(\lambda_b,\,y)$ the probability that λ_b is not included in the set when y is observed. For any realization of the data $Y = y$, an inspector can choose to object to the confidence set φ. The inspector's objection $\widetilde{b} : \mathcal{Y} \mapsto [0,\,1]$ takes value 1 if there is an objection. Denote by $\mathbf{B}$ the set of all measurable strategies $\widetilde{b}$. When $\widetilde{b} = 1$ the inspector receives 1 if φ does not contain λ_b, and she loses $\alpha/(1-\alpha)$ otherwise. For a given parameter λ_b and betting strategy $\widetilde{b}$, the inspector's expected loss is,

$$L_\alpha\left(\varphi,\,\widetilde{b},\,\lambda_b\right) = \frac{1}{1-\alpha}\int [\alpha - \varphi(\lambda_b,\,y)]\,\widetilde{b}(y)\, p(y|\lambda_b)\, d\zeta(y)\,.$$

A confidence set φ is said to be bet-proof at level $1-\alpha$ if for each $\widetilde{b} \in \mathbf{B}$, $L_\alpha\left(\varphi,\,\widetilde{b},\,\lambda_b\right) \geq 0$ for some $\lambda_b \in \Gamma^0$. If there exists a strategy $\widetilde{b}$ such that $L_\alpha\left(\varphi,\,\widetilde{b},\,\lambda_b\right) < 0$ for all $\lambda_b \in \Gamma^0$, then the inspector would be right on average and would make positive expected profits. Hence, such φ as would be an "unreasonable" confidence set. Without loss of substance, we restrict our attention to a change in the mean of a sequence of i.i.d. Gaussian variables. Let $y_t = \delta_T \mathbf{1}\{t > T_b^0\} + e_t$, where $e_t \sim i.i.d.\, \mathscr{N}(0,\,1)$. The result below can also be shown to hold also for fixed shifts $\delta_T = \delta^0$. For ease of exposition, we assume δ^0 known. The general case leads to similar results, with more lengthy derivations without any gain in intuition.

Recall that φ is such that the Quasi-posterior probability $p_T(\lambda_b|\,y) = p_T(\lambda_b)$ of excluding λ_b is less than or equal to α,

$$\int \varphi(\lambda_b,\,y)\, p_T(\lambda_b|\,y)\, d\lambda_b \leq \alpha \quad \text{for all}\, y \in \mathcal{Y}. \tag{6.1}$$

Proposition 6.1. *Assume Assumption 2.1-2.4 and 3.1-3.7, and Condition 1 hold. For* $l \in \boldsymbol{L}$ *: (i)* φ *is bet-proof at level* $1-\alpha$*; (ii) If (6.1) holds with equality, then* φ *is the shortest confidence set in the class of level* $1-\alpha$ *confidence sets, i.e., there cannot exist a level* $1-\alpha$ *confidence set* φ' *with the property that, for all* $y \in \mathcal{Y}$ $\int \varphi'(\lambda_b,\,y)\, d\lambda_b \geq \int \varphi(\lambda_b,\,y)\, d\lambda_b$*, and for all* $y \in \mathcal{Y}_0$ *with* $\zeta(\mathcal{Y}_0) > 0$, $\int \varphi'(\lambda_b,\,y)\, d\lambda_b > \int \varphi(\lambda_b,\,y)\, d\lambda_b$.

Part of the proof shows that the Quasi-posterior is asymptotically equivalent (in total variation distance) to the Bayesian posterior. Given the conservativeness allowed by Definition 4.1, the GL confidence interval is asymptotically a superset of a Bayesian credible interval. Bet-proofness is a useful criterion in change-point models where popular inference methods face some difficulties, as shown in the next section. Proposition 6.1 suggests that GL inference should not suffer from these issues; the simulations in the next section will confirm that this is indeed the case.

7 Finite-Sample Evaluations

The purpose of this section is twofold. Section 7.1 assesses the accuracy of the GL estimate of the change-point while Section 7.2 evaluates the small-sample properties of the proposed method to construct confidence sets. We consider DGPs that take the form:

$$y_t = D_t\alpha^0 + Z_t\beta^0 + Z_t\delta^0\mathbf{1}\left\{t > T_b^0\right\} + e_t, \qquad t = 1, \ldots, T, \tag{7.1}$$

with a sample size $T = 100$. Three versions of (7.1) are investigated: M1 involves a break in mean: $Z_t = 1$, D_t absent, and $e_t \sim i.i.d.\,\mathscr{N}(0,\,1)$; M2 is similar to M1 but with $e_t = 0.3e_{t-1} + u_t$, $u_t \sim \mathscr{N}(0,\,1)$; M3 is a dynamic model with $D_t = y_{t-1}$, $Z_t = 1$, $e_t \sim i.i.d.\,\mathscr{N}(0,\,0.5)$ and $\alpha^0 = 0.6$. We set $\beta^0 = 1$ in M1-M2 and $\beta^0 = 0$ in M3. We consider fractional change-points $\lambda_0 = 0.3$ and 0.5, and break magnitudes $\delta^0 = 0.3$, 0.4, 0.6 and 1.

7.1 Precision of the Change-point Estimate

We consider the following estimators of T_b^0: the least-squares estimator (OLS), the GL estimator under a least-absolute loss function (GL-LN); the GL estimator under a least-absolute loss function with a uniform prior (GL-Uni). We compare the mean absolute error (MAE), standard deviation (Std), root-mean-squared error (RMSE), and the 25% and 75% quantiles.

Table 1-3 present the results. When the magnitude of the break is small, the OLS estimator displays quite large MAE, which increases as the change-point point moves toward the tails. In contrast, the GL estimator shows substantially lower MAE uniformly over break magnitudes and break locations. In addition, the GL estimator has smaller variance as well as lower RMSE compared to the OLS estimator. Notably, the distribution of GL-LN concentrates a higher fraction of the mass around the mid-sample relative to the finite-sample distribution of the OLS estimate. This is mainly due to the fact that the Quasi-posterior essentially does not share the marked trimodality of the finite-sample distribution [cf. Casini and Perron (2019a)]. When the break magnitude is small, the objective function is quite flat with a small peak at the OLS estimate. The Quasi-posterior has higher mass close to the OLS estimate—which corresponds to the middle mode—and accordingly lower mass in the tails. The GL estimator that uses the uniform prior (GL-Uni) is also more precise than the OLS estimator, though the margin is smaller. The latter is due to the fact that the GL estimate uses information only from the OLS objective function.

7.2 Properties of the GL Confidence Sets

We now assess the performance of the suggested inference procedure for the break date. We compare it with the following existing methods: Bai's (1997) approach, Elliott and Müller's (2007) approach based on inverting a sequence of locally best invariant tests using Nyblom's (1989) statistic, the inverted likelihood-ratio (ILR) method of Eo and Morley (2015) which inverts the likelihood-ratio test of Qu and Perron (2007) and the HDR method proposed in Casini and Perron (2019a) based on continuous record asymptotics, labelled OLS-CR. These methods have been discussed in detail in Casini and Perron (2019a) and in Chang and Perron (2018). We can summarize their properties as follows. The confidence intervals obtained from Bai's (1997) method display empirical coverage rates often below the nominal level when the size of the break is small. In general, Elliott and Müller's (2007) approach achieves the most accurate coverage rates but the average length of the confidence sets is always substantially larger relative to other methods.[1] In addition, this approach breakdowns in models with serially correlated errors or lagged dependent variables, whereby the length of the confidence set approaches the whole sample as the magnitude of the break increases. The ILR has coverage rates often above the nominal level and an average length significantly longer than with the OLS-CR method when the magnitude of the shift is small. Here, we shall show that the GL inference performs well in terms of coverage probability compared with the other methods and is characterized by shorter lengths of the confidence sets.

When the errors are uncorrelated (i.e., M1 and M3) we simply estimate variances rather than long-run variances. The least-squares estimation method is employed with a trimming parameter $\epsilon = 0.15$ and we use the required degrees of freedom adjustment for the statistic $\widehat{\mathrm{U}}_T$ of Elliott and Müller (2007). To construct the OLS-CR method, we follow the steps outlined in Casini and Perron (2019a). To implement Bai's (1997) method we use the usual steps described in Bai (1997) and Bai and Perron (1998). We implement the GL estimator using a least-absolute loss with the prior from Theorem 3.2. For model M1, the estimate of the long-run variance is the pre-whitened heteroskedasticity and autocorrelation (HAC) estimator of Andrews and Monahan (1992). We consider the version $\widehat{\mathrm{U}}_T$ proposed by Elliott and Müller (2007) that allows for heterogeneity across regimes; using the restricted version when applicable leads to similar results. Finally, the last row of each panel includes the rejection probability of the 5%-level sup-Wald test using the asymptotic critical value of Andrews (1993); it serves as a statistical measure of the magnitude of the break.

Overall, the results in Table 4-6 confirm previous findings about the performance of existing methods. Bai's (1997) method has a coverage rate below the nominal level when the size of the break is small. For example, in model M2, with $\lambda_0 = 0.5$ and $\delta^0 = 0.8$, it has a coverage probability below 82% even though the Sup-Wald test rejects roughly for 70% of the samples. With smaller break sizes, it systematically fails to cover the true break date with correct probability. In contrast, the method of Elliott and Müller (2007) yields very accurate empirical coverage rates. However, the

[1]This problem is more severe when the errors are serially correlated or the model includes lagged dependent variables. Regarding the former, this in part may be due to issues with Newey and West HAC-type estimators when there are structural breaks [see Casini (2018, 2019), casini/perron:OUP-Breaks, Chang and Perron (2018), Crainiceanu and Vogelsang (2007), Deng and Perron (2006), Fossati (2018), Juhl and Xiao (2009), Kim and Perron (2009), Martins and Perron (2016), Perron and Yamamoto (2019) and Vogeslang (1999)].

average length of the confidence intervals obtained is systematically much larger than those from all other methods across all DGPs, break sizes and break locations. For large break sizes, Bai's (1997) method delivers good coverage rates and the shortest average length among all methods.

The GL method displays good coverage rates across different break magnitudes and tends to have the shortest lengths among all methods for all break magnitudes, except for $\delta^0 = 1.6$ in model M2 for which Bai's (1997) confidence interval is slightly shorter. In Model M3, the coverage rates of OLS-CR are more accurate than those with the GL method although the difference is not large. Thus the GL method strikes a good balance between adequate coverage probability and short average lengths, thus confirming the theoretical results on bet-proofness.

8 Conclusions

We developed large-sample results for a class of Generalized Laplace estimators in multiple change-points models where popular methods face some challenges due to the non-regularities of the problem. The GL method exploits the insight of Laplace who proposed to generate a density from taking an exponential transformation of a least-squares criterion. The class of GL estimators exhibits a dual limiting distribution; namely, the classical shrinkage asymptotic distribution of Bai and Perron (1998), or a Bayes-type asymptotic distribution [cf. Ibragimov and Has'minskiĭ (1981)]. Simulations show that the GL estimator is more accurate than OLS. Similarly, inference has superior finite-sample properties relative to popular methods and these properties are shown to be supported by theoretical results.

References

ANDREWS, D. W. K. (1993): "Tests for Parameter Instability and Structural Change with Unknown Change-Point," *Econometrica*, 61(4), 821–56.

ANDREWS, D. W. K., AND J. C. MONAHAN (1992): "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," *Econometrica*, 60(4), 953–966.

ANDREWS, D. W. K., AND W. PLOBERGER (1994): "Optimal Tests when a Nuisance Parameter is Present Only Under the Alternative," *Econometrica*, 62(6), 1383–1414.

BAI, J. (1994): "Least Squares Estimation of a Shift in Linear Processes," *Journal of Time Series*, 15(5), 453–472.

——— (1995): "Least Absolute Deviation Estimation of a Shift," *Econometric Theory*, 11(3), 403–436.

——— (1997): "Estimation of a Change-Point in Multiple Regression Models," *The Review of Economics and Statistics*, 79(4), 551–563.

BAI, J., AND P. PERRON (1998): "Estimating and Testing Linear Models with Multiple Structural Changes," *Econometrica*, 66(1), 47–78.

——— (2003): "Computation and Analysis of Multiple Structural Changes," *Journal of Applied Econometrics*, 18, 1–22.

BHATTACHARYA, P. K. (1987): "Maximum Likelihood Estimation of a Change-Point in the Distribution of Independent Random Variables: General Multiparameter Case," *Journal of Multivariate Analysis*, 23(2), 183–208.

——— (1994): *Some Aspects of Change-Point Analysis*vol. 23 of *Lecture Notes–Monograph Series*, pp. 28–56. Institute of Mathematical Statistics, Hayward, CA.

BICKEL, P. J., AND J. A. YAHAV (1969): "Some Contributions to the Asymptotic Theory of Bayes Solutions," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 11(4), 257–276.

BILLINGSLEY, P. (1995): *Probability and Measure*. New York: John Wiley and Sons.

——— (1999): *Convergence of Probability Measures*, Wiley Series in Probability and Statistics. John Wiley and Sons, second edn.

BUEHLER, R. J. (1959): "Some Validity Criteria for Statistical Inferences," *Annals of Mathematical Statistics*, 30(4), 845–863.

CASINI, A. (2018): "Tests for Forecast Instability and Forecast Failure under a Continuous Record Asymptotic Framework," *arXiv preprint arXiv:1803.10883*.

——— (2019): "Theory of Evolutionary Spectra for Heteroskedasticity and Autocorrelation Robust Inference in Possibly Misspecified and Nonstationary Models," *Unpublished manuscript, Department of Economics and Finance, University of Rome Tor Vergata*.

CASINI, A., AND P. PERRON (2019a): "Continuous Record Asymptotics for Structural Change

Models," *arXiv preprint arXiv:1803.10881*.

——— (2019b): "Continuous Record Laplace-based Inference in Structural Change Models," *arXiv preprint arXiv:1804.00232*.

——— (2019c): "Structural Breaks in Time Series," *Oxford Research Encyclopedia of Economics and Finance*, Oxford University Press.

——— (2020): "Supplement to Generalized Laplace Inference in Multiple Change-Points Models," .

Chang, S. Y., and P. Perron (2018): "A Comparison of Alternative Methods to Construct Confidence Intervals for the Estimate of a Break Date in Linear Regression Models," *Econometric Reviews*, 37(6), 577–601.

Chernozhukov, V., and H. Hong (2003): "An MCMC Approach to Classical Estimation," *Journal of Econometrics*, 115(2), 293–346.

——— (2004): "Likelihood Estimation and Inference in a Class of Nonregular Econometric Models," *Econometrica*, 72(5), 1445–1480.

Cornfield, J. (1969): "The Bayesian Outlook and its Application," *Biometrics*, 25(4), 617–657.

Cox, D. R. (1958): "Some Problems Connected with Statistical Inference," *Annals of Mathematical Statistics*, 29(2), 357–372.

Crainiceanu, C. M., and T. J. Vogelsang (2007): "Nonmonotonic Power for Tests of a Mean Shift in a Time Series," *Journal of Statistical Computation and Simulation*, 77(6), 457–476.

Csörgő, M., and L. Horváth (1997): *Limit Theorems in Change-Point Analysis*. New York: John Wiley and Sons.

Deng, A., and P. Perron (2006): "A Comparison of Alternative Asymptotic Frameworks to Analyse a Structural Change in a Linear Time Trend," *Econometrics Journal*, 9(3), 423–447.

Elliott, G., and U. K. Müller (2007): "Confidence Sets for the Date of a Single Break in Linear Time Series Regressions," *Journal of Econometrics*, 141(2), 1196–1218.

Eo, Y., and J. Morley (2015): "Likelihood-Ratio-Based Confidence Sets for the Timing of Structural Breaks," *Quantitative Economics*, 6(2), 463–497.

Forneron, J. J., and S. Ng (2018): "The ABC of Simulation Estimation with Auxiliary Statistics," *Journal of Econometrics*, 205(1), 112–139.

Fossati, S. (2018): "Testing for State-Dependent Predictive Ability," *Unpublished manuscript, Department of Economics, University of Alberta*.

Ghosal, S., J. K. Ghosh, and T. Samanta (1995): "On Convergence of Posterior Distributions," *Annals of Statistics*, 23(6), 2145–2152.

Hall, A., S. Han, and O. Boldea (2010): "Inference Regarding Multiple Structural Changes in Linear Models with Endogenous Regressors," *Journal of Econometrics*, 170(2), 281–302.

Hawkins, D. (1976): "Point Estimation of the Parameters of Piecewise Regression Models," *Journal of Applied Statistics*, 25(1), 51–57.

Hawkins, D. M. (1977): "Testing a Sequence of Observations for a Shift in Location," *Journal of the American Statistical Association*, 72(357), 180–186.

Hinkley, D. V. (1971): "Inference About the Change-Point from Cumulative Sum Tests," *Biometrika*, 58(3), 509–523.

Hirano, K., and J. R. Porter (2003): "Asymptotic Efficiency in Parametric Structural Models with Parameter-Dependent Support," *Econometrica*, 71(5), 1307–1338.

Horváth, L. (1993): "The Maximum Likelihood Method for Testing Changes in the Parameters of Normal Observations," *The Annals of Statistics*, 21(2), 671–680.

Hyndman, R. J. (1996): "Computing and Graphing Highest Density Regions," *The American Statistician*, 50(2), 120–126.

Ibragimov, A., and R. Z. Has'minskiĭ (1981): *Statistical Estimation: Asymptotic Theory.* Springer-Verlag New York.

Juhl, T., and Z. Xiao (2009): "Testing for Changing Mean with Monotonic Power," *Journal of Econometrics*, 148(1), 14–24.

Jun, S. J., J. Pinkse, and Y. Wan (2015): "Classical Laplace Estimation for $\sqrt[3]{n}$-consistent Estimators: Improved Convergence Rates and Rate-adaptive Inference," *Journal of Econometrics*, 187(1), 201–216.

Jurečová, J. (1977): "Asymptotic Relations of M-Estimates and R-Estimates in Linear Regression Model," *Annals of Statistics*, 5(3), 464–472.

Kim, D., and P. Perron (2009): "Assessing the Relative Power of Structural Break Tests Using a Framework Based on the Approximate Bahadur Slope," *Journal of Econometrics*, 149(1), 26–51.

Kim, H. J., and D. Pollard (1990): "Cube Root Asymptotics," *Annals of Statistics*, 18(1), 191–219.

Kim, H. J., and D. Siegmund (1989): "The Likelihood Ratio Test for a Change Point in Simple Linear Regression," *Biometrika*, 76(3), 409–423.

Laplace, P. S. (1774): "Memoir on the Probability of Causes of Events," *Mémoires de Mathématique et de Physique*, Tome Sixième, (English translation by S. M. Stigler 1986. Statist. Sci., 1(19):364-378).

Martins, L., and P. Perron (2016): "Improved Tests for Forecast Comparisons in the Presence of Instabilities," *Journal of Time Series Analysis*, 37(5), 650–659.

Mason, D. M., and W. Polonik (2008): "Asymptotic Normality of Plug-in Level Set Estimates," Extended version.

——— (2009): "Asymptotic Normality of Plug-in Level Set Estimates," *Annals of Applied Probability*, 19(3), 1108–1142.

Müller, U. K., and A. Norest (2016): "Credibility of Confidence Sets in Nonstandard Econometric Problems," *Econometrica*, 84(6), 2183–2213.

Newey, W. K., and K. D. West (1987): "A Simple Positive Semidefinite, Heteroskedastic and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

Nyblom, J. (1989): "Testing for the Constancy of Parameters over Time," *Journal of the American Statistical Association*, 89(451), 223–230.

Oka, T., and Z. Qu (2010): "Estimating Structural Changes in Regression Quantiles," *Journal of Econometrics*, 162(2), 248–267.

Perron, P. (2006): "Dealing with Structural Breaks," in *Palgrave Handbook of Econometrics*, ed. by K. Patterson, and T. Mills, vol. 1: Econometric Theory, pp. 278–352. Palgrave Macmillan.

Perron, P., and Z. Qu (2006): "Estimating Restricted Structural Change Model," *Journal of Econometrics*, 134(2), 373–399.

Perron, P., and Y. Yamamoto (2014): "A Note on Estimating and Testing for Multiple Structural Changes in Models with Endogenous Regressors via 2SLS," *Econometric Theory*, 30(2), 491–507.

——— (2015): "Using OLS to Estimate and Test for Structural Changes in Models with Endogenous Regressors," *Journal of Applied Econometrics*, 30(1), 119–144.

Perron, P., and Y. Yamamoto (2019): "Testing for Changes in Forecast Performance," *Journal of Business and Economic Statistics*, forthcoming.

Picard, D. (1985): "Testing and Estimating Change-points in Time Series," *Advances in Applied Probability*, 17(4), 841–867.

Pierce, D. A. (1973): "On Some Difficulties in a Frequency Theory of Inference," *Annals of Statistics*, 1(2), 241–250.

Qu, Z., and P. Perron (2007): "Estimating and Testing Structural Changes in Multivariate Regressions," *Econometrica*, 75(2), 459–502.

Robert, C., and G. Casella (2004): *Monte Carlo Statistical Methods*, Springer Texts in Statistics. Springer-Verlag New York, second edn.

Robinson, G. K. (1977): "Conservative Statistical Inference," *Journal of the Royal Statistical Society. Series B*, 39(3), 381–386.

Royden, H. L., and P. Fitzpatrick (2010): *Real Analysis*. Prentice Hall.

Samworth, R. J., and M. P. Wand (2010): "Asymptotics and Optimal Bandwidth Selection for Highest Density Region Estimation," *The Annals of Statistics*, 38(3), 1767–1792.

van der Vaart, A., and J. Wellner (1996): *Weak Convergence and Empirical Processes*. Springer-Verlag.

Vogeslang, T. (1999): "Sources of Nonmonotonic Power when Testing for a Shift in Mean of a Dynamic Time Series," *Journal of Econometrics*, 88(2), 283–299.

Wallace, D. L. (1959): "Conditional Confidence Level Properties," *Annals of Mathematical Statistics*, 30(4), 864–876.

Yao, Y. (1987): "Approximating the Distribution of the ML Estimate of the Change-Point in a

Sequence of Independent Random Variables," *Annals of Statistics*, 15, 1321–1328.

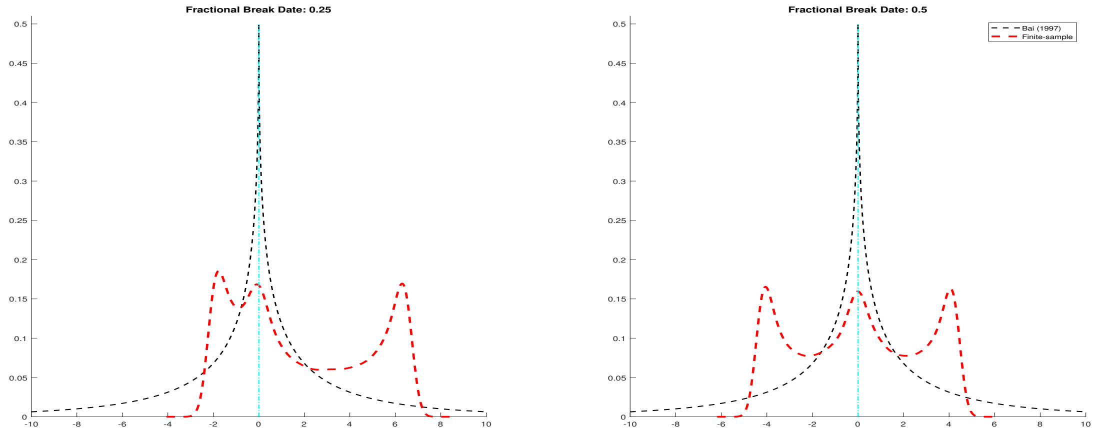

Figure 1: The probability density of the LS estimator for the model $y_t = \mu^0 + Z_t\delta_1^0 + Z_t\delta_2^0\mathbf{1}\{t > \lfloor T\lambda_0 \rfloor\} + e_t$, $Z_t = 0.3Z_{t-1} + u_t - 0.1u_{t-1}$, $u_t \sim$ i.i.d.$\mathscr{N}(0,\,1)$, $e_t \sim$ i.i.d.$\mathscr{N}(0,\,1)$, $\{u_t\}$ independent from $\{e_t\}$, $T = 100$ with $\delta^0 = 0.3$ and $\lambda_0 = 0.25$ and 0.5 (the left and right panel, respectively). The black broken line is the density of the asymptotic distribution from Bai (1997) and the red broken line break is the density of the finite-sample distribution.

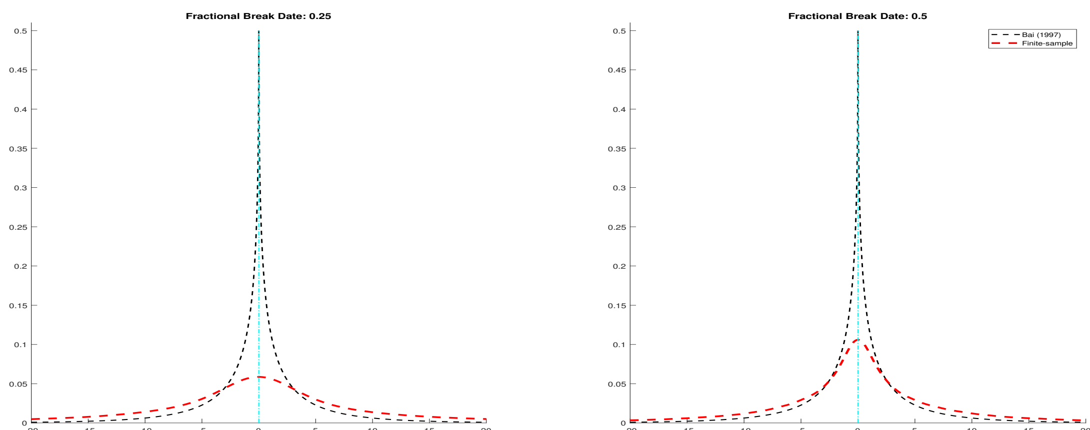

Figure 2: The comments in Figure 1 apply but with a break magnitude $\delta^0 = 1.5$.

Table 1: Small-sample accuracy of the estimate of the break point T_b^0 for model M1

		MAE	Std	RMSE	$Q_{0.25}$	$Q_{0.75}$	MAE	Std	RMSE	$Q_{0.25}$	$Q_{0.75}$
		$\lambda_0 = 0.3$					$\lambda_0 = 0.5$				
$\delta^0 = 0.3$	OLS	21.99	27.51	30.53	24	66	21.51	26.85	26.79	34	71
	GL-LN	13.44	15.03	18.99	28	54	11.85	14.51	14.93	38	60
	GL-Uni	17.56	22.88	25.51	26	56	16.90	22.03	22.13	38	61
$\delta^0 = 0.4$	OLS	20.48	26.30	28.51	23	57	15.64	21.79	21.23	40	61
	GL-LN	13.02	15.52	18.29	29	51	9.46	11.84	12.30	44	56
	GL-Uni	17.68	22.30	24.64	27	54	12.38	17.69	17.15	42	57
$\delta^0 = 0.6$	OLS	13.04	20.82	15.92	28	41	11.06	16.05	16.89	45	55
	GL-LN	9.20	13.67	13.67	28	40	7.04	9.92	10.46	47	53
	GL-Uni	11.49	18.59	14.23	27	39	9.11	13.92	13.48	45	55
$\delta^0 = 1$	OLS	3.49	4.61	4.61	28	32	2.92	5.24	5.23	48	52
	GL-LN	3.41	4.53	4.52	28	32	2.89	5.44	5.20	49	51
	GL-Uni	3.63	4.56	4.61	28	32	2.90	5.21	5.22	48	52

The model is $y_t = \delta^0 \mathbf{1}\{t > \lfloor T\lambda_0 \rfloor\} + e_t$, $e_t \sim i.i.d. \mathscr{N}(0, 1)$, $T = 100$. The columns refer to Mean Absolute Error (MAE), standard deviation (Std), Root Mean Squared Error (RMSE) and the 25% and 75% empirical quantiles. OLS is the least-squares estimator; GL-LN is the GL estimator under a least-absolute loss function with the density of the long-span asymptotic distribution as the prior; GL-Uni is the GL estimator under a least-absolute loss function with a uniform prior. The number of simulations is 3,000.

Table 2: Small-sample accuracy of the estimates of the break point T_b^0 for model M2

		MAE	Std	RMSE	$Q_{0.25}$	$Q_{0.75}$	MAE	Std	RMSE	$Q_{0.25}$	$Q_{0.75}$
		$\lambda_0 = 0.3$					$\lambda_0 = 0.5$				
$\delta^0 = 0.3$	OLS	26.61	22.85	33.03	23	76	24.09	28.29	28.08	23	73
	GL-LN	19.33	10.17	24.87	29	61	16.01	18.78	19.81	29	62
	GL-Uni	24.76	21.05	31.34	26	70	20.93	25.37	25.39	28	65
$\delta^0 = 0.4$	OLS	23.10	27.99	30.85	21	68	20.47	25.55	25.54	33	70
	GL-LN	16.59	18.59	22.75	29	60	13.68	17.06	17.12	38	61
	GL-Uni	21.51	25.87	28.83	24	61	17.91	22.94	22.91	37	62
$\delta^0 = 0.6$	OLS	17.64	23.51	25.01	24	50	15.51	20.93	20.91	41	59
	GL-LN	13.42	16.63	18.63	28	47	11.06	14.90	14.38	46	54
	GL-Uni	16.01	21.54	22.75	25	47	13.92	19.11	19.91	40	58
$\delta^0 = 1$	OLS	8.71	15.87	15.79	27	34	7.24	10.73	10.72	47	54
	GL-LN	8.25	15.27	15.61	27	34	6.88	9.21	9.19	47	52
	GL-Uni	8.65	14.96	15.21	27	33	6.96	10.44	10.45	46	53

The model is $y_t = \delta^0 \mathbf{1}\{t > \lfloor T\lambda_0 \rfloor\} + e_t$, $e_t = 0.3e_{t-1} + u_t$, $u_t \sim i.i.d. \mathscr{N}(0, 1)$, $T = 100$. The notes of Table 1 apply.

Table 3: Small-sample accuracy of the estimates of the break point T_b^0 for model M3

		MAE	Std	RMSE	$Q_{0.25}$	$Q_{0.75}$	MAE	Std	RMSE	$Q_{0.25}$	$Q_{0.75}$
		$\lambda_0 = 0.3$					$\lambda_0 = 0.5$				
$\delta^0 = 0.3$	OLS	23.66	28.14	31.32	22	69	22.01	26.61	26.59	33	72
	GL-LN	19.31	19.22	26.28	30	57	14.89	18.18	19.08	39	61
	GL-Uni	21.38	24.12	28.08	24	64	18.76	22.88	22.01	31	66
$\delta^0 = 0.4$	OLS	19.31	25.76	27.71	23	57	18.14	23.43	23.44	38	60
	GL-LN	15.04	17.64	21.29	29	51	12.36	16.43	16.52	40	60
	GL-Uni	18.46	22.74	25.18	25	58	15.91	20.42	20.42	37	62
$\delta^0 = 0.6$	OLS	12.02	19.02	19.82	25	37	10.28	15.51	15.58	45	55
	GL-LN	9.29	12.86	14.61	29	40	8.46	11.84	11.86	45	55
	GL-Uni	12.33	18.43	19.54	27	41	8.90	14.54	14.53	45	55
$\delta^0 = 1$	OLS	3.72	6.88	6.89	28	32	3.85	6.98	6.98	48	52
	GL-LN	3.49	6.44	6.57	28	32	3.45	6.09	6.10	48	52
	GL-Uni	4.37	8.12	8.24	28	32	3.86	6.97	6.96	48	52

The model is $y_t = \delta^0 \mathbf{1}\{t > \lfloor T\lambda_0 \rfloor\} + \alpha^0 y_{t-1} + e_t$, $e_t \sim i.i.d.\, \mathscr{N}(0, 0.5)$, $\alpha^0 = 0.6$, $T = 100$. The notes of Table 1 apply.

Table 4: Small-sample coverage rates and lengths of the confidence sets for model M1

		$\delta^0 = 0.4$		$\delta^0 = 0.8$		$\delta^0 = 1.2$		$\delta^0 = 1.6$	
		Cov.	Lgth.	Cov.	Lgth.	Cov.	Lgth.	Cov.	Lgth.
$\lambda_0 = 0.5$	OLS-CR	0.922	77.52	0.934	49.46	0.946	22.51	0.938	10.48
	Bai (1997)	0.812	58.12	0.862	28.75	0.928	13.78	0.928	8.16
	$\widehat{U}_T(T_{\mathrm{m}})$.neq	0.950	75.45	0.950	41.68	0.950	21.78	0.950	14.79
	ILR	0.959	76.14	0.973	35.79	0.976	14.44	0.977	7.15
	GL-LN	0.942	49.76	0.948	22.45	0.958	10.47	0.965	5.15
	sup-W	0.384		0.916		1.000		1.000	
$\lambda_0 = 0.3$	OLS-CR	0.928	74.95	0.928	46.68	0.930	21.47	0.958	10.22
	Bai (1997)	0.830	56.64	0.870	28.72	0.904	13.89	0.962	8.27
	$\widehat{U}_T(T_{\mathrm{m}})$.neq	0.952	77.51	0.952	44.72	0.952	22.51	0.952	14.21
	ILR	0.952	78.28	0.966	39.78	0.969	31.29	0.968	18.23
	GL-LN	0.942	49.60	0.948	23.89	0.958	11.14	0.980	5.60
	sup-W	0.316		0.866		0.992		1.000	

The model is $y_t = \delta^0 \mathbf{1}_{\{t > \lfloor T\lambda_0 \rfloor\}} + e_t$, $e_t \sim i.i.d.\, \mathscr{N}(0, 1)$, $T = 100$. Cov. and Lgth. refer to the coverage probability and the average length of the confidence set (i.e., the average number of dates in the confidence set). sup-W refers to the rejection probability of the sup-Wald test using a 5% asymptotic critical value. The number of simulations is 3,000.

Table 5: Small-sample coverage rates and lengths of the confidence sets for model M2

		$\delta^0=0.4$		$\delta^0=0.8$		$\delta^0=1.2$		$\delta^0=1.6$	
		Cov.	Lgth.	Cov.	Lgth.	Cov.	Lgth.	Cov.	Lgth.
$\lambda_0=0.5$	OLS-CR	0.952	80.29	0.954	57.70	0.957	30.04	0.963	15.10
	Bai (1997)	0.804	64.64	0.824	43.53	0.907	13.03	0.930	7.81
	$\widehat{U}_T(T_{\mathrm{m}})$.neq	0.967	87.30	0.967	72.70	0.957	36.70	0.957	30.20
	ILR	0.937	81.88	0.945	57.43	0.972	21.99	0.972	18.96
	GL-LN	0.933	55.13	0.912	32.97	0.935	20.03	0.961	10.62
	sup-W	0.316		0.699		1.000		1.000	
$\lambda_0=0.3$	OLS-CR	0.945	79.25	0.957	54.93	0.962	29.91	0.970	15.37
	Bai (1997)	0.823	63.79	0.851	26.33	0895	13.07	0.946	7.87
	$\widehat{U}_T(T_{\mathrm{m}})$.neq	0.966	88.23	0.953	59.66	0.950	39.65	0.951	32.39
	ILR	0.945	84.37	0.945	62.97	0.971	33.74	0.987	17.92
	GL-LN	0.945	53.79	0.923	34.75	0.934	19.92	0.944	10.04
	sup-W	0.314		0.881		0.999		1.000	

The model is $y_t=\delta^0\mathbf{1}\{t>\lfloor T\lambda_0\rfloor\}+e_t$, $e_t=0.3e_{t-1}+u_t$, $u_t\sim i.i.d.\,\mathscr{N}(0,1)$, $T=100$. The notes of Table 4 apply.

Table 6: Small-sample coverage rates and lengths of the confidence sets for model M3

		$\delta^0=0.4$		$\delta^0=0.8$		$\delta^0=1.2$		$\delta^0=1.6$	
		Cov.	Lgth.	Cov.	Lgth.	Cov.	Lgth.	Cov.	Lgth.
$\lambda_0=0.5$	OLS-CR	0.954	80.29	0.952	57.23	0.957	30.21	0.963	15.20
	Bai (1997)	0.781	55.85	0.845	26.23	0.902	13.03	0.932	7.81
	$\widehat{U}_T(T_{\mathrm{m}})$.neq	0.958	81.28	0.959	55.34	0.957	36.71	0.957	30.20
	ILR	0.934	65.96	0.956	33.73	0.975	21.96	0.984	17.45
	GL-LN	0.912	60.90	0.925	32.93	0.964	19.23	0.971	9.23
	sup-W	0.407		0.931		1.000		1.000	
$\lambda_0=0.3$	OLS-CR	0.968	83.69	0.951	54.13	0.962	29.31	0.970	15.37
	Bai (1997)	0.795	64.06	0.853	26.33	0.896	13.07	0.946	7.85
	$\widehat{U}_T(T_{\mathrm{m}})$.neq	0.960	86.42	0.953	59.13	0.950	39.65	0.951	32.28
	ILR	0.934	67.73	0.964	35.30	0.971	33.74	0.987	17.92
	GL-LN	0.912	60.28	0.945	36.08	0.974	22.72	0.975	12.71
	sup-W	0.232		0.884		0.999		1.000	

The model is $y_t=\delta^0\mathbf{1}\{t>\lfloor T\lambda_0\rfloor\}+\alpha^0y_{t-1}+e_t$, $e_t\sim i.i.d.\,\mathscr{N}(0,0.5)$, $\alpha^0=0.6$, $T=100$. The notes of Table 4 apply.

Supplemental Material to

Generalized Laplace Inference in Multiple Change-Points Models

Alessandro Casini
University of Rome Tor Vergata

Pierre Perron
Boston University

2nd March 2020

Abstract

This supplemental material is structured as follows. Section A contains the Mathematical Appendix which includes all proofs of the results in the paper.

A Mathematical Appendix

The mathematical appendix is structured as follows. Section A.2 presents some preliminary lemmas which will be used in the sequel. The proofs of the theoretical results in the paper are in Section A.3-A.5.

A.1 Additional Notation

The $(i,\, j)$ element of A is denoted by $A^{(i,j)}$. For a matrix A, the orthogonal projection matrices P_A, M_A are defined as $P_A = A\left(A'A\right)^{-1} A'$ and $M_A = I - P_A$, respectively. Also, for a projection matrix P, $\|PA\| \leq \|A\|$. We denote the d-dimensional identity matrix by I_d. When the context is clear we omit the subscript notation in the projection matrices. We denote the $i \times j$ upper-left (resp., lower-right) sub-block of A as $[A]_{\{i\times j,\cdot\}}$ (resp., $[A]_{\{\cdot,i\times j\}}$). Note that the norm of A is equal to the square root of the maximum eigenvalue of $A'A$, and thus, $\|A\| \leq [\mathrm{tr}\,(A'A)]^{1/2}$. For a sequence of matrices $\{A_T\}$, we write $A_T = o_{\mathbb{P}}(1)$ if each of its elements is $o_{\mathbb{P}}(1)$ and likewise for $O_{\mathbb{P}}(1)$. For a random variable ξ and a number $r \geq 1$, $\|\xi\|_r = \left(\mathbb{E}\,\|\xi\|^r\right)^{1/r}$. K is a generic constant that may vary from line to line; we may sometime write K_r to emphasize the dependence of K on a number r. For two scalars a and b, $a \wedge b = \inf\{a,\, b\}$. We may use $\sum_k$ when the limit of the summation are clear from the context. Unless otherwise sated $\mathbf{A}^c$ denotes the complementary set of $\mathbf{A}$.

A.2 Preliminary Lemmas

We first present results related to the extremum criterion function $Q_T\left(\delta\left(T_b\right),\, T_b\right)$ under the following assumption (Assumption 3.1-3.2 are not needed in this section).

Assumption A.1. *We consider model (2.3) with Assumption 2.1-2.4 and 3.3-3.5.*

Lemma A.1. *The following inequalities hold $\mathbb{P}$-a.s.:*

$$\left(Z_0'MZ_0\right) - \left(Z_0'MZ_2\right)\left(Z_2'MZ_2\right)^{-1}\left(Z_2'MZ_0\right) \geq D'\left(X_\Delta'X_\Delta\right)\left(X_2'X_2\right)^{-1}\left(X_0'X_0\right)D, \qquad T_b < T_b^0 \tag{A.1}$$

$$\left(Z_0'MZ_0\right) - \left(Z_0'MZ_2\right)\left(Z_2'MZ_2\right)^{-1}\left(Z_2'MZ_0\right) \geq D'\left(X_\Delta'X_\Delta\right)\left(X'X - X_2'X_2\right)^{-1}\left(X'X - X_0'X_0\right)D, \qquad T_b \geq T_b^0 \tag{A.2}$$

Proof. See Lemma A.1 in Bai (1997). □

Recall that $Q_T\left(\delta\left(T_b\right),\, T_b\right) = \delta\left(T_b\right)\left(Z_2'MZ_2\right)\delta\left(T_b\right)$. We decompose $Q_T\left(\delta\left(T_b\right),\, T_b\right) - Q_T\left(\delta\left(T_b^0\right),\, T_b^0\right)$ into a "deterministic" and a "stochastic" component. It follows by definition that,

$$\delta\left(T_b\right) = \left(Z_2'MZ_2\right)^{-1}\left(Z_2'MY\right) = \left(Z_2'MZ_2\right)^{-1}\left(Z_2'MZ_0\right)\delta_T + \left(Z_2'MZ_2\right)^{-1} Z_2Me,$$

and

$$\delta\left(T_b^0\right) = \left(Z_0'MZ_0\right)^{-1}\left(Z_0'MY\right) = \delta_T + \left(Z_0'MZ_0\right)^{-1}\left(Z_0'Me\right).$$

Therefore

$$Q_T\left(\delta\left(T_b\right),\, T_b\right) - Q_T\left(\delta\left(T_b^0\right),\, T_b^0\right) = \delta\left(T_b\right)'\left(Z_2'MZ_2\right)\delta\left(T_b\right) - \delta\left(T_b^0\right)'\left(Z_0'MZ_0\right)\delta\left(T_b^0\right) \tag{A.3}$$

$$\triangleq g_d\left(\delta_T,\, T_b\right) + g_e\left(\delta_T,\, T_b\right), \tag{A.4}$$

where

$$g_d\left(\delta_T,\, T_b\right) = \delta_T'\left\{\left(Z_0'MZ_2\right)\left(Z_2'MZ_2\right)^{-1}\left(Z_2'MZ_0\right) - Z_0'MZ_0\right\}\delta_T, \tag{A.5}$$

and

$$g_e(\delta_T, T_b) = 2\delta_T'(Z_0'MZ_2)(Z_2'MZ_2)^{-1}Z_2Me - 2\delta_T'(Z_0'Me) \tag{A.6}$$
$$+ e'MZ_2(Z_2'MZ_2)^{-1}Z_2Me - e'MZ_0(Z_0'MZ_0)^{-1}Z_0'Me. \tag{A.7}$$

(A.5) constitutes the deterministic component and $g_e(\delta_T, T_b)$ the stochastic one. Denote

$$X_\Delta \triangleq X_2 - X_0 = \left(0, \ldots, 0, x_{T_b+1}, \ldots, x_{T_b^0}, 0, \ldots,\right)', \qquad \text{for } T_b < T_b^0$$
$$X_\Delta \triangleq -(X_2 - X_0) = \left(0, \ldots, 0, x_{T_b^0+1}, \ldots, x_{T_b}, 0, \ldots,\right)', \qquad \text{for } T_b > T_b^0$$

whereas $X_\Delta \triangleq 0$ when $T_b = T_b^0$. Observe that $X_2 = X_0 + X_\Delta \text{sign}(T_b^0 - T_b)$. When the sign is immaterial, we simply write $X_2 = X_0 + X_\Delta$. Next, let $Z_\Delta = X_\Delta D$, and define

$$\overline{g}_d(\delta_T, T_b) \triangleq -\frac{g_d(\delta_T, T_b)}{|T_b - T_b^0|}. \tag{A.8}$$

We arbitrarily define $\overline{g}_d(\delta^0, T_b) = \delta_T'\delta_T$ when $T_b = T_b^0$. Observe that $\overline{g}_d(\delta_T, T_b)$ is non-negative because the matrix inside the braces in (A.5) is negative semidefinite. (A.3) can be written as

$$Q_T(\delta(T_b), T_b) - Q_T\left(\delta\left(T_b^0\right), T_b^0\right) = -\left|T_b - T_b^0\right|\overline{g}_d(\delta_T, T_b) + g_e(\delta_T, T_b), \qquad \text{for all } T_b. \tag{A.9}$$

We use the notation $u = T\|\delta_T\|^2(\lambda_b - \lambda_0)$ and $T_b = T\lambda_b$. For $\eta > 0$, let $B_{T,\eta} \triangleq \{T_b : |T_b - T_b^0| \leq T\eta\}$, $B_{T,K} \triangleq \left\{T_b : |T_b - T_b^0| \leq K/\|\delta_T\|^2\right\}$ and $B_{T,K}^c \triangleq \left\{T_b : T\eta \geq |T_b - T_b^0| > K/\|\delta_T\|^2\right\}$, with $K > 0$. Note that $B_{T,\eta} = B_{T,K} \cup B_{T,K}^c$. Further, let $B_{T,\eta}^c \triangleq \{T_b : |T_b - T_b^0| > T\eta\}$.

Lemma A.2. *Under Assumption A.1,* $Q_T(\delta(T_b), T_b) - Q_T(\delta(T_b^0), T_b^0) = -\delta_T'Z_\Delta'Z_\Delta\delta_T + 2\text{sgn}(T_b^0 - T_b)\delta_T'Z_\Delta'e + o_{\mathbb{P}}(1)$*, uniformly on* $B_{T,K}$ *for* K *large enough.*

Proof. It follows from Lemma A.5 in Bai (1997). □

Lemma A.3. *Under Assumption A.1, for* $T_b = T_b^0 + \left\lfloor u/\|\delta_T\|^2\right\rfloor$*, we have* $\delta_T'Z_\Delta'Z_\Delta\delta_T = \delta_T'\sum_{t=T_b+1}^{T_b^0} z_t z_t'\delta_T = |u|(\delta^0)'\,\overline{V}\delta^0 + o_{\mathbb{P}}(1)$*, where* $\overline{V} = V_1$ *if* $u \leq 0$ *and* $\overline{V} = V_2$ *if* $u > 0$.

Proof. It follows from basic arguments (cf. Assumptions 3.4-3.5). □

Lemma A.4. *Under Assumption A.1, for any* $\epsilon > 0$ *there exists a* $C < \infty$ *and a positive sequence* $\{\nu_T\}$*, with* $\nu_T \to \infty$ *as* $T \to \infty$*, such that*

$$\liminf_{T\to\infty} \mathbb{P}\left[\sup_{K \leq |u| \leq \eta T\|\delta_T\|^2} Q_T(\delta(T_b), T_b) - Q_T\left(\delta\left(T_b^0\right), T_b^0\right) < -C\nu_T\right] \geq 1 - \epsilon,$$

for all sufficiently large K *and a sufficiently small* $\eta > 0$.

Proof. Note that on $\left\{K \leq |u| \leq \eta T\|\delta_T\|^2\right\}$ we have $K/\|\delta_T\|^2 \leq |T_b - T_b^0| \leq \eta T$. In view of (A.8), the statement $Q_T(\delta(T_b), T_b) - Q_T(\delta(T_b^0), T_b^0) < -C\nu_T$ follows from showing that as $T \to \infty$,

$$\mathbb{P}\left(\sup_{T_b \in B_{K,T}^c} g_e(\delta_T, T_b) \geq \inf_{T_b \in B_{K,T}^c}\left|T_b - T_b^0\right|^\kappa \overline{g}_d(\delta_T, T_b)\right) < \epsilon,$$

where $\kappa \in (1/2,\,1)$. Suppose $T_b < T_b^0$. We show that

$$\mathbb{P}\left(\sup_{T_b \in B^c_{K,T}} \frac{\|\delta_T\|}{K} g_e\left(\delta_T,\,T_b\right) \geq \frac{1}{\|\delta_T\|^{2\kappa-1}} \left(\frac{1}{K}\right)^{1-\kappa} \inf_{T_b \in B^c_{K,T}} \overline{g}_d\left(\delta_T,\,T_b\right)\right) < \epsilon. \tag{A.10}$$

Lemma A.5-(ii) stated below implies that $\inf_{T_b \in B^c_{T,K}} \overline{g}_d\left(\delta_T,\,T_b\right)$ is bounded away from zero as $T \to \infty$ for large K and small η. Next, we show that

$$\sup_{T_b \in B^c_{K,T}} K^{-1} \|\delta_T\| \, g_e\left(\delta_T,\,T_b\right) = o_{\mathbb{P}}(1). \tag{A.11}$$

Consider the first term of (A.6),

$$\begin{aligned} 2\delta_T'\left(Z_0'MZ_2\right)\left(Z_2'MZ_2\right)^{-1} Z_2Me &= 2\delta_T'\left(Z_0'MZ_2/T\right)\left(Z_2'MZ_2/T\right)^{-1} Z_2Me \\ &= 2C\|\delta_T\| O_{\mathbb{P}}(1)\, O_{\mathbb{P}}(1)\, O_{\mathbb{P}}\left(T^{1/2}\right) = CO_{\mathbb{P}}\left(\|\delta_T\| T^{1/2}\right). \end{aligned}$$

When multiplied by $\|\delta_T\|/K$, this term is $O_{\mathbb{P}}\left(\|\delta_T\|^2 T^{1/2}/K\right)$ which goes to zero for large K.. The second term in (A.6), when multiplied by $\|\delta_T\|/K$, is

$$2K^{-1}\|\delta_T\|\,\delta_T'\left(Z_0'Me\right) = K^{-1}\|\delta_T\| O_{\mathbb{P}}\left(\|\delta_T\| T^{1/2}\right) = K^{-1} O_{\mathbb{P}}\left(\|\delta_T\|^2 T^{1/2}\right),$$

which converges to zero using the same argument as for the first term. Consider now the first term of (A.7), $T^{-1/2}e'MZ_2\left(Z_2'MZ_2/T\right)^{-1} T^{-1/2} Z_2Me = O_{\mathbb{P}}(1)$. A similar argument can be used for the second term which is also $O_{\mathbb{P}}(1)$. The latter two terms multiplied by $\|\delta_T\|/K$ is $O_{\mathbb{P}}(\|\delta_T\|/K) = o_{\mathbb{P}}(1)$. This proves (A.11) and thus (A.10). To conclude the proof, note that $\kappa \in (1/2,\,1)$ implies $\|\delta_T\|^{-(2\kappa-1)} \to \infty$, so that we can choose $\nu_T = \left(\|\delta_T\|^2/K\right)^{-(1-\kappa)}$. □

Lemma A.5. *Let* $\widetilde{g}_d \triangleq \inf_{\left|T_b - T_b^0\right| > K\|\delta_T\|^{-2}} \overline{g}_d\left(\delta_T,\,T_b\right)$. *Under Assumption A.1,*
(i) for any $\epsilon > 0$ *there exists some* $C > 0$ *such that* $\liminf_{T\to\infty} \mathbb{P}\left(\widetilde{g}_d > C\|\delta_T\|^2\right) \leq 1 - \epsilon$*;*
(ii) with $B^c_{T,K} = \left\{T_b : T\eta \geq \left|T_b - T_b^0\right| \geq K/\|\delta_T\|^2\right\}$, *for any* $\epsilon > 0$ *there exists a* $C > 0$ *such that* $\liminf_{T\to\infty} \mathbb{P}\left(\inf_{T_b \in B^c_{T,K}} \overline{g}_d\left(\delta_T,\,T_b\right) > C\right) \leq 1 - \epsilon$.

Proof. Part (i) was proved in Lemma A.2 of Bai (1997). As for part (ii), by Lemma A.1,

$$\overline{g}_d\left(\delta^0,\,T_b\right) \geq \delta_T D' \frac{X_\Delta' X_\Delta}{T_b^0 - T_b} \left(X_2'X_2\right)^{-1}\left(X_0'X_0\right) D\delta_T \geq \lambda_{J,T_b},$$

where λ_{J,T_b} is the minimum eigenvalue of $D'J\left(T_b\right)D$, with $J\left(T_b\right) \triangleq \|\delta_T\|^2 \left(T_b^0 - T_b\right)^{-1} X_\Delta' X_\Delta \left(X_2'X_2\right)^{-1}\left(X_0'X_0\right)$. It is sufficient to show that, for $T_b \in B^c_{T,K}$, λ_{J,T_b} is bounded away from zero with large probability for large K and small η. We have $\left\|J\left(T_b\right)^{-1}\right\| \leq \left\|\left[\|\delta_T\|^2 \left(T_b^0 - T_b\right)^{-1} X_\Delta' X_\Delta\right]^{-1}\right\| \left\|\left(X_2'X_2\right)\left(X_0'X_0\right)^{-1}\right\|$ and by Assumption 2.3-2.4 $\left\|\left(X_2'X_2\right)\left(X_0'X_0\right)^{-1}\right\| \leq \|X'X\| \left\|\left(X_0'X_0\right)^{-1}\right\|$ is bounded. Next, note that $\left(T_b^0 - T_b\right)^{-1} X_\Delta' X_\Delta = \left(T_b^0 - T_b\right)^{-1} \sum_{t=T_b+1}^{T_b^0} x_t x_t'$ is larger than $(T\eta)^{-1} \sum_{t=T_b^0 - \lfloor K/\|\delta_T\|^2 \rfloor}^{T_b^0} x_t x_t'$ on $B^c_{T,K}$, and for all K, $\left(\|\delta_T\|^2/K\right) \sum_{t=T_b^0 - \lfloor K/\|\delta_T\|^2 \rfloor}^{T_b^0} x_t x_t'$ is positive definite with large probability as $T \to \infty$ by Assumption 2.3. Now, $(K/T\eta)\left(\|\delta_T\|^2/K\right) \sum_{t=T_b^0 - \lfloor K/\|\delta_T\|^2 \rfloor}^{T_b^0} x_t x_t' = O_{\mathbb{P}}(1)$, by choosing sufficiently large K and small η. Thus, $\left\|\left[\|\delta_T\|^2 \left(T_b^0 - T_b\right)^{-1} X_\Delta' X_\Delta\right]^{-1}\right\|$ is bounded with large probability for such large K

and small η, which in turn implies that $\|J(T_b)^{-1}\|$ is bounded. Since D has full column rank, λ_{J,T_b} is bounded away from zero for sufficiently large K and small η. □

Lemma A.6. *Under Assumption A.1, for any $\epsilon > 0$ there exists a $C > 0$ such that*

$$\liminf_{T\to\infty} \mathbb{P}\left[\sup_{|u|\geq T\|\delta_T\|^2\eta} Q_T\left(\delta\left(T_b\right), T_b\right) - Q_T\left(\delta\left(T_b^0\right), T_b^0\right) < -C\nu_T\right] \geq 1-\epsilon,$$

for every $\eta > 0$, where $\nu_T \to \infty$.

Proof. Fix any $\eta > 0$. Note that on $\left\{|u| \geq T\|\delta_T\|^2 \eta\right\}$ we have $|T_b - T_b^0| \geq T\eta$. We proceed in a similar manner to Lemma A.4. Let $B_{T,\eta}^c \triangleq \{T_b : |T_b - T_b^0| \geq T\eta\}$ and recall (A.8). First, as in Lemma A.5-(i), we have $\inf_{T_b\in B_{T,\eta}^c} \overline{g}_d(\delta_T, T_b) \geq C\|\delta_T\|^2$ with large probability for some $C > 0$. Noting that $T\eta \inf_{T_b\in B_{T,\eta}^c} \overline{g}_d(\delta_T, T_b)$ diverges at rate $\tau_T = T\|\delta_T\|^2$, the claim follows if we can show that $g_e(\delta_T, T_b) = O_{\mathbb{P}}(\tau_T^{\varpi})$, with $0 \leq \varpi < 1$ uniformly on $B_{T,\eta}^c$. This is shown in Lemma A.7 below, which suggests setting $\varpi \in (1/2, 1)$. Then, choose $\nu_T = \left(T\|\delta_T\|^2\right)^{1-\varpi}$. □

Lemma A.7. *Under Assumption A.1, uniformly on $B_{T,\eta}^c$, $|g_e(\delta_T, T_b)| = O_{\mathbb{P}}\left(\|\delta_T\| T^{1/2}\log T\right)$.*

Proof. We show that $T^{-1}|g_e(\delta^0, T_b)| = O_{\mathbb{P}}\left(\|\delta_T\| T^{-1/2}\log T\right)$ uniformly on $B_{T,\eta}^c$. Note that

$$\sup_{T_b\in B_{T,\eta}^c} |g_e(\delta_T, T_b)| \leq \sup_{q\leq T_b\leq T-q} |g_e(\delta_T, T_b)|,$$

and recall that $q = \dim(z_t)$ is needed for identification. Observe that

$$\sup_{q\leq T_b\leq T-q} \left\|\left(Z_2'MZ_2\right)^{-1/2} Z_2'Me\right\| = O_{\mathbb{P}}(\log T), \tag{A.12}$$

by the law of iterated logarithms [cf. Billingsley (1995), Ch. 1, Theorem 9.5]. Next,

$$\sup_{q\leq T_b\leq T-q} T^{-1/2}\left(Z_0'MZ_2\right)\left(Z_2'MZ_2\right)^{-1/2} = O_{\mathbb{P}}(1), \tag{A.13}$$

which can be proved using the inequality $(Z_0'MZ_2)(Z_2'MZ_2)(Z_0'MZ_2) \leq Z_0'MZ_0 = O_{\mathbb{P}}(T)$ (valid for all T_b). Thus, by (A.12) and (A.13), the first term on the right-hand side of (A.6) multiplied by T^{-1} is such that

$$\sup_{q\leq T_b\leq T-q} 2\delta_T' T^{-1}\left(Z_0'MZ_2\right)\left(Z_2'MZ_2\right)^{-1} Z_2'Me = O_{\mathbb{P}}\left(\|\delta_T\| T^{-1/2}\log T\right). \tag{A.14}$$

The second term on the right-hand side of (A.6) is $2\delta_T' Z_0'Me = O_{\mathbb{P}}\left(\|\delta_T\| T^{1/2}\right)$. Using (A.12), and dividing by T, the first term of (A.7) is $O_{\mathbb{P}}\left((\log T)^2/T\right)$ while the last term is $O_{\mathbb{P}}(T^{-1})$. When divided by T, they are of order $O_{\mathbb{P}}\left((\log T)^2/T\right)$ and $O_{\mathbb{P}}(T^{-1})$, respectively. Therefore, $|g_e(T_b, \delta^0)| = O_{\mathbb{P}}\left(\|\delta_T\| T^{1/2}\log T\right)$, uniformly on $B_{T,\eta}^c$. □

A.3 Proofs of Results in Section 3

We denote by $\boldsymbol{P}$ the class of polynomial functions $p : \mathbb{R} \to \mathbb{R}$. Let $\mathbf{U}_T \triangleq \{u \in \mathbb{R} : \lambda_b^0 + u/\psi_T \in \Gamma^0\}$, $\Gamma_{T,\psi} \triangleq \{u \in \mathbb{R} : |u| \leq \psi_T\}$, $\Gamma_{T,\psi}^c \triangleq \mathbb{R} - \Gamma_{T,\psi}$, and $\widetilde{\mathbf{U}}_T^c \triangleq \mathbf{U}_T - \Gamma_{T,\psi}$. For $u \in \mathbb{R}$, let $R_{T,v}(u) \triangleq Q_{T,v}(u) -$

$\Lambda^0(u)$ and $\overline{G}_{T,v}(u) \triangleq \sup_{\widetilde{v}\in\mathbf{V}} \widetilde{G}_{T,v}(u, \widetilde{v})$. The generic constant $0 < C < \infty$ used below may change from line to line. Finally, let $\widetilde{\gamma}_T \triangleq \gamma_T / T \|\delta_T\|^2$.

A.3.1 Proof of Proposition 3.1

We begin with the proof for the case of a fixed shift.

Lemma A.8. *Under Assumption 2.1-2.4, 3.1-3.3 (except that $\delta_T = \delta^0$) and 3.6-(i), $\widehat{\lambda}_b^{\mathrm{GL}} = \lambda_b^0 + o_{\mathbb{P}}(1)$.*

Proof. Let $\overline{S}_T(\delta(\lambda_b), \lambda_b) \triangleq Q_T(\delta(\lambda_b), \lambda_b) - Q_T(\delta(\lambda_b^0), \lambda_b^0)$. From (A.9),

$$\overline{S}_T\left(\widehat{\delta}(\lambda_b), \lambda_b\right) = -\left|T_b - T_b^0\right| \overline{g}_d\left(\delta^0, T_b\right) + g_e\left(\delta^0, T_b\right),$$

where $g_e(\delta^0, T_b)$ and $\overline{g}_d(\delta^0, T_b)$ are defined in (A.6)-(A.8). By Lemma A.24 in Bai (1997), $\liminf_{T\to\infty} \overline{g}_d$ $(\delta^0, T_b) > 0$ and $T^{-1}\sup_{T_b} |g_e(\delta^0, T_b)| = O_{\mathbb{P}}\left(T^{-1/2}\log T\right)$. Thus, for any $B > 0$ if $\left|\widehat{\lambda}_b^{\mathrm{GL}} - \lambda_b^0\right| > B$ we have that,

$$-\overline{S}_T\left(\widehat{\delta}(\lambda_b), \lambda_b\right) \to \infty \,\text{at rate}\, TB. \tag{A.15}$$

Let $p_T(u) \triangleq p_{1,T}(u)/\overline{p}_T$ with $p_{1,T}(u) = \exp(Q_T(\delta(u), u))$ and $\overline{p}_T \triangleq \int_{\mathbf{U}_T} p_{1,T}(w)\,dw$. By definition, $\widehat{\lambda}_b^{\mathrm{GL}}$ is the minimum of the function $\int_{\Gamma^0} l(s-u)\,p_{1,T}(u)\,\pi(u)\,du$ with $s \in \Gamma^0$. Using a change in variables,

$$\begin{aligned}&\int_{\Gamma^0} l(s-u)\,p_{1,T}(u)\,\pi(u)\,du \\ &\quad = T^{-1}\overline{p}_T \int_{\mathbf{U}_T} l\left(T\left(s - \lambda_b^0\right) - u\right) p_T\left(\lambda_b^0 + T^{-1}u\right) \pi\left(\lambda_b^0 + T^{-1}u\right) du,\end{aligned}$$

where $\mathbf{U}_T \triangleq \{u \in \mathbb{R} : \lambda_b^0 + T^{-1}u \in \Gamma^0\}$. Thus, $\lambda_{\delta,T} \triangleq T\left(\widehat{\lambda}_b^{\mathrm{GL}} - \lambda_b^0\right)$ is the minimum of the function,

$$\mathcal{S}_T(s) \triangleq \int_{\mathbf{U}_T} l(s-u) \frac{p_T\left(\lambda_b^0 + T^{-1}u\right)\pi\left(\lambda_b^0 + T^{-1}u\right)}{\int_{\mathbf{U}_T} p_T\left(\lambda_b^0 + T^{-1}w\right)\pi\left(\lambda_b^0 + T^{-1}w\right)dw} du,$$

where the optimization is over $\mathbf{U}_T$. We shall show that for any $B > 0$,

$$\mathbb{P}\left[\left|\widehat{\lambda}_b^{\mathrm{GL}} - \lambda_b^0\right| > B\right] \leq \mathbb{P}\left[\inf_{|s|>TB} \mathcal{S}_T(s) \leq \mathcal{S}_T(0)\right] \to 0. \tag{A.16}$$

By assumption the prior is bounded and so we can proceed the proof for the case $\pi(u) = 1$ for all u. By the properties of the family $\boldsymbol{L}$ of loss functions, we can find $\overline{u}_1, \overline{u}_2 \in \mathbb{R}$, with $0 < \overline{u}_1 < \overline{u}_2$ such that as T increases,

$$\bar{l}_{1,T} \triangleq \sup\{l(u) : u \in \Gamma_{1,T}\} < \bar{l}_{2,T} \triangleq \inf\{l(u) : u \in \Gamma_{2,T}\},$$

where $\Gamma_{1,T} \triangleq \mathbf{U}_T \cap (|u| \leq \overline{u}_1)$ and $\Gamma_{2,T} \triangleq \mathbf{U}_T \cap (|u| > \overline{u}_2)$. With this notation,

$$\mathcal{S}_T(0) \leq \bar{l}_{1,T} \int_{\Gamma_{1,T}} p_T(u)\,du + \int_{\mathbf{U}_T \cap (|u| > \overline{u}_1)} l(u)\,p_T(u)\,du.$$

If $l \in \boldsymbol{L}$ then for a sufficiently large T the following relationship holds: $l(u) - \inf_{|v|>TB/2} l(v) \leq 0$, $|u| \leq (TB/2)^{\vartheta}$ for some $\vartheta > 0$. It also follows that for large T we have $TB > 2\overline{u}_2$ and $(TB/2)^{\vartheta} > \overline{u}_2$. Let

$\Gamma_{T,B} \triangleq \{u : (|u| > TB/2) \cap \mathbf{U}_T\}$. Then, whenever $|s| > TB$ and $|u| \leq TB/2$, we have,

$$|u - s| > TB/2 > \overline{u}_2 \qquad \text{and} \qquad \inf_{u \in \Gamma_{T,B}} l\left(u\right) \geq \bar{l}_{2,T}. \tag{A.17}$$

With this notation,

$$\begin{aligned}\inf_{|s|>TB} \mathcal{S}_T\left(s\right) &\geq \inf_{u \in \Gamma_{T,B}} l_T\left(u\right) \int_{(|w| \leq TB/2) \cap \mathbf{U}_T} p_T\left(w\right) dw \\ &\geq \bar{l}_{2,T} \int_{(|w| \leq TB/2) \cap \mathbf{U}_T} p_T\left(w\right) dw,\end{aligned}$$

from which it follows that

$$\begin{aligned}\mathcal{S}_T\left(0\right) - \inf_{|s|>TB} \mathcal{S}_T\left(s\right) \leq &-\varpi \int_{\Gamma_{1,T}} p_T\left(u\right) du \\ &+ \int_{\mathbf{U}_T \cap \left((TB/2)^{\vartheta} \geq |u| \geq \overline{u}_1\right)} \left(l\left(u\right) - \inf_{|s|>TB/2} l_T\left(s\right)\right) p_T\left(u\right) du \\ &+ \int_{\mathbf{U}_T \cap \left(|u| > (TB/2)^{\vartheta}\right)} l\left(u\right) p_T\left(u\right) du,\end{aligned}$$

where $\varpi \triangleq \bar{l}_{2,T} - \bar{l}_{1,T}$. The last inequality can be manipulated further using (A.17),

$$\begin{aligned}\mathcal{S}_T\left(0\right) - \inf_{|s|>TB} \mathcal{S}_T\left(s\right) \leq &-\varpi \int_{\Gamma_{1,T}} p_T\left(u\right) du \\ &+ \int_{\mathbf{U}_T \cap \left(|u| > (TB/2)^{\vartheta}\right)} l_T\left(u\right) p_T\left(u\right) du.\end{aligned} \tag{A.18}$$

Since $l \in \boldsymbol{L}$, we have $l\left(u\right) \leq |u|^a$, $a > 0$ when u is large enough. Thus, given (A.15), the second term of (A.18) converges to zero. Since $\int_{\Gamma_{1,T}} p_T\left(u\right) du > 0$ the first term of (A.18) is negative which then leads to $\mathcal{S}_T\left(0\right) - \inf_{|s|>TB} \mathcal{S}_T\left(s\right) < 0$ or $\mathcal{S}_T\left(0\right) < \inf_{|s|>TB} \mathcal{S}_T\left(s\right)$. Thus, we have (A.16). □

Lemma A.9. *Under Assumption 2.1-2.4, 3.1-3.3 and 3.6-(i), for $l \in \boldsymbol{L}$ and any $B > 0$ and $\varepsilon > 0$, we have for all large T,* $\mathbb{P}\left[\left|\widehat{\lambda}_b^{\mathrm{GL}} - \lambda_b^0\right| > B\right] < \varepsilon$.

Proof. The structure of the proof is similar to that of Lemma A.8. By Proposition 1 in Bai (1997), eq. (A.15) holds with $O_{\mathbb{P}}\left(T\left\|\delta_T\right\|^2\right)$ in place of $O_{\mathbb{P}}\left(TB\right)$, $B > 0$. One can then follow the same steps as in the previous lemma to yield the result. □

Lemma A.10. *Under Assumption 2.1-2.4, 3.1-3.3 and 3.6-(i), for $l \in \boldsymbol{L}$ and for every $\varepsilon > 0$ there exists a $B < \infty$ such that for all large T,* $\mathbb{P}\left[Tv_T^2\left|\widehat{\lambda}_b^{\mathrm{GL}} - \lambda_b^0\right| > B\right] < \varepsilon$.

Proof. See Lemma A.29 which proves a stronger result needed for Theorem 3.2. □

Parts (i) and (ii) of Proposition 3.1 follow from Lemma A.9 and Lemma A.10, respectively.

A.3.2 Proof of Theorem 3.1

We start with the following lemmas.

Lemma A.11. *For any $a \in \mathbb{R}$, $|c| \leq 1$, and integer $i \geq 0$,* $\left|\exp\left(ca\right) - \sum_{j=0}^{i}\left(ca\right)^j/j!\right| \leq |c|^{i+1}\exp\left(|a|\right)$.

Proof. The proof is immediate and the same as the one in Jun, Pinkse, and Wan (2015). Using simple manipulations,

$$\left|\exp(ca) - \sum_{j=0}^{i} (ca)^j / j!\right| \leq \left|\sum_{j=i+1}^{\infty} \frac{(ca)^j}{j!}\right| \leq |c|^{i+1} \left|\sum_{j=i+1}^{\infty} \frac{(a)^j}{j!}\right| \leq |c|^{i+1} \exp(|a|).$$

□

Lemma A.12. *$\widetilde{G}_{T,v}(u, \widetilde{v}) \Rightarrow \mathscr{W}(u)$ in $\mathbb{D}_b(\mathbf{C} \times \mathbf{V})$, where $\mathbf{C} \subset \mathbb{R}$ and $\mathbf{V} \subset \mathbb{R}^{p+2q}$ are both compact sets, and*

$$\mathscr{W}(u) \triangleq \begin{cases} 2\left(\left(\delta^0\right)' \Sigma_1 \delta^0\right)^{1/2} W_1(-u), & \textit{if } u < 0 \\ 2\left(\left(\delta^0\right)' \Sigma_2 \delta^0\right)^{1/2} W_2(u), & \textit{if } u \geq 0. \end{cases}$$

Proof. Consider $u < 0$. According to the expansion of the criterion function given in Lemma A.2, for any $(u, \widetilde{v}) \in \mathbf{C}\times\mathbf{V}$, $\widetilde{G}_{T,v}(u, \widetilde{v})$ satisfies $2\text{sgn}\left(T_b^0 - T_b(u)\right) \delta_T' Z_\Delta' e + o_{\mathbb{P}}(1)$. Then, $\delta_T' Z_\Delta' e = \left(\delta^0\right)' v_T \sum_{t=\lfloor u/v_T^2 \rfloor}^{T_b^0} z_t e_t \Rightarrow \left(\delta^0\right)' \mathscr{G}_1(-u)$, where $\mathscr{G}_1$ is a multivariate Gaussian process. In particular, $\left(\delta^0\right)' \mathscr{G}_1(-u)$ is equivalent in law to $\left(\left(\delta^0\right)' \Sigma_1 \delta^0\right)^{1/2} W_1(-u)$, where $W_1(\cdot)$ is a standard Wiener process on $[0, \infty)$. Similarly, for $u \geq 0$, $\delta_T' Z_\Delta' e \Rightarrow \left(\left(\delta^0\right)' \Sigma_2 \delta^0\right)^{1/2} W_2(u)$, where $W_2(\cdot)$ is another standard Wiener process on $[0, \infty)$ which is independent of W_1. Hence, $\widetilde{G}_{T,v}(u, \widetilde{v}) \Rightarrow \mathscr{W}(u)$ in $\mathbb{D}_b(\mathbf{C} \times \mathbf{V})$. □

Lemma A.13. *Fix any $a > 0$ and let $\varpi \in (1/2, 1]$. (i) For any $\nu > 0$ and any $\varepsilon > 0$,*

$$\limsup_{T\to\infty} \mathbb{P}\left[\sup_{u\in\Gamma_{T,\psi}^c} \left\{\overline{G}_{T,v}(u) - a\left\|\delta^0\right\|^2 |u|^{\varpi}\right\} > \nu\right] < \varepsilon.$$

(ii) For $\widetilde{u} \in \mathbb{R}_+$ let $\widetilde{\Gamma} \triangleq \{u \in \mathbb{R} : |u| > \widetilde{u}\}$. Then, for every $\epsilon > 0$,

$$\lim_{\widetilde{u}\to\infty} \lim_{T\to\infty} \mathbb{P}\left[\sup_{u\in\widetilde{\Gamma}} \left\{\overline{G}_{T,v}(u) - a\left\|\delta^0\right\|^2 |u|^{\varpi}\right\} > \epsilon\right] = 0.$$

Proof. We begin with part (i). Upon using Lemma A.12 and the continuous mapping theorem, with any nonnegative integer i,

$$\begin{aligned}
\limsup_{T\to\infty} \mathbb{P}\left[\sup_{u\in\Gamma_{T,\psi}^c} \left\{\overline{G}_{T,v}(u) - a\left\|\delta^0\right\|^2 |u|^{\varpi}\right\} > \nu\right] &\leq \lim_{T\to\infty} \mathbb{P}\left[\sup_{|u|>\overline{u}} \left\{\overline{G}_{T,v}(u) - a\left\|\delta^0\right\| |u|^{\varpi}\right\} > \nu\right] \\
&\leq \lim_{T\to\infty} \mathbb{P}\left[\sup_{|u|\geq i} \left\{\overline{G}_{T,v}(u) > a\left\|\delta^0\right\| |u|^{\varpi}\right\} > \nu\right] \\
&\leq \mathbb{P}\left[\sup_{|u|\geq i} \left\{|\mathscr{W}(u)| - a\left\|\delta^0\right\| |u|^{\varpi}\right\} > \nu\right] \\
&\leq \sum_{r=i+1}^{\infty} \mathbb{P}\left[\sup_{r-1\leq|u|<r} \left\{|\mathscr{W}(u)| - a\left\|\delta^0\right\| |u|^{\varpi}\right\} > \nu\right].
\end{aligned}$$

Then,

$$\begin{aligned}
&\sum_{r=i+1}^{\infty} \mathbb{P}\left[\sup_{r-1\leq |u|<r} \frac{1}{\sqrt{r}} \left|\mathscr{W}\left(u\right)\right| > \inf_{r-1<|u|<r} a\frac{1}{\sqrt{r}} \left\|\delta^0\right\| |u|^{\varpi}\right] \\
&\quad = \sum_{r=i+1}^{\infty} \mathbb{P}\left[\sup_{1-1/r\leq |u|/r\leq 1} \left|\mathscr{W}\left(u/r\right)\right| > \inf_{1-1/r<|u|/r\leq 1} a\left(\frac{r}{r}\right)^{\varpi-1/2} \frac{|u|^{\varpi}}{\sqrt{r}} \left\|\delta^0\right\|\right] \\
&\quad = \sum_{r=i+1}^{\infty} \mathbb{P}\left[\sup_{1-1/r<s\leq 1} \left|\mathscr{W}\left(s\right)\right| > \inf_{c<s\leq 1} ar^{\varpi-1/2}s^{\varpi} \left\|\delta^0\right\|\right] \\
&\quad = \sum_{r=i+1}^{\infty} \mathbb{P}\left[\sup_{s\leq 1} \left|\mathscr{W}\left(s\right)\right| > r^{\varpi-1/2}c^{\varpi}C \left\|\delta^0\right\|\right], \qquad \text{(A.19)}
\end{aligned}$$

where $0<c\leq 1$. By Markov's inequality,

$$\sum_{r=i+1}^{\infty} \mathbb{P}\left[\sup_{c<s\leq 1} \left|\mathscr{W}\left(s\right)\right|^4 > C^4 \left\|\delta^0\right\|^4 r^{4(\varpi-1/2)}c^{4\varpi}\right] \leq \frac{C}{\left\|\delta^0\right\|^4} \frac{\mathbb{E}\left(\sup_{s\leq 1}\left|\mathscr{W}\left(s\right)\right|^4\right)}{c^{4\varpi}} \sum_{r=i+1}^{\infty} r^{-(4\varpi-2)}. \qquad \text{(A.20)}$$

By Proposition A.2.4 in van der Vaart and Wellner (1996), $\mathbb{E}(\sup_{s\leq 1}|\mathscr{W}(s)|^4) \leq C\mathbb{E}\left(\sup_{s\leq 1}|\mathscr{W}(s)|\right)^4$ for some $C<\infty$, which is finite by Corollary 2.2.8 in van der Vaart and Wellner (1996). Choose K (thus $\overline{u}$) large enough such that the right-hand side in (A.20) can be made arbitrarily smaller than $\varepsilon>0$. The proof of the second part is similar and omitted. □

Lemma A.14. *Fix any $a>0$. For any $\varepsilon>0$ there exists a $C<\infty$ such that*

$$\mathbb{P}\left[\sup_{u\in\mathbb{R}}\left\{\overline{G}_{T,v}\left(u\right) - a\left\|\delta^0\right\|^2|u|\right\} > C\right] < \varepsilon, \qquad \text{for all } T.$$

Proof. For any finite T, $\overline{G}_{T,v}\left(u\right)\in\mathbb{D}_b$ by definition. As for the limiting case, fix any $0<\overline{u}<\infty$,

$$\begin{aligned}
\limsup_{T\to\infty} \mathbb{P}\left[\sup_{u\in\mathbb{R}}\left\{\overline{G}_{T,v}\left(u\right) - a\left\|\delta^0\right\|^2|u|\right\} > C\right] &\leq \limsup_{T\to\infty}\mathbb{P}\left[\sup_{|u|\leq\overline{u}} \overline{G}_{T,v}\left(u\right) > C\right] \\
&\quad + \limsup_{T\to\infty}\mathbb{P}\left[\sup_{|u|>\overline{u}} \overline{G}_{T,v}\left(u\right) > a\left\|\delta^0\right\|^2\overline{u}\right].
\end{aligned}$$

The second term converges to zero letting $\overline{u}\to\infty$ from Lemma A.13-(ii). For the first term, let $C\to\infty$, use the continuous mapping theorem and Lemma A.12 to deduce that it converges to zero by the properties of $\mathscr{W}\in\mathbb{D}_b$. □

Lemma A.15. *Let*

$$\begin{aligned}
A_1\left(u,\,\widetilde{v}\right) &= u^m\pi_{T,v}\left(u\right)\exp\left(\widetilde{\gamma}_T\widetilde{G}_{T,v}\left(u,\,\widetilde{v}\right)+Q_{T,v}\left(u\right)\right), \qquad \text{(A.21)} \\
A_2\left(u,\,\widetilde{v}\right) &= u^m\pi^0\exp\left(\widetilde{\gamma}_T\widetilde{G}_{T,v}\left(u,\,\widetilde{v}\right)-\Lambda_0\left(u\right)\right).
\end{aligned}$$

For $m\geq 0$,

$$\liminf_{T\to\infty}\mathbb{P}\left[\sup_{\widetilde{v}\in\mathbf{V}}\left|\int_{\Gamma^c_{T,\psi}}\left(A_1\left(u,\,\widetilde{v}\right)-A_2\left(u,\,\widetilde{v}\right)\right)\right|<\epsilon\right]\geq 1-\epsilon.$$

Proof. We consider each integrand $A_i(u, \widetilde{v})$ $(i = 1, 2)$ separately on $\Gamma^c_{T,\psi}$. Let us consider A_1 first. Lemma A.4 yields that whenever $\widetilde{\gamma}_T \to \kappa_\gamma < \infty$, $A_1(u, \widetilde{v}) \leq C_1 \exp(-C_2\nu_T)$ where $0 < C_1, C_2 < \infty$ and ν_T is a divergent sequence. Note that the number C_1 follows from Assumption 3.2 (cf. $\pi(\cdot) < \infty$). The argument for $A_2(u, \widetilde{v})$ relies on Lemma A.13-(i), which shows that $G_{T,v}(u, \widetilde{v})$ is always less than $C|u|^{\varpi}$ uniformly on $\Gamma^c_{T,\psi}$, with $C > 0$ and $\varpi \in (1/2, 1)$. Thus, $A_2(u, \widetilde{v}) = o_{\mathbb{P}}(1)$ uniformly on $\mathbf{V}$. $\square$

Let $\Gamma_{T,K} \triangleq \{u \in \mathbb{R} : |u| < K, K > 0\}$, and $\Gamma_{T,\eta} \triangleq \{u \in \mathbb{R} : K \leq |u| \leq \eta\psi_T, K, \eta > 0\}$.

Lemma A.16. *For any polynomial function $p \in \boldsymbol{P}$ and any $C < \infty$, let*

$$D_T \triangleq \sup_{\widetilde{v}\in\mathbf{W}} \int_{\Gamma_{T,K}} |p(u)| \exp\left\{C\widetilde{G}_{T,v}(u, \widetilde{v})\right\} |\exp(R_{T,v}(u)) - 1| \exp\left(-\Lambda^0(u)\right) du = o_{\mathbb{P}}(1).$$

Proof. Let $0 < \epsilon < 1$. We shall use Lemma A.11 with $i = 0$, $a = R_{T,v}(u)/c$, and $c = \epsilon$ to deduce that $D_T = O_{\mathbb{P}}(\epsilon)$ and then let $\epsilon \to 0$. Note that

$$\epsilon^{-1} D_T \leq C \int_{\Gamma_{T,K}} |p(u)| \exp\left(C\overline{G}_{T,v}(u, \widetilde{v}) + \left|\epsilon^{-1} R_{T,v}(u)\right| - \Lambda^0(u)\right) du.$$

By definition, $K \geq u = \|\delta_T\|^2 \left(T_b - T_b^0\right)$ on $\Gamma_{T,K}$. By Lemma A.2-A.3, on $\Gamma_{T,K}$ we have $R_{T,v}(u) = O_{\mathbb{P}}\left(\|\delta_T\|^2\right)$ for each u. Thus, for large enough T, the right-hand side above is $O_{\mathbb{P}}(1)$ and does not depend on ϵ. Thus, $D_T = \epsilon O_{\mathbb{P}}(1)$. The claim of the lemma follows by letting ϵ approach zero. $\square$

Lemma A.17. *For $p \in \boldsymbol{P}$,*

$$D_{2,T} \triangleq \sup_{\widetilde{v}\in\mathbf{V}} \int_{\Gamma_{T,\eta}} |p(u)| \exp\left\{\widetilde{\gamma}_T \widetilde{G}_{T,v}(u, \widetilde{v})\right\} \exp\left(-\Lambda^0(u)\right) \left|\pi_{T,v}(u) - \pi^0\right| du = o_{\mathbb{P}}(1).$$

Proof. By the differentiability of $\pi(\cdot)$ at λ_b^0 (cf. Assumption 3.2), for any $u \in \mathbb{R}$ $\left|\pi_{T,v}(u) - \pi^0\right| \leq \left|\pi\left(\lambda_{b,T}^0(v)\right) - \pi^0\right| + C\psi_T^{-1}|u|$, with $C > 0$. The first term on the right-hand side is $o(1)$ and does not depend on u. Recalling that $\overline{G}_{T,v}(u, \widetilde{v}) = \sup_{\widetilde{v}\in\mathbf{V}} \left|\widetilde{G}_{T,v}(u, \widetilde{v})\right|$,

$$D_{2,T} \leq K\left[o(1) \int_{\Gamma_{T,\eta}} d_T(u)\, du + \psi_T^{-1} \int_{\Gamma_{T,\eta}} |u|\, d_T(u)\, du\right] \leq K\left[o(1)\, O_{\mathbb{P}}(1) + \psi_T^{-1} O_{\mathbb{P}}(1)\right],$$

where $d_T(u) \triangleq |p(u)| \exp\left\{\widetilde{\gamma}_T \overline{G}_{T,v}(u, \widetilde{v})\right\} \left|\exp\left(-\Lambda^0(u)\right)\right|$ and the $O_{\mathbb{P}}(1)$ terms follows from Lemma A.14 and $\widetilde{\gamma}_T \to \kappa_\gamma < \infty$. Since $\psi_T \to \infty$, we have $D_{2,T} = o_{\mathbb{P}}(1)$. $\square$

Lemma A.18. *For any $p \in \boldsymbol{P}$ and constants $C_1, C_2 > 0$, $\int_{\Gamma^c_{T,\psi}} |p(u)| \exp\left(C_1 \overline{G}_T(u) - C_2|u|\right) du = o_{\mathbb{P}}(1)$.*

Proof. It follows from Lemma A.13. $\square$

Lemma A.19. *For $p \in \boldsymbol{P}$ and constants $a_1, a_2, a_3 \geq 0$, with $a_2 + a_3 > 0$, let*

$$D_{3,T} \triangleq \int_{\widetilde{\mathbf{U}}^c_T} |p(u)| \exp\left(\widetilde{\gamma}_T \left\{a_1 \overline{G}_{T,v}(u) + a_2 Q_{T,v}(u) - a_3 \Lambda^0(u)\right\}\right) du = o_{\mathbb{P}}(1).$$

Proof. It follows from Lemma A.6. $\square$

Lemma A.20. *For any integer* $m \geq 0$,

$$\sup_{\widetilde{v}\in\mathbf{V}}\left|\int_{\mathbb{R}} u^{m}\exp\left(\widetilde{\gamma}_{T}\widetilde{G}_{T,v}\left(u,\,\widetilde{v}\right)\right)\left[\pi_{T,v}\left(u\right)\exp\left(Q_{T,v}\left(u\right)\right)-\pi^{0}\exp\left(-\Lambda^{0}\left(u\right)\right)\right]du\right|$$
$$=\sup_{\widetilde{v}\in\mathbf{V}}\left|\int_{\mathbb{R}}\left(A_{1}\left(u,\,\widetilde{v}\right)-A_{2}\left(u,\,\widetilde{v}\right)\right)du\right|$$
$$=o_{\mathbb{P}}\left(1\right).$$

Proof. By Assumption 3.2, $A_1\left(u,\,\widetilde{v}\right)=0$ for $u\in\Gamma_{T,\psi}^{c}-\widetilde{\mathbf{U}}_{T}^{c}$. Then, omitting arguments, we can write,

$$\sup\left|\int_{\mathbb{R}}\left(A_{1}-A_{2}\right)\right|\leq\sup\left|\int_{\Gamma_{T,\psi}}\left(A_{1}-A_{2}\right)\right|+\sup\left|\int_{\Gamma_{T,\psi}^{c}}A_{2}\right|+\sup\left|\int_{\widetilde{\mathbf{U}}_{T}^{c}}A_{1}\right|. \tag{A.22}$$

The first right-hand side term above converges in probability to zero by Lemma A.16-A.17. The second and the last term are each $o_{\mathbb{P}}\left(1\right)$ by, receptively, Lemma A.18 and Lemma A.19. □

We are now in a position to conclude the proof of Theorem 3.1.

Proof. Let $\mathbf{V}\subset\mathbb{R}^{p+2q}$ be a compact set. From (3.11),

$$\psi_{T}\left(\widehat{\lambda}_{b}^{\mathrm{GL},*}\left(\widetilde{v},\,v\right)-\lambda_{b,T}^{0}\left(v\right)\right)=\frac{\int_{\mathbb{R}}u\exp\left(\widetilde{\gamma}_{T}\left[\widetilde{G}_{T,v}\left(u,\,\widetilde{v}\right)+Q_{T,v}\left(u\right)\right]\right)\pi_{T,v}\left(u\right)du}{\int_{\mathbb{R}}\exp\left(\widetilde{\gamma}_{T}\left[\widetilde{G}_{T,v}\left(u,\,\widetilde{v}\right)+Q_{T,v}\left(u\right)\right]\right)\pi_{T,v}\left(u\right)du}.$$

For a large enough T, by Lemma A.20 the right-hand is uniformly in $\widetilde{v}\in\mathbf{V}$ equal to

$$\frac{\int_{\mathbb{R}}u\exp\left(\widetilde{\gamma}_{T}\widetilde{G}_{T,v}\left(u,\,\widetilde{v}\right)\right)\exp\left(-\Lambda^{0}\left(u\right)\right)du}{\int_{\mathbb{R}}\exp\left(\widetilde{\gamma}_{T}\widetilde{G}_{T,v}\left(u,\,\widetilde{v}\right)\right)\exp\left(-\Lambda^{0}\left(u\right)\right)du}+o_{\mathbb{P}}\left(1\right).$$

The first term is integrable with large probability by Lemma A.13-A.14. Thus, by Lemma A.12 and the continuous mapping theorem, we have for each $v\in\mathbf{V}$,

$$T\left\Vert \delta_{T}\right\Vert ^{2}\left(\widehat{\lambda}_{b}^{\mathrm{GL},*}\left(\widetilde{v},\,v\right)-\lambda_{b,T}^{0}\left(v\right)\right)\Rightarrow\frac{\int_{\mathbb{R}}u\exp\left(\mathscr{W}\left(u\right)\right)\exp\left(-\Lambda^{0}\left(u\right)\right)du}{\int_{\mathbb{R}}\exp\left(\mathscr{W}\left(u\right)\right)\exp\left(-\Lambda^{0}\left(u\right)\right)du}. \tag{A.23}$$

Note that $\partial_{\theta}Q_{T}^{0}\left(\theta,\,\cdot\right)$ is monotonic and bounded for all $\theta\in\mathbf{S}$. The argument of Theorem 4.1 in Jurečová (1977) can be used in (A.23) to achieve uniformity in v. □

A.3.3 Proof of Proposition 3.2

We first need to introduce further notation. For a scalar $\overline{u}>0$ define $\Gamma_{\overline{u}}\triangleq\left\{ u:\in\mathbb{R}:\,\left|u\right|\leq\overline{u}\right\}$. Note that $\widetilde{\gamma}_{T}^{-1}=o\left(1\right)$. We shall be concerned with the asymptotic properties of the following statistic:

$$\xi_{T}\left(\widetilde{v}\right)=\frac{\int_{\Gamma_{\overline{u}}}u\exp\left(\widetilde{\gamma}_{T}\left(\widetilde{G}_{T,v}\left(u,\,\widetilde{v}\right)+Q_{T,v}\left(u\right)\right)\right)\pi_{T,v}\left(u\right)du}{\int_{\Gamma_{\overline{u}}}\exp\left(\widetilde{\gamma}_{T}\left(\widetilde{G}_{T,v}\left(u,\,\widetilde{v}\right)+Q_{T,v}\left(u\right)\right)\right)\pi_{T,v}\left(u\right)du}.$$

Furthermore, for every $\widetilde{v}\in\mathbf{V}$, let $\xi_{0}\left(\widetilde{v}\right)=\arg\max_{u\in\Gamma_{\overline{u}}}\mathscr{V}\left(u\right)$. It turns out that $\xi_{0}\left(\widetilde{v}\right)$ is flat in $\widetilde{v}$ and thus we write $\xi_{0}=\xi_{0}\left(\widetilde{v}\right)$. Finally, recall that $u=T\left\Vert \delta_{T}\right\Vert ^{2}\left(\lambda_{b}-\lambda_{b,T}^{0}\left(v\right)\right)$.

Lemma A.21. *Let $\Gamma^c_{T,\overline{u}} = \mathbf{U}_T - \Gamma_{\overline{u}}$. Then for any $\epsilon > 0$ and $m = 0, 1$,*

$$\lim_{\overline{u}\to\infty}\lim_{T\to\infty}\mathbb{P}\left(\frac{\sup_{\widetilde{v}\in\mathbf{V}}\int_{\Gamma^c_{T,\overline{u}}}|u|^m\exp\left(\widetilde{\gamma}_T\left(\widetilde{G}_{T,v}(u,\,\widetilde{v})+Q_{T,v}(u)\right)\right)\pi_{T,v}(u)\,du}{\sup_{\widetilde{v}\in\mathbf{V}}\int_{\mathbb{R}}\exp\left(\widetilde{\gamma}_T\left(\widetilde{G}_{T,v}(u,\,\widetilde{v})+Q_{T,v}(u)\right)\right)\pi_{T,v}(u)\,du}>\epsilon\right)=0.$$

Proof. Let J_1 and J_2 denote the numerator and denominator, respectively, in the display of the lemma. Then,

$$\mathbb{P}(J_1/J_2>\epsilon)\leq\mathbb{P}(J_2\leq\exp(-\overline{a}\widetilde{\gamma}_T))+\mathbb{P}(J_1>\epsilon\exp(-\overline{a}\widetilde{\gamma}_T)),\tag{A.24}$$

for any constant $\overline{a} > 0$. Let us consider the second term term in (A.24). For an arbitrary $a > 0$, let $\mathbf{H}(\overline{u},\,a) = \left\{u\in\Gamma^c_{T,\overline{u}} : \sup_{\widetilde{v}\in\mathbf{V}}\left|\widetilde{G}_{T,v}(u,\,\widetilde{v})\right|\leq a|u|\right\}$. Let $\overline{\lambda} = 2\sup_{\lambda_b\in\varGamma^0}|\lambda_b|$. Note that $\overline{\lambda} < 2$ and $\sup_{u\in\mathbf{H}(\overline{u},\,a)}|u|\leq\overline{\lambda}T\|\delta_T\|^2$. By Assumption 2.4 and 3.4, and Lemma A.6, $Q_{T,v}(u)\leq-\min\left(\varLambda^0(u)/2,\,\eta\overline{\lambda}\|\delta_T\|^2T\right)$ uniformly for all large T where $\eta > 0$. Thus,

$$\begin{aligned}\sup_{u\in\mathbf{H}(\overline{u},\,a)}&\sup_{\widetilde{v}\in\mathbf{V}}\exp\left(\widetilde{\gamma}_T\left[\widetilde{G}_{T,v}(u,\,\widetilde{v})+Q_{T,v}(u)\right]\right)\\&\leq\sup_{u\in\mathbf{H}(\overline{u},\,a)}\sup_{\widetilde{v}\in\mathbf{V}}\exp\left(\widetilde{\gamma}_T\left[a|u|-\varLambda^0(u)/4+\left[\varLambda^0(u)/2+Q_{T,v}(u)\right]\right]\right)\\&\leq\sup_{u\in\mathbf{H}(\overline{u},\,a)}\exp\left(\widetilde{\gamma}_T\left[a|u|-\varLambda^0(u)-\min\left(\varLambda^0(u)/4,\,\varLambda^0(u)/4+\eta\|\delta_T\|^2T\right)\right]\right)\\&\leq\sup_{u\in\mathbf{H}(\overline{u},\,c)}\exp(\widetilde{\gamma}_T[a|u|-C_2|u|])+\exp\left(\gamma_T\left[a\overline{\lambda}-\eta C\right]\right)\\&\leq\sup_{u\in\mathbf{H}(\overline{u},\,c)}\exp(\gamma_T[a-C_2])+\exp\left(\gamma_T\left[a\overline{\lambda}-\eta C\right]\right)=o(\exp(-\gamma_T\overline{a}_1)),\end{aligned}\tag{A.25}$$

when $a > 0$ is chosen sufficiently small and for some $\overline{a}_1 > 0$. Furthermore, by Lemma A.13-(ii) below with $\varpi = 1$,

$$\lim_{\overline{u}\to\infty}\lim_{T\to\infty}\mathbb{P}\left(u\in\left\{\Gamma^c_{T,\overline{u}}-\mathbf{H}(\overline{u},\,c)\right\}\right)\leq\lim_{\overline{u}\to\infty}\lim_{T\to\infty}\mathbb{P}\left(\sup_{|u|>\overline{u}}\frac{\widetilde{G}_{T,v}(u,\,\widetilde{v})}{|u|}>a\right)=0.\tag{A.26}$$

By combining (A.25)-(A.26), $\mathbb{P}(J_1>\epsilon\exp(-\overline{a}\widetilde{\gamma}_T))\to 0$ as $T\to\infty$. Next, we consider the first right-hand side term in (A.24). Recall the definition of λ_+ from Assumption 3.5 and let $0 < b\leq\overline{a}/4\lambda_+$. Note that for $G_{T,v}(b)\triangleq\sup_{|u|\leq b}\sup_{\widetilde{v}\in\mathbf{V}}\left|\widetilde{G}_{T,v}(u,\,\widetilde{v})\right|$,

$$\mathbb{P}(J_2\leq\exp(-\overline{a}\widetilde{\gamma}_T))\leq\mathbb{P}(G_{T,v}(b)\leq\overline{a},\,J_2\leq\exp(-\overline{a}\widetilde{\gamma}_T))+\mathbb{P}(G_{T,v}(b)>\overline{a}).\tag{A.27}$$

Under Assumption 3.2 and the second part of Assumption 3.5, using the definition of b,

$$\begin{aligned}\mathbb{P}(G_{T,v}(b)\leq\overline{a},\,J_2\leq\exp(-\overline{a}\widetilde{\gamma}_T))&\leq\mathbb{P}\left(C_\pi\int_{|u|\leq b}\exp(\widetilde{\gamma}_T(-\overline{a}/2-\lambda_+b))\,du\leq\exp(-\overline{a}\widetilde{\gamma}_T)\right)\\&\leq\mathbb{P}(C_\pi b\exp(\overline{a}\widetilde{\gamma}_T/2)\leq 1)\to 0,\end{aligned}$$

as $T\to\infty$. We shall use the uniform convergence in Lemma A.12 for the second right-hand side term in (A.27) to deduce that (recall that $\overline{a}$ was chosen sufficiently small and $b\leq\overline{a}/4\lambda_+$),

$$\lim_{b\to 0}\lim_{T\to\infty}\mathbb{P}(G_{T,v}(b)>\overline{a})\leq\lim_{b\to 0}\mathbb{P}\left(\sup_{|u|\leq b}|\mathscr{W}(u)|>\overline{a}\right)=0.$$

□

Lemma A.22. *As* $T \to \infty$, $\xi_T(\widetilde{v}) \Rightarrow \xi_0$ *in* $\mathbb{D}_b(\mathbf{V})$.

Proof. Let $\mathbf{B} = \Gamma_{\overline{u}} \times \mathbf{V}$. For any fixed $\overline{u}$, Lemma A.12 and the result $\sup_{(u,\widetilde{v}) \in \mathbf{B}} |Q_{T,v}(u) - \Lambda^0(u)| = o_{\mathbb{P}}(1)$ (cf. Lemma A.3), imply that $\overline{Q}_T \Rightarrow \mathscr{V}$ in $\mathbb{D}_b(\mathbf{B})$. By the Skorokhod representation theorem [cf. Theorem 6.4 in Billingsley (1999)] we can find a probability space $\left(\widetilde{\Omega}, \widetilde{\mathscr{F}}, \widetilde{\mathbb{P}}\right)$ on which there exist processes $\widetilde{Q}_T(u, \widetilde{v})$ and $\widetilde{\mathscr{V}}(u)$ which have the same law as $\overline{Q}_T(u, \widetilde{v})$ and $\mathscr{V}(u)$, respectively, and with the property that

$$\sup_{(u,\widetilde{v}) \in \mathbf{B}} \left| \widetilde{Q}_T(u, \widetilde{v}) - \widetilde{\mathscr{V}}(u) \right| \to 0 \qquad \widetilde{\mathbb{P}} - \text{a.s.} \tag{A.28}$$

Let

$$\widetilde{\xi}_T(\widetilde{v}) \triangleq \frac{\int_{\Gamma_{\overline{u}}} u \exp\left(\widetilde{\gamma}_T \widetilde{Q}_{T,v}(u, \widetilde{v})\right) \pi_{T,v}(u) \, du}{\int_{\Gamma_{\overline{u}}} \exp\left(\widetilde{\gamma}_T \widetilde{Q}_{T,v}(u, \widetilde{v})\right) \pi_{T,v}(u) \, du},$$

and $\widetilde{\xi}_0 \triangleq \arg\max_{u \in \Gamma_{\overline{u}}} \widetilde{\mathscr{V}}(u)$. We shall rely on (A.28) to establish that

$$\sup_{\widetilde{v} \in \mathbf{V}} \left| \widetilde{\xi}_T(\widetilde{v}) - \widetilde{\xi}_0 \right| \to 0 \qquad \widetilde{\mathbb{P}} - \text{a.s.}. \tag{A.29}$$

Let us indicate any pair of sample paths of $\widetilde{Q}_T(u, \widetilde{v})$ and $\widetilde{\mathscr{V}}$, for which (A.28) holds with a superscript ω, by $\widetilde{Q}^{\omega}_{T,v}$ and $\widetilde{\mathscr{V}}^{\omega}$, respectively. For arbitrary sets $\mathbf{S}_1, \mathbf{S}_2 \subset \mathbf{B}$, let $\widetilde{\rho}(\mathbf{S}_1, \mathbf{S}_2) \triangleq \text{Leb}(\mathbf{S}_1 - \mathbf{S}_2) + \text{Leb}(\mathbf{S}_2 - \mathbf{S}_1)$ where $\text{Leb}(\mathbf{A})$ is the Lebesgue measure of the set $\mathbf{A}$. Further, for an arbitrary scalar $c > 0$ and function $\Upsilon : \mathbf{B} \to \mathbb{R}$, define $\mathbf{S}(\Upsilon, c) \triangleq \left\{(u, \widetilde{v}) \in \mathbf{B} : \left|\Upsilon(u, \widetilde{v}) - \widetilde{\mathscr{V}}_{\text{M}}\right| \leq c\right\}$ where $\widetilde{\mathscr{V}}_{\text{M}} \triangleq \max_{u \in \Gamma_{\overline{u}}} \widetilde{\mathscr{V}}^{\omega}(u)$. The first step is to show that

$$\widetilde{\rho}\left(\mathbf{S}\left(\widetilde{Q}^{\omega}_{T,v}, c\right), \mathbf{S}\left(\widetilde{\mathscr{V}}^{\omega}, c\right)\right) = o(1). \tag{A.30}$$

Let $\mathbf{S}_{1,T}(c) = \mathbf{S}\left(\widetilde{Q}^{\omega}_{T,v}, c\right) - \mathbf{S}\left(\widetilde{\mathscr{V}}^{\omega}, c\right)$ and $\mathbf{S}_{2,T}(c) = \mathbf{S}\left(\widetilde{\mathscr{V}}^{\omega}, c\right) - \mathbf{S}\left(\widetilde{Q}^{\omega}_{T,v}, c\right)$. We first establish that $\text{Leb}(\mathbf{S}_{2,T}(c)) = o(1)$. For an arbitrary $\overline{c} > 0$, define the set $\widetilde{\mathbf{S}}_T(\overline{c}) \triangleq \left\{(u, \widetilde{v}) \in \mathbf{B} : \left|\widetilde{Q}^{\omega}_{T,v}(u, \widetilde{v}) - \widetilde{\mathscr{V}}^{\omega}(u)\right| \leq \overline{c}\right\}$ and its complement (relative to $\mathbf{B}$) $\widetilde{\mathbf{S}}^c_T(\overline{c}) \triangleq \left\{(u, \widetilde{v}) \in \mathbf{B} : \left|\widetilde{Q}^{\omega}_{T,v}(u, \widetilde{v}) - \widetilde{\mathscr{V}}^{\omega}(u)\right| > \overline{c}\right\}$. We have

$$\begin{aligned}
\text{Leb}(\mathbf{S}_{2,T}(c)) &= \text{Leb}\left(\mathbf{S}_{2,T}(c) \cap \widetilde{\mathbf{S}}_T(\overline{c})\right) + \text{Leb}\left(\mathbf{S}_{2,T}(c) \cap \widetilde{\mathbf{S}}^c_T(\overline{c})\right) \\
&\leq \text{Leb}\left(\mathbf{S}_{2,T}(c) \cap \widetilde{\mathbf{S}}_T(\overline{c})\right) + \text{Leb}\left(\widetilde{\mathbf{S}}^c_T(\overline{c})\right).
\end{aligned}$$

Note that $\text{Leb}\left(\widetilde{\mathbf{S}}^c_T(\overline{c})\right) = o(1)$ since the path ω satisfies (A.28). Furthermore, $\mathbf{S}_{2,T}(c) \cap \widetilde{\mathbf{S}}_T(\overline{c}) \subset \mathbf{C}_T(c, \overline{c})$ where $\mathbf{C}_T(c, \overline{c}) \triangleq \left\{(u, \widetilde{v}) \in \mathbf{B} : c \leq \left|\widetilde{Q}^{\omega}_{T,v}(u, \widetilde{v}) - \widetilde{\mathscr{V}}_{\text{M}}\right| \leq c + \overline{c}\right\}$. In view of (A.28),

$$\begin{aligned}
\lim_{\overline{c} \downarrow 0} \lim_{T \to \infty} \text{Leb}(\mathbf{C}_T(c, \overline{c})) &= \lim_{\overline{c} \downarrow 0} \text{Leb}\left\{(u, \widetilde{v}) \in \mathbf{B} : c \leq \left|\widetilde{\mathscr{V}}^{\omega}(u) - \widetilde{\mathscr{V}}_{\text{M}}\right| \leq c + \overline{c}\right\} \\
&= \text{Leb}\left\{(u, \widetilde{v}) \in \mathbf{B} : \left|\widetilde{\mathscr{V}}^{\omega}(u) - \widetilde{\mathscr{V}}_{\text{M}}\right| = c\right\} = 0,
\end{aligned}$$

by the path properties of $\widetilde{\mathscr{V}}^{\omega}$. Since $\text{Leb}(\mathbf{S}_{1,T}(c)) = o(1)$ can be proven in a similar fashion, (A.30) holds.

For $m = 0,\, 1$, $C_1 < \infty$ and by Assumption 3.2 we know there exists some $C_2 < \infty$ such that

$$\sup_{\widetilde{v}\in\mathbf{V}} \int_{\mathbf{S}^c\left(\widetilde{Q}^\omega_{T,v}(u,\,\widetilde{v}),\,c\right)} |u|^m \exp\left(\widetilde{\gamma}_T\left(\widetilde{Q}^\omega_{T,v}\left(u,\,\widetilde{v}\right) - \mathscr{V}_{\mathrm{M}}\right)\right)\pi_{T,v}\left(u\right)du \leq C_1 \exp\left(-c\widetilde{\gamma}_T\right) C_2 \int_{\Gamma_{\overline{u}}} |u|^m\, du = o\left(1\right),$$

since $\{u \leq \overline{u}\}$ on $\Gamma_{\overline{u}}$ and recalling that $\widetilde{\gamma}_T \to \infty$. This gives an upper bound to the same function where u replaces $|u|$. Then,

$$\sup_{\widetilde{v}\in\mathbf{V}} \frac{\int_{\Gamma_{\overline{u}}} u \exp\left(\widetilde{\gamma}_T \widetilde{Q}^\omega_{T,v}\left(u,\,\widetilde{v}\right)\right)\pi_{T,v}\left(u\right)du}{\int_{\Gamma_{\overline{u}}} \exp\left(\widetilde{\gamma}_T \widetilde{Q}^\omega_{T,v}\left(u,\,\widetilde{v}\right)\right)\pi_{T,v}\left(u\right)du} \leq \operatorname{ess\,sup} \mathbf{S}\left(\widetilde{Q}^\omega_{T,v},\,c\right) + o\left(1\right).$$

By (A.28) we deduce $\operatorname{ess\,sup} \mathbf{S}\left(\widetilde{Q}^\omega_{T,v},\,c\right) + o\left(1\right) = \operatorname{ess\,sup} \mathbf{S}\left(\widetilde{\mathscr{V}}^\omega,\,c\right) + o\left(1\right)$. The same argument yields

$$\inf_{\widetilde{v}\in\mathbf{V}} \frac{\int_{\Gamma_{\overline{u}}} u \exp\left(\widetilde{\gamma}_T \widetilde{Q}^\omega_{T,v}\left(u,\,\widetilde{v}\right)\right)\pi_{T,v}\left(u\right)du}{\int_{\Gamma_{\overline{u}}} \exp\left(\widetilde{\gamma}_T \widetilde{Q}^\omega_{T,v}\left(u,\,\widetilde{v}\right)\right)\pi_{T,v}\left(u\right)du} \geq \operatorname{ess\,inf} \mathbf{S}\left(\widetilde{\mathscr{V}}^\omega,\,c\right) + o\left(1\right).$$

Since almost every path ω of the Gaussian process $\widetilde{\mathscr{V}}$ achieves its maximum at a unique point on compact sets [cf. Bai (1997) and Lemma 2.6 in Kim and Pollard (1990)], we have

$$\lim_{c\downarrow 0} \operatorname{ess\,inf} \mathbf{S}\left(\widetilde{\mathscr{V}}^\omega,\,c\right) = \lim_{c\downarrow 0} \operatorname{ess\,sup} \mathbf{S}\left(\widetilde{\mathscr{V}}^\omega,\,c\right) = \arg\max_{u\in\Gamma_{\overline{u}}} \widetilde{\mathscr{V}}^\omega\left(u\right).$$

Hence, we have proved (A.29) which by the dominated convergence theorem then implies the weak convergence of $\widetilde{\xi}_T$ toward $\widetilde{\xi}_0$. Since the law of $\widetilde{\xi}_T$ $(\widetilde{\xi}_0)$ under $\widetilde{\mathbb{P}}$ is the same as the law of ξ_T (ξ_0) under $\mathbb{P}$, the claim of the Lemma follows. □

We are now in a position to conclude the proof of Proposition 3.2. For a set $\mathbf{T} \subset \mathbb{R}$ and $m = 0,\, 1$ we define $J_m\left(\mathbf{T}\right) \triangleq \int_{\mathbf{T}} u^m \exp\left(\widetilde{\gamma}_T\left(\widetilde{G}_{T,v}\left(u,\,\widetilde{v}\right) + Q_{T,v}\left(u\right)\right)\right)\pi_{T,v}\left(u\right)du$. Hence, with this notation equation (3.11) can be rewritten as $T\left\|\delta_T\right\|^2\left(\widehat{\lambda}_b^{\mathrm{GL},*}\left(\widetilde{v},\,v\right) - \lambda^0_{b,T}\left(v\right)\right) = J_1\left(\mathbb{R}\right)/J_0\left(\mathbb{R}\right)$. Applying simple manipulations, we obtain,

$$J_1\left(\mathbb{R}\right)/J_0\left(\mathbb{R}\right) = \frac{J_1\left(\Gamma_{\overline{u}}\right) + J_1\left(\Gamma^c_{\overline{u},T}\right)}{J_0\left(\Gamma_{\overline{u}}\right) + J_0\left(\Gamma^c_{\overline{u},T}\right)} = \frac{J_1\left(\Gamma_{\overline{u}}\right)}{J_0\left(\Gamma_{\overline{u}}\right)}\left[1 - \frac{J_0\left(\Gamma^c_{\overline{u},T}\right)}{J_0\left(\mathbb{R}\right)}\right] + \frac{J_1\left(\Gamma^c_{\overline{u},T}\right)}{J_0\left(\mathbb{R}\right)}. \tag{A.31}$$

By Lemma A.21, $J_m\left(\Gamma^c_{\overline{u},T}\right)/J_0\left(\mathbb{R}\right) = o_{\mathbb{P}}\left(1\right)$ $(m = 0,\, 1)$ uniformly in $\widetilde{v} \in \mathbf{V}$. By Lemma A.22, with $\xi_T\left(\widetilde{v}\right) = J_1\left(\Gamma_{\overline{u}}\right)/J_0\left(\Gamma_{\overline{u}}\right)$, the first right-hand side term in (A.31) converges weakly to $\arg\max_{u\in\mathbb{R}} \mathscr{V}\left(u\right)$ in $\mathbb{D}_b\left(\mathbf{V}\right)$.

A.3.4 Proof of Corollary 3.1

The proof involves a simple change in variable. We refer to Proposition 3 in Bai (1997).

A.3.5 Proof of Theorem 3.2

We begin by introducing some notation. Since $l \in \boldsymbol{L}$, for all real numbers B sufficiently large and ϑ sufficiently small the following relationship holds

$$\inf_{|u|>B} l\left(u\right) - \sup_{|u|\leq B^\vartheta} l\left(u\right) \geq 0. \tag{A.32}$$

Let $\zeta_{T,v}(u, \widetilde{v}) = \exp\left(G_{T,v}(u, \widetilde{v}) - \Lambda^0(u)\right)$, $\Gamma_T \triangleq \{u \in \mathbb{R} : \lambda_b \in \Gamma^0\}$ and

$$\Gamma_M = \{u \in \mathbb{R} : M \leq |u| < M + 1\} \cap \Gamma_T,$$

and define

$$J_{1,M} \triangleq \int_{\Gamma_M} \zeta_{T,v}(u, \widetilde{v}) \pi_{T,v}(u) du, \qquad J_2 \triangleq \int_{\Gamma_T} \zeta_{T,v}(u, \widetilde{v}) \pi_{T,v}(u) du. \tag{A.33}$$

In some steps in the proof we shall be working with elements of the following families of functions. A function $f_T : \mathbb{R} \to \mathbb{R}$ is said to belong to the family $\boldsymbol{F}$ if it satisfies the following properties: (1) For fixed T, $f_T(x)$ increases monotocically to infinity with $x \in [0, \infty)$; (2) For any $b < \infty$, $x^b \exp(-f_T(x)) \to 0$ as both T and x diverge to infinity.

Proof. The random variable $T \|\delta_T\|^2 \left(\widehat{\lambda}_b^{\mathrm{GL}} - \lambda_0\right) = \widetilde{\tau}_T$ is a minimizer of the function

$$\Psi_{l,T}(s) = \int_{\Gamma_T} l(s-u) \frac{\exp\left(\widetilde{G}_{T,v}(u, \widetilde{v}) + Q_{T,v}(u)\right) \pi_{T,v}(u)}{\int_{\Gamma_T} \exp\left(\widetilde{G}_{T,v}(w, \widetilde{v}) + Q_{T,v}(u)\right) \pi_{T,v}(w) dw} du.$$

Observe that Lemma A.16-A.20 apply to any polynomial $p \in \boldsymbol{P}$; therefore, they are still valid for $l \in \boldsymbol{L}$. We then have that the asymptotic behavior of $\Psi_{l,T}(s)$ only matters when u (and thus s) varies on $\Gamma_K = \{u \in \mathbb{R} : u \leq K\}$. By Lemma A.27-A.28, for any $\vartheta > 0$, there exists a $\overline{T}$ such that for all $T > \overline{T}$,

$$\mathbb{E}\left[\int_{\Gamma_K} \frac{\exp\left(\widetilde{G}_{T,v}(u, \widetilde{v}) + Q_{T,v}(u)\right)}{\int_{\Gamma_T} \exp\left(\widetilde{G}_{T,v}(w, \widetilde{v}) + Q_{T,v}(w)\right) dw} du\right] \leq \frac{c_\vartheta}{K^\vartheta}. \tag{A.34}$$

Therefore, for all $T > \overline{T}$,

$$\Psi_{l,T}(s) = \frac{\int_{|u| \leq K} l(s-u) \exp\left(\widetilde{G}_{T,v}(u, \widetilde{v}) + Q_{T,v}(u)\right) du}{\int_{|w| \leq K} \exp\left(\widetilde{G}_{T,v}(w, \widetilde{v}) + Q_{T,v}(w)\right) dw} + o_{\mathbb{P}}(1), \tag{A.35}$$

where the $o_{\mathbb{P}}(1)$ term is uniform in $T > \overline{T}$ as K increases to infinity. By Assumption (3.2), $\left|\pi_{T,v}(u) - \pi^0\right| \leq \left|\pi\left(\lambda_{b,T}^0(v)\right) - \pi^0\right| + C\psi_T^{-1}|u|$, with $C > 0$. On $\{|u| \leq K\}$, the first term on the right-hand side is $o(1)$ and does not depend on u. The second term is negligible when T is large. Thus, without loss of generality we set $\pi_{T,v}(u) = 1$ for all u in what follows.

Next, we show the convergence of the marginal distributions of the estimate $\Psi_{l,T}(s)$ to the marginals of the random function $\Psi_l(s)$, where the region of integration in the definition of both the numerator and denominator of $\Psi_{l,T}(s)$ and $\Psi_l(s)$ is restricted to $\{|u| \leq K\}$ only, in view of (A.35). For a finite integer n, choose arbitrary real numbers a_j $(j = 0, \ldots, n)$ and introduce the following estimate:

$$\sum_{j=1}^{n} a_j \int_{|u| \leq K} l(s_j - u) \zeta_{T,v}(u, \widetilde{v}) du + a_0 \int_{|u| \leq K} l(s_0 - u) \zeta_{T,v}(u, \widetilde{v}) du. \tag{A.36}$$

By Lemma A.24 and A.30, we can invoke Theorem I.A.22 in Ibragimov and Has'minskiĭ (1981) which gives the convergence in distribution of the estimate in (A.36) towards the distribution of the following random variable:

$$\sum_{j=1}^{n} a_j \int_{|u| \leq K} l(s_j - u) \exp(\mathscr{V}(u)) du + a_0 \int_{|u| \leq K} l(s_0 - u) \exp(\mathscr{V}(u)) du.$$

By the Cramer-Wold Theorem [cf. Theorem 29.4 in Billingsley (1995)] this suffices for the convergence in distribution of the vector

$$\int_{|u|\leq K} l\left(s_{i}-u\right)\zeta_{T,v}\left(u,\,\widetilde{v}\right)du,\ldots,\int_{|u|\leq K} l\left(s_{n}-u\right)\zeta_{T,v}\left(u,\,\widetilde{v}\right)du,\qquad \int_{|u|\leq K} l\left(s_{0}-u\right)\zeta_{T,v}\left(u,\,\widetilde{v}\right)du,$$

to the distribution of the vector

$$\int_{|u|\leq K} l\left(s_{i}-u\right)\exp\left(\mathscr{V}\left(u\right)\right)du,\ldots,\int_{|u|\leq K} l\left(s_{n}-u\right)\exp\left(\mathscr{V}\left(u\right)\right)du,\qquad \int_{|u|\leq K} l\left(s_{0}-u\right)\exp\left(\mathscr{V}\left(u\right)\right)du.$$

As a consequence, for any $K_1,\,K_2<\infty$, the marginal distributions of

$$\frac{\int_{|u|\leq K_{1}} l\left(s-u\right)\exp\left(\widetilde{G}_{T,v}\left(u,\,\widetilde{v}\right)+Q_{T,v}\left(u\right)\right)du}{\int_{|w|\leq K_{2}}\exp\left(\widetilde{G}_{T,v}\left(w,\,\widetilde{v}\right)+Q_{T,v}\left(w\right)\right)dw},$$

converge to the marginals of $\int_{|u|\leq K_{1}} l\left(s-u\right)\exp\left(\mathscr{V}\left(u\right)\right)du/\left(\int_{|w|\leq K_{2}}\exp\left(\mathscr{V}\left(w\right)\right)dw\right)$. The same convergence result extends to the distribution of

$$\int_{M\leq|u|<M+1}\frac{\exp\left(\widetilde{G}_{T,v}\left(u,\,\widetilde{v}\right)+Q_{T,v}\left(u\right)\right)}{\int_{|w|\leq K_{2}}\exp\left(\widetilde{G}_{T,v}\left(w,\,\widetilde{v}\right)+Q_{T,v}\left(w\right)\right)dw}du,$$

towards the distribution of $\int_{M\leq|u|<M+1}(\exp\left(\mathscr{V}\left(u\right)\right)du/\int_{|w|\leq K_{2}}\exp\left(\mathscr{V}\left(w\right)\right)dw)$. By choosing $K_2>M+1$ we deduce

$$\mathbb{E}\left[\int_{M\leq|u|<M+1}\frac{\exp\left(\mathscr{V}\left(u\right)\right)}{\int_{\mathbb{R}}\exp\left(\mathscr{V}\left(w\right)\right)dw}du\right]\leq\lim_{T\rightarrow\infty}\mathbb{E}\left[\int_{\Gamma_{M}}\frac{\exp\left(\widetilde{G}_{T,v}\left(u,\,\widetilde{v}\right)+Q_{T,v}\left(u\right)\right)}{\int_{|w|\leq K_{2}}\exp\left(\widetilde{G}_{T,v}\left(w,\,\widetilde{v}\right)+Q_{T,v}\left(w\right)\right)dw}du\right]\leq c_{\vartheta}M^{-\vartheta},$$

in view of (A.34). This leads to

$$\Psi_{l}\left(s\right)=\int_{|u|\leq K} l\left(s-u\right)\frac{\exp\left(\mathscr{V}\left(u\right)\right)du}{\int_{|w|\leq K}\exp\left(\mathscr{V}\left(w\right)\right)dw}+o_{\mathbb{P}}\left(1\right),\tag{A.37}$$

where the $o_{\mathbb{P}}(1)$ term is uniform as K increases to infinity. We then have the convergence of the finite-dimensional distributions of $\Psi_{l,T}\left(s\right)$ toward $\Psi_{l}\left(s\right)$. Next, we need to prove the tightness of the sequence $\{\Psi_{l,T}\left(s\right),\,T\geq1\}$. More specifically, we shall show that the family of distributions on the space of continuous functions $\mathbb{C}_{b}\left(K\right)$ generated by the contractions of $\Psi_{l,T}\left(s\right)$ on $\{|s|\leq K\}$ are dense. For any $l\in\boldsymbol{L}$ the inequality $l\left(u\right)\leq2^{r}\left(1+|u|^{2}\right)^{r}$ holds for some r. Let

$$\Upsilon_{K}\left(\varpi\right)\triangleq\int_{\mathbb{R}}\sup_{|s|\leq K,\,|y|\leq\varpi}\left|l\left(s+y-u\right)-l\left(s-u\right)\right|\left(1+|u|^{2}\right)^{-r-1}du.$$

Fix $K<\infty$. We show $\lim_{\varpi\downarrow0}\Upsilon_{K}\left(\varpi\right)=0$. Note that for any $\kappa>0$, we can choose a M such that

$$\int_{|u|>M}\sup_{|s|\leq K,\,|y|\leq\varpi}\left|l\left(s+y-u\right)-l\left(s-u\right)\right|\left(1+|u|^{2}\right)^{-r-1}du<\kappa.$$

We now use Lusin's Theorem [cf. Section 3.3 in Royden and Fitzpatrick (2010)]. Since $l\left(\cdot\right)$ is measurable, there exists a continuous function $g\left(u\right)$ in the interval $\{u\in\mathbb{R}:\,|u|\leq K+2M\}$ which agrees with $l\left(u\right)$ except on a set whose measure does not exceed $\kappa\left(2\overline{L}\right)^{-1}$, where $\overline{L}$ is the upper bound of $l\left(\cdot\right)$ on

$\{u \in \mathbb{R} : |u| \leq K + 2M\}$. Denote the modulus of continuity of $g(\cdot)$ by $w_g(\varpi)$. Without loss of generality assume $|g(u)| \leq \overline{L}$ for all u satisfying $|u| \leq K + 2M$. Then,

$$\begin{aligned}
&\int_{|u|>M} \sup_{|s|\leq K,\, |y|\leq\varpi} |l(s+y-u) - l(s-u)| \left(1+|u|^2\right)^{-r-1} du \\
&\qquad \leq \int_{\mathbb{R}} \sup_{|s|\leq K,\, |y|\leq\varpi} |l(s+y-u) - l(s-u)| \left(1+|u|^2\right)^{-r-1} du \\
&\qquad \leq w_g(\varpi) \int_{\mathbb{R}} \sup_{|s|\leq K,\, |y|\leq\varpi} \left(1+|u|^2\right)^{-r-k} du + 2\overline{L}\mathrm{Leb}\{u \in \mathbb{R} : |u| \leq K + 2M,\, l \neq g\},
\end{aligned}$$

and $\overline{L} \leq C w_g(\varpi) + \kappa$ for some C. Hence, $\Upsilon_K(\varpi) \leq C w_g(\varpi) + 2\kappa$ since κ can be chosen arbitrary small and (for each fixed κ) $w_g(\varpi) \to 0$ as $\varpi \downarrow 0$ by definition. By Assumption 3.7, there exists a number $C < \infty$ such that

$$\begin{aligned}
&\mathbb{E}\left[\sup_{|s|\leq K,\, |y|\leq\varpi} |\Psi_{l,T}(s+y) - \Psi_{l,T}(s)|\right] \\
&\quad \leq \int_{\mathbb{R}} \sup_{|s|\leq K,\, |y|\leq\varpi} |l(s+y-u) - l(s-u)| \,\mathbb{E}\left(\frac{\exp\left(\widetilde{G}_{T,v}(u, \widetilde{v}) + Q_{T,v}(u)\right)}{\int_{\mathbf{U}_T} \exp\left(\widetilde{G}_{T,v}(w, \widetilde{v}) + Q_{T,v}(w)\right) dw}\right) du \\
&\quad \leq C\Upsilon_K(\varpi).
\end{aligned}$$

Markov's inequality together with the above bound establish that the family of distributions generated by the contractions of $\Psi_{T,l}$ is dense in $\mathbb{C}_b(K)$. Since the finite-dimensional convergence in distribution was demonstrated above, we can deduce the weak convergence $\Psi_{l,T} \Rightarrow \Psi_l$ in $\mathbb{D}_b(\mathbf{V})$ uniformly in $\lambda_b^0 \in \mathbf{K}$. Finally, we examine the oscillations of the minimum points of the sample criterion $\Psi_{l,T}$. Consider an open bounded interval $\mathbf{A}$ that satisfies $\mathbb{P}\{\xi_l^0 \in b(\mathbf{A})\} = 0$, where $b(\mathbf{A})$ denotes the boundary of the set $\mathbf{A}$. Choose a real number K sufficiently large such that $\mathbf{A} \subset \{s : |s| \leq K\}$ and define for $|s| \leq K$ the functionals $H_{\mathbf{A}}(\Psi) = \inf_{s\in\mathbf{A}} \Psi_l(s)$ and $H_{\mathbf{A}^c}(\Psi) = \inf_{s\in\mathbf{A}^c} \Psi_l(s)$. Let $\mathbf{M}_T$ denote the set of minimum points of $\Psi_{l,T}$. We have

$$\begin{aligned}
\mathbb{P}[\mathbf{M}_T \subset \mathbf{A}] &= \mathbb{P}[H_{\mathbf{A}}(\Psi) < H_{\mathbf{A}^c}(\Psi),\, \mathbf{M}_T \subset \{s : |s| \leq K\}] \\
&\geq \mathbb{P}[H_{\mathbf{A}}(\Psi) < H_{\mathbf{A}^c}(\Psi)] - \mathbb{P}[\mathbf{M}_T \not\subseteq \{s : |s| \leq K\}].
\end{aligned}$$

Therefore,

$$\liminf_{T\to\infty} \mathbb{P}[\mathbf{M}_T \subset \mathbf{A}] \geq \mathbb{P}[H_{\mathbf{A}}(\Psi) < H_{\mathbf{A}^c}(\Psi)] - \sup_T \mathbb{P}[\mathbf{M}_T \not\subseteq \{s : |s| \leq K\}],$$

and $\limsup_{T\to\infty} \mathbb{P}[\mathbf{M}_T \subset \mathbf{A}] \leq \mathbb{P}[H_{\mathbf{A}}(\Psi) < H_{\mathbf{A}^c}(\Psi)]$. Moreover, the minimum of the population criterion $\Psi_l(\cdot)$ satisfies $\mathbb{P}[\xi_l^0 \in \mathbf{A}] \leq \mathbb{P}[H_{\mathbf{A}}(\Psi) < H_{\mathbf{A}^c}(\Psi)]$ and $\mathbb{P}[\xi_l^0 \in \mathbf{A}] + \mathbb{P}[|\xi_l^0| > K] \geq \mathbb{P}[H_{\mathbf{A}}(\Psi) \leq H_{\mathbf{A}^c}(\Psi)]$. Lemma A.29 shall be used to deduce that the following relationship holds,

$$\limsup_{T\to\infty} \mathbb{E}\left[l\left(T\|\delta_T\|^2\left(\widehat{\lambda}_b^{\mathrm{GL}} - \lambda_b^0\right)\right)\right] < \infty,$$

for any loss function $l \in \boldsymbol{L}$. Hence, the set $\mathbf{M}_T$ of absolute minimum points of the function $\Psi_{l,T}(s)$ are uniformly stochastically bounded for all T large enough: $\lim_{K\to\infty} \mathbb{P}[\mathbf{M}_T \not\subseteq \{s : |s| \leq K\}] = 0$. The latter result together with the uniqueness assumption (cf. Assumption 3.7) yield

$$\lim_{K\to\infty}\left\{\sup_{T}\mathbb{P}\left[\mathbf{M}_T\nsubseteq\{s:\,|s|\leq K\}\right]+\mathbb{P}\left[\left|\xi_l^0\right|>K\right]\right\}=0.$$

Hence, we have

$$\lim_{T\to\infty}\mathbb{P}\left[\mathbf{M}_T\subset\mathbf{A}\right]=\mathbb{P}\left[\xi_l^0\in\mathbf{A}\right]. \tag{A.38}$$

The last step involves showing that the length of the set $\mathbf{M}_T$ approaches zero in probability as $T\to\infty$. Let $\mathbf{A}_d$ denote an interval in $\mathbb{R}$ centered at the origin and of length $d<\infty$. Equation (A.38) guarantees that $\lim_{d\to\infty}\sup_{T\to\infty}\mathbb{P}\left[\mathbf{M}_T\nsubseteq\mathbf{A}_d\right]=0$. Choose any $\epsilon>0$ and divide $\mathbf{A}_d$ into admissible subintervals whose lengths do not exceed $\epsilon/2$. Then,

$$\mathbb{P}\left[\sup_{s_i,s_j\in\mathbf{M}_T}|s_i-s_j|>\epsilon\right]\leq\mathbb{P}\left[\mathbf{M}_T\nsubseteq\mathbf{A}_d\right]+(1+2d/\epsilon)\sup\mathbb{P}\left[H_{\mathbf{A}}\left(\Psi_{l,T}\right)=H_{\mathbf{A}^c}\left(\Psi_{l,T}\right)\right],$$

where the term $1+2d/\epsilon$ is an upper bound on the admissible number of subintervals and the supremum in the second term is over all possible open bounded subintervals $\mathbf{A}\subset\mathbf{A}_d$. The weak convergence result implies $\mathbb{P}\left[H_{\mathbf{A}}\left(\Psi_{l,T}\right)=H_{\mathbf{A}^c}\left(\Psi_{l,T}\right)\right]\to\mathbb{P}\left[H_{\mathbf{A}}\left(\Psi_l\right)=H_{\mathbf{A}^c}\left(\Psi_l\right)\right]$ as $T\to\infty$. Since $\mathbb{P}\left[H_{\mathbf{A}}\left(\Psi_l\right)=H_{\mathbf{A}^c}\left(\Psi_l\right)\right]=0$ and $\mathbb{P}\left[\mathbf{M}_T\nsubseteq\mathbf{A}_d\right]\to0$ for large d, then $\mathbb{P}\left[\sup_{s_i,s_j\in\mathbf{M}_T}|s_i-s_j|>\epsilon\right]=o(1)$. Since $\epsilon>0$ can be chosen arbitrary small we deduce that the distribution of $T\left\|\delta_T\right\|^2\left(\widehat{\lambda}_b^{\mathrm{GL}}-\lambda_b^0\right)$ converges to the distribution of ξ_l^0. □

Lemma A.23. *Let $u_1,\,u_2\in\mathbb{R}$ be of the same sign with $0<|u_1|<|u_2|$. For any integer $r>0$ and some constants c_r and C_r which depend on r only, we have uniformly in $\widetilde{v}\in\mathbf{V}$,*

$$\mathbb{E}\left[\left(\zeta_{T,v}^{1/2r}\left(u_2,\,\widetilde{v}\right)-\zeta_{T,v}^{1/2r}\left(u_1,\,\widetilde{v}\right)\right)^{2r}\right]\leq c_r\left|\left(\delta^0\right)'\left(\left|u_2-u_1\right|\Sigma_i\right)\delta^0\right|^r\leq C_r\left|u_2-u_1\right|^r,$$

where Σ_i is defined in Assumption 3.5 and $i=1$ if $u_1<0$ and $i=2$ if $u_1>0$.

Proof. The proof is given for the case $u_2>u_1>0$. The other case is similar and thus omitted. We follow closely the proof of Lemma III.5.2 in Ibragimov and Has'minskiĭ (1981). Let $\mathcal{V}\left(u_i\right)=\exp\left(\mathscr{V}\left(u_i\right)\right)$, $i=1,\,2$. We have $\mathbb{E}\left[\left(\mathcal{V}^{1/2r}\left(u_2\right)-\mathcal{V}^{1/2r}\left(u_1\right)\right)^{2r}\right]=\sum_{j=0}^{2r}\binom{2r}{j}(-1)^j\,\mathbb{E}_{u_1}\left[\mathcal{V}_{u_1}^{j/2r}\left(u_2\right)\right]$, where $\mathcal{V}_{u_1}\left(u_2\right)\triangleq\exp\left(\mathscr{V}\left(u_2\right)-\mathscr{V}\left(u_1\right)\right)$. Using the Gaussian property of $\mathscr{V}(u)$, for each $u\in\mathbb{R}$, we have

$$\mathbb{E}_{u_1}\left[\mathcal{V}^{j/2r}\left(u_2\right)\right]=\exp\left(\frac{1}{2}\left(\frac{j}{2r}\right)^2 4\left(\delta^0\right)'\left(\left|u_2-u_1\right|\Sigma_2\right)\delta^0-\frac{j}{2r}\left|\Lambda^0\left(u_2\right)-\Lambda^0\left(u_1\right)\right|\right). \tag{A.39}$$

Then, $\mathbb{E}\left[\left(\mathcal{V}^{1/2r}\left(u_2\right)-\mathcal{V}^{1/2r}\left(u_1\right)\right)^{2r}\right]=\sum_{j=0}^{2r}\binom{2r}{j}(-1)^j\,d^{j/2r}$ with

$$d\triangleq\exp\left(\frac{j}{2r}2\left(\delta^0\right)'\left(\left|u_2-u_1\right|\Sigma_2\right)\delta^0-\left|\Lambda^0\left(u_2\right)-\Lambda^0\left(u_1\right)\right|\right).$$

Let $B\triangleq2\left(\delta^0\right)'\left(\left|u_2-u_1\right|\Sigma_2\right)\delta^0-\left|\Lambda^0\left(u_2\right)-\Lambda^0\left(u_1\right)\right|$. There are different cases to be considered:
(1) $B<0$. Note that

$$d=\exp\left(\frac{j}{2r}2\left(\delta^0\right)'\left(\left|u_2-u_1\right|\Sigma_2\right)\delta^0-\left|\left(\delta^0\right)'\left(\left|u_2-u_1\right|\Sigma_2\right)\delta^0\right|+B\right)$$

$$= \exp\left(-\frac{2r-j}{r}\left(\delta^0\right)'\left(|u_2-u_1|\,\Sigma_2\right)\delta^0\right)e^B,$$

which then results in

$$\mathbb{E}\left[\left(\mathcal{V}^{1/2r}\left(u_2\right)-\mathcal{V}^{1/2r}\left(u_1\right)\right)^{2r}\right]\leq p_r\left(a\right), \tag{A.40}$$

where $p_r\left(a\right)\triangleq\sum_{j=0}^{2r}\binom{2r}{j}\left(-1\right)^j a^{(2r-j)}$ and $a=e^{B/2r}\exp\left(-r^{-1}\left(\delta^0\right)'\left(|u_2-u_1|\,\Sigma_2\right)\delta^0\right)$.

(2) $2\left(\delta^0\right)'\left(|u_2-u_1|\,\Sigma_2\right)\delta^0=\left|\Lambda^0\left(u_2\right)-\Lambda^0\left(u_1\right)\right|$. This case is the same as the previous one but with $a=\exp\left(-r^{-1}\left(\delta^0\right)'\left(|u_2-u_1|\,\Sigma_2\right)\delta^0\right)$.

(3) $B>0$. Upon simple manipulations, $\mathbb{E}\left[\left(\mathcal{V}^{1/2r}\left(u_2\right)-\mathcal{V}^{1/2r}\left(u_1\right)\right)^{2r}\right]\leq p_r\left(a\right)$, where

$$p_r\left(a\right)=e^{-B/2r}\sum_{j=0}^{2r}\binom{2r}{j}\left(-1\right)^j a^{(2r-j)},$$

with $a=\exp\left(-r^{-1}\left(\delta^0\right)'\left(|u_2-u_1|\,\Sigma_2\right)\delta^0\right)$. We can thus proceed with the same proof for all the above cases. Let us consider the first case. We show that at the point $a=1$, the polynomial $p_r\left(a\right)$ admits a root of multiplicity r. This can be established by verifying the equalities $p_r\left(1\right)=p_r^{(1)}\left(1\right)=\cdots=p_r^{(r-1)}\left(1\right)=0$. One then recognizes that $p_r^{(i)}\left(a\right)$ is a linear combination of summations $\mathcal{S}_k$ ($k=0,\,1,\ldots,\,2i$) given by $\mathcal{S}_k=e^B\sum_{j=0}^{2r}\binom{2r}{j}j^k$. Thus, one only needs to verify that $\mathcal{S}_k=0$ for $k=0,\,1,\ldots,\,2r-2$. This follows because the expression for $\mathcal{S}_k$ is found by applying the operator $e^B a\left(d/da\right)$ to the function $\left(1-a^2\right)^{2r}$ and evaluating it at $a=1$. Consequently, $\mathcal{S}_k=0$ for $k=0,\,1,\ldots,\,2r-1$. Using this result into (A.40) we find, with $\widetilde{p}_r\left(a\right)$ being a polynomial of degree r^2-r,

$$\mathbb{E}\left[\left(\mathscr{V}^{1/2r}\left(u_2\right)-\mathscr{V}^{1/2r}\left(u_1\right)\right)^{2r}\right]=\left(1-a\right)^r\widetilde{p}_r\left(a\right)\leq\left(r^{-1}\left(\delta^0\right)'\left(|u_2-u_1|\,\Sigma_2\right)\delta^0\right)^r\widetilde{p}_r\left(a\right), \tag{A.41}$$

where the last inequality follows from $1-e^{-c}\leq c$, for $c>0$. Next, let $\overline{\zeta}_{T,v}^{1/2r}\left(u_2,\,u_1\right)=\zeta_{T,v}^{1/2r}\left(u_2\right)-\zeta_{T,v}^{1/2r}\left(u_1\right)$. By Lemma A.3 and A.12, the continuous mapping theorem and (A.41), $\lim_{T\to\infty}\mathbb{E}\left[\overline{\zeta}_{T,v}^{1/2r}\left(u_2,\,u_1\right)\right]\leq\left(1-a\right)^r\widetilde{p}_r\left(a\right)$, uniformly in $\widetilde{v}\in\mathbf{V}$. Noting that $j\leq 2r$, we can set $C_r=\max_{0\leq a\leq 1}e^B\widetilde{p}_r\left(a\right)/r^r$ to prove the lemma. □

Lemma A.24. *For $u_1,\,u_2\in\mathbb{R}$ being of the same sign and satisfying $0<|u_1|<|u_2|<K<\infty$. Then, for all T sufficiently large, we have*

$$\mathbb{E}\left[\left(\zeta_{T,v}^{1/4}\left(u_2,\,\widetilde{v}\right)-\zeta_{T,v}^{1/4}\left(u_1,\,\widetilde{v}\right)\right)^4\right]\leq C_1\left|u_2-u_1\right|^2, \tag{A.42}$$

where $0<C_1<\infty$. Furthermore, for the constant C_1 from Lemma A.23, we have

$$\mathbb{P}\left[\zeta_{T,v}\left(u,\,\widetilde{v}\right)>\exp\left(-3C_1\left|u\right|/2\right)\right]\leq\exp\left(-C_1\left|u\right|/4\right). \tag{A.43}$$

Both relationships are valid uniformly in $\widetilde{v}\in\mathbf{V}$.

Proof. Suppose $u>0$. The relationship in (A.42) follows from Lemma A.23 with $r=2$. By Markov's inequality and Lemma A.23,

$$\mathbb{P}\left[\zeta_{T,v}\left(u,\,\widetilde{v}\right)>\exp\left(-3C_1\left|u\right|/2\right)\right]\leq\exp\left(3C_1\left|u\right|/4\right)\mathbb{E}\left[\zeta_{T,v}^{1/2}\left(u,\,\widetilde{v}\right)\right]$$

$$\leq \exp\left(3C_1 |u| /4 - \left(\delta^0\right)' (|u| \Sigma_2) \delta^0\right) \leq \exp\left(-C_1 |u| /4\right).$$

□

Lemma A.25. *Under the conditions of Lemma A.24, for any $\vartheta > 0$ there exists a finite real number c_ϑ and a $\overline{T}$ such that for all $T > \overline{T}$, $\sup_{\widetilde{v}\in\mathbf{V}} \mathbb{P}\left[\sup_{|u|>M} \zeta_{T,v}(u, \widetilde{v}) > M^{-\vartheta}\right] \leq c_\vartheta M^{-\vartheta}$.*

Proof. It can be shown using Lemma A.23-A.24. □

Lemma A.26. *For every sufficiently small $\epsilon \leq \overline{\epsilon}$, where $\overline{\epsilon}$ depends on the smoothness of $\pi(\cdot)$, there exists $0 < C < \infty$ such that*

$$\mathbb{P}\left[\int_0^\epsilon \zeta_{T,v}(u, \widetilde{v}) \pi\left(\lambda_b^0 + u/\psi_T\right) du < \epsilon\pi\left(\lambda_b^0\right)\right] < C\epsilon^{1/2}. \tag{A.44}$$

Proof. Since $\mathbb{E}(\zeta_{T,v}(0, \widetilde{v})) = 1$ and $\mathbb{E}(\zeta_{T,v}(u, \widetilde{v})) \leq 1$ for sufficiently large T, we have

$$\mathbb{E}\left|\zeta_{T,v}(u, \widetilde{v}) - \zeta_{T,v}(0, \widetilde{v})\right| \leq \left(\mathbb{E}\left|\zeta_{T,v}^{1/2}(u, \widetilde{v}) + \zeta_{T,v}^{1/2}(0, \widetilde{v})\right|^2 \mathbb{E}\left|\zeta_{T,v}^{1/2}(u, \widetilde{v}) - \zeta_{T,v}^{1/2}(0, \widetilde{v})\right|^2\right)^{1/2} \leq C|u|^{1/2}, \tag{A.45}$$

by Lemma A.23 with $r = 1$. By Assumption 3.2, $\left|\pi_{T,v}(u) - \pi^0\right| \leq \left|\pi\left(\lambda_{b,T}^0(v)\right) - \pi^0\right| + C\psi_T^{-1}|u|$, with $C > 0$. The first term on the right-hand side is $o(1)$ (and independent of u) while the second is asymptotically negligible for small u. Thus, for a sufficiently small $\overline{\epsilon} > 0$,

$$\int_0^\epsilon \zeta_{T,v}(u, \widetilde{v}) \pi_{T,v}(u) du > \frac{\pi^0}{2} \int_0^\epsilon \zeta_{T,v}(u, \widetilde{v}) du.$$

Next, using $\zeta_{T,v}(0, \widetilde{v}) = 1$,

$$\begin{aligned}\mathbb{P}\left[\int_0^\epsilon \zeta_{T,v}(u, \widetilde{v}) \pi_{T,v}(u) du < \epsilon/2\right] &\leq \mathbb{P}\left[\int_0^\epsilon \left(\zeta_{T,v}(u, \widetilde{v}) - \zeta_{T,v}(0, \widetilde{v})\right) du < -\epsilon/2\right] \\ &\leq \mathbb{P}\left[\int_0^\epsilon \left|\zeta_{T,v}(u, \widetilde{v}) - \zeta_{T,v}(0, \widetilde{v})\right| du > \epsilon/2\right],\end{aligned}$$

and by Markov's inequality together with (A.45) the last expression is less than or equal to

$$(2/\epsilon) \int_0^\epsilon \mathbb{E}\left|\zeta_{T,v}(u, \widetilde{v}) - \zeta_{T,v}(0, \widetilde{v})\right| du < 2C\epsilon^{1/2}.$$

□

Lemma A.27. *For $f_T \in \boldsymbol{F}$, and M sufficiently large, there exist constants c, $C > 0$ such that*

$$\mathbb{P}\left[J_{1,M} > \exp\left(-cf_T(M)\right)\right] \leq C\left(1 + M^C\right) \exp\left(-cf_T(M)\right), \tag{A.46}$$

uniformly in $\widetilde{v} \in \mathbf{V}$.

Proof. In view of the smotheness property of $\pi(\cdot)$, without loss of generality we consider the case of the uniform prior (i.e., $\pi_{T,v}(u) = 1$ for all u). We begin by dividing the open interval $\{u : M \leq |u| < M+1\}$ into I disjoint segments denoting the i-th one by Π_i. For each segment Π_i choose a point u_i and define $J_{1,M}^{\Pi} \triangleq \sup_{\widetilde{v}\in\mathbf{V}} \sum_{i\in I} \zeta_{T,v}(u_i, \widetilde{v}) \text{Leb}(\Pi_i) = \sup_{\widetilde{v}\in\mathbf{V}} \sum_{i\in I} \int_{\Pi_i} \zeta_{T,v}(u_i, \widetilde{v}) du$. Then,

$$\mathbb{P}\left[J_{1,M}^{\Pi} > (1/4) \exp\left(-cf_T(M)\right)\right] \leq \mathbb{P}\left[\max_{i\in I} \sup_{\widetilde{v}\in\mathbf{V}} \zeta_{T,v}^{1/2}(u_i, \widetilde{v}) \left(\text{Leb}(\Gamma_M)\right)^{1/2} > (1/2) \exp\left(-f_T(M)/2\right)\right]$$

$$
\begin{aligned}
&\le \sum_{i\in I} \mathbb{P}\left[\zeta_{T,v}^{1/2}\left(u_i,\,\widetilde{v}\right) > (1/2)\left(\mathrm{Leb}\left(\Gamma_M\right)\right)^{-1/2}\exp\left(-f_T\left(M\right)/2\right)\right]\\
&\le 2I\left(\mathrm{Leb}\left(\Gamma_M\right)\right)^{1/2}\exp\left(-f_T\left(M\right)/12\right), \qquad \text{(A.47)}
\end{aligned}
$$

where the last inequality follows from applying Lemma A.24 to each summand. Upon using the inequality $\exp\left(-f_T\left(M\right)/2\right) < 1/2$ (which is valid for sufficiently large M), we have

$$
\mathbb{P}\left[J_{1,M} > \exp\left(-f_T\left(M\right)/2\right)\right] \le \mathbb{P}\left[\left|J_{1,M} - J_{1,M}^{\Pi}\right| > (1/2)\exp\left(-f_T\left(M\right)/2\right)\right] + \mathbb{P}\left[J_{1,M}^{\Pi} > \exp\left(-f_T\left(M\right)\right)\right].
$$

Focusing on the first term,

$$
\begin{aligned}
\mathbb{E}\left[J_{1,M} - J_{1,M}^{\Pi}\right] &\le \sum_{i\in I}\int_{\Pi_i}\mathbb{E}\left|\zeta_{T,v}^{1/2}\left(u,\,\widetilde{v}\right) - \zeta_{T,v}^{1/2}\left(u_i,\,\widetilde{v}\right)\right| du\\
&\le \sum_{i\in I}\int_{\Pi_i}\left(\mathbb{E}\left|\zeta_{T,v}^{1/2}\left(u,\,\widetilde{v}\right) + \zeta_{T,v}^{1/2}\left(u_i,\,\widetilde{v}\right)\right|\mathbb{E}\left|\zeta_{T,v}^{1/2}\left(u,\,\widetilde{v}\right) - \zeta_{T,v}^{1/2}\left(u_i,\,\widetilde{v}\right)\right|\right)^{1/2} du\\
&\le C\left(1+M\right)^{C}\sum_{i\in I}\int_{\Pi_i}\left|u_i - u\right|^{1/2} du,
\end{aligned}
$$

where for the last inequality we have used Lemma A.24 since we can always choose the partition of the segments such that each Π_i contains either positive or negative u_i. Since each summand on the right-hand side above is less than $C\left(MI^{-1}\right)^{3/2}$ there exist numbers C_1 and C_2 such that

$$
\mathbb{E}\left[J_{1,M} - J_{1,M}^{\Pi}\right] \le C_1\left(1+M^{C_2}\right)I^{-1/2}. \qquad \text{(A.48)}
$$

Using (A.47) and (A.48) we have

$$
\mathbb{P}\left[J_{1,M} > \exp\left(-f_T\left(M\right)/2\right)\right] \le C_1\left(1+M^{C_2}\right)I^{-1/2} + 2I\left(\mathrm{Leb}\left(\Gamma_M\right)\right)^{1/2}\exp\left(-f_T\left(M\right)/12\right).
$$

The relationship in the last display leads to the claim of the lemma if we choose I satisfying $1 \le I^{3/2}\exp\left(-f_T\left(M\right)/4\right) \le 2$. □

Lemma A.28. *For $f_T \in \boldsymbol{F}$, and M sufficiently large, there exist constants $c,\,C > 0$ such that*

$$
\mathbb{E}\left[J_{1,M}/J_2\right] \le C\left(1+M^{C}\right)\exp\left(-cf_T\left(M\right)\right), \qquad \text{(A.49)}
$$

uniformly in $\widetilde{v} \in \mathbf{V}$.

Proof. Note that $J_{1,M}/J_2 \le 1$. Thus, for any $\epsilon > 0$,

$$
\mathbb{E}\left[J_{1,M}/J_2\right] \le \mathbb{P}\left[J_{1,M} > \exp\left(-cf_T\left(M\right)/2\right)\right] + (4/\epsilon)\exp\left(-cf_T\left(M\right)\right) + \mathbb{P}\left[\int_{\Gamma_T}\zeta_{T,v}\left(u,\,\widetilde{v}\right)du < \epsilon/4\right].
$$

By Lemma A.27, the first term is bounded by $C\left(1+M^{C}\right)\exp\left(-cf_T\left(M\right)/4\right)$ while for the last term we can use (A.44) to deduce

$$
\mathbb{E}\left[J_{1,M}/J_2\right] \le C\left(1+M^{C}\right)\exp\left(-cf_T\left(M\right)\right) + (4/\epsilon)\exp\left(-cf_T\left(M\right)\right) + C\epsilon^{1/2}.
$$

Finally, choose $\epsilon = \exp\left(\left(-2c/3\right)f_T\left(M\right)\right)$ to complete the proof of the lemma. □

Lemma A.29. *For $l \in \boldsymbol{L}$ and and any $\vartheta > 0$, $\lim_{B\to\infty}\lim_{T\to\infty}B^{\vartheta}\mathbb{P}\left[\psi_T\left(\widehat{\lambda}_b^{\mathrm{GL}} - \lambda_b^0\right) > B\right] = 0$.*

Proof. Let $p_T(u) \triangleq p_{1,T}(u)/\overline{p}_T$ where $p_{1,T}(u) = \exp\left(\widetilde{G}_{T,v}(u, \widetilde{v}) + Q_{T,v}(u)\right)$ and $\overline{p}_T \triangleq \int_{\mathbf{U}_T} p_{1,T}(w)\, dw$. By definition, $\widehat{\lambda}_b^{\mathrm{GL}}$ is the minimum of the function $\int_{\Gamma^0} l\left(T\|\delta_T\|^2(s-u)\right) p_{1,T}(u)\,\pi_{T,v}(u)\,du$ with $s \in \Gamma^0$. Upon using a change in variables,

$$\begin{aligned}
&\int_{\Gamma^0} l\left(T\|\delta_T\|^2(s-u)\right) p_{1,T}(u)\,\pi_{T,v}(u)\,du \\
&= \left(T\|\delta_T\|^2\right)^{-1} \overline{p}_T \int_{\mathbf{U}_T} l\left(T\|\delta_T\|^2\left(s-\lambda_b^0\right)-u\right) p_T\left(\lambda_{b,T}^0(v) + \left(T\|\delta_T\|^2\right)^{-1} u\right) \\
&\quad \times \pi_{T,v}\left(\lambda_{b,T}^0(v) + \left(T\|\delta_T\|^2\right)^{-1} u\right) du.
\end{aligned}$$

Thus, $\lambda_{\delta,T} \triangleq T\|\delta_T\|^2\left(\widehat{\lambda}_b^{\mathrm{GL}} - \lambda_b^0\right)$ is the minimum of the function

$$\mathcal{S}_T(s) \triangleq \int_{\mathbf{U}_T} l(s-u) \frac{p_T\left(\lambda_b^0 + \left(T\|\delta_T\|^2\right)^{-1} u\right) \pi_{T,v}\left(\lambda_b^0 + \left(T\|\delta_T\|^2\right)^{-1} u\right)}{\int_{\mathbf{U}_T} p_T\left(\lambda_b^0 + \left(T\|\delta_T\|^2\right)^{-1} w\right) \pi_{T,v}\left(\lambda_b^0 + \left(T\|\delta_T\|^2\right)^{-1} w\right) dw} du,$$

where the optimization is over $\mathbf{U}_T$. The random function $\mathcal{S}_T(\cdot)$ converges with probability one in view of Lemma A.27-A.28 together with the properties of the loss function l [cf. (A.35) and the discussion surrounding it]. Therefore, we shall show that the random function $\mathcal{S}_T(s)$ is strictly larger than $\mathcal{S}_T(0)$ on $\{|s| > B\}$ with high probability as $T \to \infty$. This reflects that

$$\mathbb{P}\left[\left|T\|\delta_T\|^2\left(\widehat{\lambda}_b^{\mathrm{GL}} - \lambda_b^0\right)\right| > B\right] \leq \mathbb{P}\left[\inf_{|s|>B} \mathcal{S}_T(s) \leq \mathcal{S}_T(0)\right]. \tag{A.50}$$

We present the proof for the case $\pi_{T,v}(u) = 1$ for all u. The general case follows with no additional difficulties due to the assumptions satisfied by the prior $\pi(\cdot)$. By the properties of the family $\boldsymbol{L}$ of loss functions, we can find $\overline{u}_1, \overline{u}_2 \in \mathbb{R}$, with $0 < \overline{u}_1 < \overline{u}_2$ such that as T increases,

$$\bar{l}_{1,T} \triangleq \sup\{l(u) : u \in \Gamma_{1,T}\} < \bar{l}_{2,T} \triangleq \inf\{l(u) : u \in \Gamma_{2,T}\},$$

where $\Gamma_{1,T} \triangleq \mathbf{U}_T \cap (|u| \leq \overline{u}_1)$ and $\Gamma_{2,T} \triangleq \mathbf{U}_T \cap (|u| > \overline{u}_2)$. With this notation,

$$\mathcal{S}_T(0) \leq \bar{l}_{1,T} \int_{\Gamma_{1,T}} p_T(u)\,du + \int_{\mathbf{U}_T \cap (|u| > \overline{u}_1)} l(u)\, p_T(u)\,du.$$

Furthermore, if $l \in \boldsymbol{L}$, then for sufficiently large B the following relationships hold: (i) $l(u) - \inf_{|v|>B/2} l(v) \leq 0$; (ii) $|u| \leq (B/2)^{\vartheta}$, $\vartheta > 0$. We shall assume that B is chosen so that $B > 2\overline{u}_2$ and $(B/2)^{\vartheta} > \overline{u}_2$ hold. Let $\Gamma_{T,B} \triangleq \{u : (|u| > B/2) \cap \mathbf{U}_T\}$. Then, whenever $|s| > B$ and $|u| \leq B/2$, we have,

$$|u - s| > B/2 > \overline{u}_2 \qquad \text{and} \qquad \inf_{u \in \Gamma_{T,B}} l(u) \geq \bar{l}_{2,T}. \tag{A.51}$$

With this notation,

$$\begin{aligned}
\inf_{|s|>B} \mathcal{S}_T(s) &\geq \inf_{u \in \Gamma_{T,B}} l_T(u) \int_{(|w| \leq B/2) \cap \mathbf{U}_T} p_T(w)\,dw \\
&\geq \bar{l}_{2,T} \int_{(|w| \leq B/2) \cap \mathbf{U}_T} p_T(w)\,dw,
\end{aligned}$$

from which it follows that

$$\mathcal{S}_T(0) - \inf_{|s|>B} \mathcal{S}_T(s) \leq -\varpi \int_{\Gamma_{1,T}} p_T(u)\,du + \int_{\mathbf{U}_T \cap \left((B/2)^{\vartheta} \geq |u| \geq \overline{u}_1\right)} \left(l(u) - \inf_{|s|>B/2} l_T(s) \right) p_T(u)\,du$$
$$+ \int_{\mathbf{U}_T \cap \left(|u| > (B/2)^{\vartheta}\right)} l(u)\,p_T(u)\,du,$$

where $\varpi \triangleq \bar{l}_{2,T} - \bar{l}_{1,T}$. The last inequality can be manipulated further using (A.51), so that

$$\mathcal{S}_T(0) - \inf_{|s|>B} \mathcal{S}_T(s) \leq -\varpi \int_{\Gamma_{1,T}} p_T(u)\,du + \int_{\mathbf{U}_T \cap \left(|u| > (B/2)^{\vartheta}\right)} l_T(u)\,p_T(u)\,du. \tag{A.52}$$

Let $B_{\vartheta} \triangleq (B/2)^{\vartheta}$ and fix an arbitrary number $\overline{a} > 0$. For the first term of (A.52), Lemma A.26 implies that for sufficiently large T, we have

$$\mathbb{P}\left[\int_{\Gamma_{1,T}} p_T(u)\,du < 2\left(\varpi B^{\overline{a}}\right)^{-1} \right] \leq c\left(\varpi B^{\overline{a}}\right)^{-1/2}, \tag{A.53}$$

where $0 < c < \infty$. Next, let us consider the second term of (A.52). We show that for large enough T, an arbitrary number $\overline{a} > 0$,

$$\mathbb{P}\left[\int_{\mathbf{U}_T \cap \{|u| > B_{\vartheta}\}} l(u)\,p_T(u)\,du > B^{-\overline{a}} \right] \leq cB^{-\overline{a}}. \tag{A.54}$$

Since $l \in \boldsymbol{L}$, we have $l(u) \leq |u|^a$, $a > 0$ when u is large enough. Choosing B large leads to

$$\mathbb{E}\left[\int_{\mathbf{U}_T \cap \{|u| > B_{\vartheta}\}} l(u)\,p_T(u)\,du \right] \leq \sum_{i=0}^{\infty} (B_{\vartheta} + i + 1)^a\, \mathbb{E}\left(J_{1,B_{\vartheta}+i}/J_2\right),$$

where $J_{1,B_{\vartheta}+i}$, J_2 are defined as in (A.33). By Lemma A.28,

$$\mathbb{E}\left(J_{1,B_{\vartheta}+i}/J_2\right) \leq c\left(1 + (B_{\vartheta} + i)^a\right) \exp\left(-bf_T(B_{\vartheta} + i)\right),$$

where $f_T \in \boldsymbol{F}$ and thus for some for some b, $0 < c < \infty$,

$$\mathbb{E}\left[\int_{\mathbf{U}_T \cap \{|u| > B_{\vartheta}\}} l(u)\,p_T(u)\,du \right] \leq c \int_{B_{\vartheta}}^{\infty} (1 + v^a) \exp\left(-bf_T(v)\right) dv \leq c \exp\left(-bf_T(B_{\vartheta})\right).$$

By property (ii) of the function f_T in the class $\boldsymbol{F}$, for any $d \in \mathbb{R}$, $\lim_{v\to\infty} \lim_{T\to\infty} v^d e^{-bf_T(v)} = 0$. Thus, we know that for T large enough and some $0 < c < \infty$,

$$\mathbb{E}\left[\int_{\mathbf{U}_T \cap \{|u| > B_{\vartheta}\}} l(u)\,p_T(u)\,du \right] \leq cB^{-2\overline{a}},$$

from which we deduce (A.54) after applying Markov's inequality. Therefore, for sufficiently large T and large B, combining equation (A.50), and (A.53)-(A.54), we have

$$\mathbb{P}\left[T\left\|\delta_T\right\|^2 \left(\widehat{\lambda}_b^{\mathrm{GL}} - \lambda_b^0\right) > B \right]$$
$$\leq \mathbb{P}\left[-\varpi \int_{\Gamma_{1,T}} p_T(u)\,du + \int_{\mathbf{U}_T \cap \{|u| > B_{\vartheta}\}} l_T(u)\,p_T(u)\,du \leq 0 \right]$$

$$\leq \mathbb{P}\left[\int_{\Gamma_{1,T}} p_T(u)\,du < 2\left(\varpi B^{\overline{a}}\right)^{-1}\right] + \mathbb{P}\left[\int_{\mathbf{U}_T\cap\{|u|>B_\vartheta\}} l(u)\,p_T(u)\,du > B^{-\overline{a}}\right]$$
$$\leq c\left(B^{-\overline{a}/2} + B^{-\overline{a}}\right),$$

which can be made arbitrarily small choosing B large enough. □

Lemma A.30. *As $T\to\infty$, the marginal distributions of $\zeta_{T,v}(u,\,\widetilde{v})$ converge to the marginal distributions of $\exp(\mathscr{V}(u))$.*

Proof. The results follows from Lemma A.3, Lemma A.12 and the continuous mapping theorem. □

A.4 Proofs of Section 4

A.4.1 Proof of Proposition 4.1

The preliminary lemmas below consider the Gaussian process $\mathscr{W}$ on the positive half-line with $s>0$. The case $s\leq 0$ is similar and omitted. The generic constant $C>0$ used in the proofs of this section may change from line to line.

Lemma A.31. *For $\varpi>3/4$, we have $\lim_{T\to\infty}\limsup_{|s|\to\infty}\left|\widehat{\mathscr{W}}_T(s)\right|/|s|^{\varpi}=0$, $\mathbb{P}$-a.s.*

Proof. For any $\epsilon>0$, if we can show that

$$\sum_{i=1}^{\infty}\mathbb{P}\left[\sup_{i-1\leq|s|<i}\left|\widehat{\mathscr{W}}_T(s)\right|/|s|^{\varpi}>\epsilon\right]<\infty, \tag{A.55}$$

then by the Borel-Cantelli lemma, $\mathbb{P}\left[\limsup_{|s|\to\infty}\left|\widehat{\mathscr{W}}_T(s)\right|/|s|^{\varpi}>\epsilon\right]=0$. Proceeding as in the proof of Lemma A.13,

$$\mathbb{P}\left[\sup_{i-1\leq|s|<i}\left|\widehat{\mathscr{W}}_T(s)\right|/|s|^{\varpi}>\epsilon\right]\leq\mathbb{P}\left[\sup_{|s|\leq1}\left|\widehat{\mathscr{W}}_T(s)\right|>\epsilon i^{\varpi-1/2}\right]$$
$$\leq\frac{1}{\epsilon^4}\mathbb{E}\left[\mathbb{E}\left(\sup_{|s|\leq1}\left(\widehat{\mathscr{W}}_T(s)\right)^4\mid\widehat{\Sigma}_T\right)\right]\frac{1}{i^{4\varpi-2}}.$$

The series $\sum_{i=1}^{\infty}i^{-p}$ is a Riemann's zeta function and satisfies $\sum_{i=1}^{\infty}i^{-p}<\infty$ if $p>1$. Then,

$$\sum_{i=1}^{\infty}\mathbb{P}\left[\sup_{i-1\leq|s|<i}\left|\widehat{\mathscr{W}}_T(s)\right|/|s|^{\varpi}>\epsilon\right]\leq\left(C/\epsilon^4\right)\mathbb{E}\left[\mathbb{E}\left(\sup_{|s|\leq1}\left(\widehat{\mathscr{W}}_T(s)\right)^4\mid\widehat{\Sigma}_T\right)\right]$$
$$\leq\left(C/\epsilon^4\right)\mathbb{E}\left[\mathbb{E}\left(\sup_{|s|\leq1}\widehat{\mathscr{W}}_T(s)\mid\widehat{\Sigma}_T\right)\right]^4, \tag{A.56}$$

where $C>0$ and the last inequality follows from Proposition A.2.4 in van der Vaart and Wellner (1996). The process $\widehat{\mathscr{W}}_T$, conditional on $\widehat{\Sigma}_T$, is sub-Gaussian with respect to the semimetric $d_{VW}^2(t,\,s)=\widehat{\Sigma}_T(t,\,t)+\widehat{\Sigma}_T(s,\,s)$, which by invoking Assumption 4.1-(ii,iii) is bounded by

$$\widehat{\Sigma}_T(t-s,\,t-s)\leq|t-s|\sup_{|s|=1}\widehat{\Sigma}_T(s,\,s).$$

Theorem 2.2.8 in van der Vaart and Wellner (1996) then implies

$$\mathbb{E}\left(\sup_{|s|\leq1}\widehat{\mathscr{W}}_T(s)\mid\widehat{\Sigma}_T\right)\leq C\sup_{|s|=1}\widehat{\Sigma}_T^{1/2}(s,\,s).$$

The above inequality can be used into the right-hand side of (A.56) to deduce that the latter is bounded by $C\mathbb{E}\left(\sup_{|s|=1}\widehat{\Sigma}_T^2(s,\,s)\right)$. By Assumption 4.1-(iv) $C\mathbb{E}\left(\sup_{|s|=1}\widehat{\Sigma}_T^2(s,\,s)\right)<\infty$, and the proof is concluded. □

Lemma A.32. *$\left\{\widehat{\mathscr{W}_T}\right\}$ converges weakly toward $\mathscr{W}$ on compact subsets of $\mathbb{D}_b$.*

Proof. By the definition of $\widehat{\mathscr{W}_T}(\cdot)$, we have the finite-dimensional convergence in distribution of $\widehat{\mathscr{W}_T}$ toward $\mathscr{W}$. Hence, it remains to show the (asymptotic) stochastic equicontinuity of the sequence of processes $\left\{\widehat{\mathscr{W}_T},\,T\geq1\right\}$. Let $\mathbf{C}\subset\mathbb{R}_+$ be any compact set. Fix any $\eta>0$ and $\epsilon>0$. We show that for any positive sequence $\{d_T\}$, with $d_T\downarrow0$, and for every $t,\,s\in\mathbf{C}$,

$$\limsup_{T\to\infty}\mathbb{P}\left(\sup_{|t-s|<d_T}\left|\widehat{\mathscr{W}_T}(t)-\widehat{\mathscr{W}_T}(s)\right|>\eta\right)<\epsilon. \tag{A.57}$$

By Markov's inequality, $\mathbb{P}\left(\sup_{|t-s|<d_T}\left|\widehat{\mathscr{W}_T}(t)-\widehat{\mathscr{W}_T}(s)\right|>\eta\right)\leq\mathbb{E}\left(\sup_{|t-s|<d_T}\left|\widehat{\mathscr{W}_T}(t)-\widehat{\mathscr{W}_T}(s)\right|\right)/\eta$. Let $\widehat{\Upsilon}_T(t,\,s)$ denote the covariance matrix of $\left(\widehat{\mathscr{W}_T}(t),\,\widehat{\mathscr{W}_T}(s)\right)'$ and $\mathcal{N}$ be a two-dimensional standard normal vector. Letting $\imath\triangleq\begin{bmatrix}1 & -1\end{bmatrix}'$, we have

$$\begin{aligned}\left[\mathbb{E}\sup_{|t-s|<d_T}\left|\widehat{\mathscr{W}_T}(t)-\widehat{\mathscr{W}_T}(s)\right|\right]^2&=\left[\mathbb{E}\sup_{|t-s|<d_T}\left|\imath'\widehat{\Upsilon}_T^{1/2}(t,\,s)\,\mathcal{N}\right|\right]^2\leq\mathbb{E}\left[\sup_{|t-s|<d_T}\imath'\widehat{\Upsilon}_T(t,\,s)\,\imath\right]\\&=\mathbb{E}\left[\sup_{|t-s|<d_T}\widehat{\Sigma}_T(t-s,\,t-s)\right]\\&\leq d_T\mathbb{E}\left[\sup_{|s|=1}\widehat{\Sigma}_T(s,\,s)\right],\end{aligned}$$

and so $\mathbb{E}\left[\sup_{|t-s|<d_T}\widehat{\Sigma}_T(t-s,\,t-s)\right]\leq2d_T\mathbb{E}\left[\sup_{|s|=1}\widehat{\Sigma}_T(s,\,s)\right]$ where we have used Assumption 4.1-(iii) in the last step. As $d_T\downarrow0$ the right-hand side goes to zero since $\mathbb{E}\left[\sup_{|s|=1}\widehat{\Sigma}_T(s,\,s)\right]=O(1)$ by Assumption 4.1-(iv). □

Lemma A.33. *Fix $0<a<\infty$. For any $p\in\boldsymbol{P}$ and for any positive sequence $\{a_T\}$ satisfying $a_T\overset{\mathbb{P}}{\to}a$,*

$$\int_{\mathbb{R}}|p(s)|\exp\left(\widehat{\mathscr{W}_T}(s)\right)\exp\left(-a_T|s|\right)ds\overset{d}{\to}\int_{\mathbb{R}}|p(s)|\exp\left(\mathscr{W}(s)\right)\exp\left(-a|s|\right)ds.$$

Proof. Let $\mathbf{B}_+$ be a compact subset of $\mathbb{R}_+/\{0\}$. Let

$$\boldsymbol{G}=\left\{(W,\,a_T)\in\mathbb{D}_b(\mathbb{R},\,\mathscr{B},\,\mathbb{P})\times\mathbf{B}_+:\,\limsup_{|s|\to\infty}|W(s)|/|s|^{\varpi}=0,\,\varpi>3/4,\,a_T=a+o_{\mathbb{P}}(1)\right\},$$

and denote by $f:\boldsymbol{G}\to\mathbb{R}$ the functional given by $f(\boldsymbol{G})=\int|p(s)|\exp(W(s))\exp(-a_T|s|)\,ds$. In view of the continuity of $f(\cdot)$ and $a_T\overset{\mathbb{P}}{\to}a$, the claim of the lemma follows by Lemma A.31-A.32 and the continuous mapping theorem. □

We are now in a position to conclude the proof of Proposition 4.1. Suppose $\gamma_T=CT\left\|\widehat{\delta}_T\right\|^2$ for some $C>0$. Under mean-squared loss function, $\widehat{\xi}_T$ admits a closed form:

$$\widehat{\xi}_T=\frac{\int u\exp\left(\widehat{\mathscr{W}_T}(u)-\widehat{\Lambda}_T(u)\right)du}{\int\exp\left(\widehat{\mathscr{W}_T}(u)-\widehat{\Lambda}_T(u)\right)du}.$$

By Lemma A.33, we deduce that $\widehat{\xi}_T$ converges in law to the distribution stated in (3.12). For general loss functions, a result corresponding to Lemma A.33 can be shown to hold since $l(\cdot)$ is assumed to be continuous.

A.5 Proofs of Section 5

Rewrite the GL estimator $\widehat{\boldsymbol{\lambda}}_b^{\mathrm{GL}}$ as the minimizer of

$$\mathcal{R}_{l,T} \triangleq \int_{\Gamma^0} l\left(s - \boldsymbol{\lambda}_b\right) \frac{\exp\left(-Q_T\left(\delta\left(\boldsymbol{\lambda}_b\right), \boldsymbol{\lambda}_b\right)\right)\pi\left(\boldsymbol{\lambda}_b\right)}{\int_{\Gamma^0}\exp\left(-Q_T\left(\delta\left(\boldsymbol{\lambda}_b\right), \boldsymbol{\lambda}_b\right)\right)\pi\left(\boldsymbol{\lambda}_b\right)d\boldsymbol{\lambda}_b} d\boldsymbol{\lambda}_b. \tag{A.58}$$

We show with the following lemma that, for each i, $\widehat{\lambda}_i^{\mathrm{GL}} \xrightarrow{\mathbb{P}} \lambda_i^0$ no matter whether the magnitude of the shifts is fixed or not. Then, the proof of Theorem 3.2 can be repeated for each $i = 1, \ldots, m$ separately. We begin with the proof for the case of fixed shifts.

Lemma A.34. *Under Assumption 5.1-5.2, except that $\Delta_{T,i} = \Delta_i^0$ for all i, for $l \in \boldsymbol{L}$ and any $B > 0$ and $\varepsilon > 0$, we have for all large T, $\mathbb{P}\left[\left|\widehat{\lambda}_i^{\mathrm{GL}} - \lambda_i^0\right| > B\right] < \varepsilon$ for each i.*

Proof. Let $S_T\left(\delta\left(\boldsymbol{\lambda}_b\right), \boldsymbol{\lambda}_b\right) \triangleq Q_T\left(\delta\left(\boldsymbol{\lambda}_b\right), \boldsymbol{\lambda}_b\right) - Q_T\left(\delta\left(\boldsymbol{\lambda}_b^0\right), \boldsymbol{\lambda}_b^0\right)$. Without loss of generality, we assume there are only three change-points and provide a proof by contradiction for the consistency result. In particular, we suppose that all but the second change-point are consistently estimated. That is, consider the case $T_2 < T_2^0$ and for some finite $C > 0$ assume that $|\lambda_2 - \lambda_2^0| > C$. $Q_T\left(\delta\left(\boldsymbol{\lambda}_b\right), \boldsymbol{\lambda}_b\right)$ can be decomposed as,

$$Q_T\left(\delta\left(\boldsymbol{\lambda}_b\right), \boldsymbol{\lambda}_b\right) = \sum_{t=1}^T e_t^2 + \sum_{t=1}^T d_t^2 - 2\sum_{t=1}^T e_t d_t,$$

where $d_t = w_t'\left(\widehat{\phi} - \phi^0\right) + z_t'\left(\widehat{\delta}_k - \delta_j^0\right)$, for $t \in \left[\widehat{T}_{k-1} + 1, \widehat{T}_k\right] \cap \left[T_{j-1}^0 + 1, T_j^0\right]$ $(k, j = 1, \ldots, m+1)$ where $\widehat{\phi}$ and $\widehat{\delta}_k$ are asymptotically equivalent to the corresponding least-squares estimates. Bai and Perron (1998) showed that

$$T^{-1}\sum_{t=1}^T d_t^2 \xrightarrow{\mathbb{P}} K > 0 \qquad \text{and} \qquad T^{-1}\sum_{t=1}^T e_t d_t = o_{\mathbb{P}}(1).$$

Note that $Q_T\left(\delta\left(\boldsymbol{\lambda}_b^0\right), \boldsymbol{\lambda}_b^0\right) = S_T\left(T_1^0, T_2^0, T_3^0\right)$, where $S_T\left(T_1^0, T_2^0, T_3^0\right)$ denotes the sum of squared residuals evaluated at $\left(T_1^0, T_2^0, T_3^0\right)$. Since $T^{-1}S_T\left(T_1^0, T_2^0, T_3^0\right)$ is asymptotically equivalent to $T^{-1}\sum_{t=1}^T e_t^2$, this implies that $T^{-1}S_T\left(\delta\left(\boldsymbol{\lambda}_b\right), \boldsymbol{\lambda}_b\right) > 0$ for all large T. For some finite $K > 0$, this implies

$$S_T\left(\delta\left(\boldsymbol{\lambda}_b\right), \boldsymbol{\lambda}_b\right) \geq TK. \tag{A.59}$$

Let $\mathbf{U}_T \triangleq \left\{u \in \mathbb{R} : \boldsymbol{\lambda}_b^0 + T^{-1}u \in \Gamma^0\right\}$. Define $p_T(u) \triangleq p_{1,T}(u)/\overline{p}_T$ where $p_{1,T}(u) = \exp\left(-Q_T\left(\delta(u), u\right)\right)$ and $\overline{p}_T \triangleq \int_{\mathbf{U}_T} p_{1,T}(w)\,dw$. By definition, $\widehat{\boldsymbol{\lambda}}_b^{\mathrm{GL}}$ is the minimum of the function $\int_{\Gamma^0} l(s-u)\,p_{1,T}(u)\,\pi(u)\,du$ with $s \in \Gamma^0$. Upon using a change in variables,

$$\begin{aligned}
&\int_{\Gamma^0} l(s-u)\,p_{1,T}(u)\,\pi(u)\,du \\
&\quad = T^{-1}\overline{p}_T \int_{\mathbf{U}_T} l\left(T\left(s - \boldsymbol{\lambda}_b^0\right) - u\right) p_T\left(\boldsymbol{\lambda}_b^0 + T^{-1}u\right)\pi\left(\boldsymbol{\lambda}_b^0 + T^{-1}u\right)du.
\end{aligned}$$

Thus, $\boldsymbol{\lambda}_{\delta,T} \triangleq T\left(\widehat{\boldsymbol{\lambda}}_b^{\mathrm{GL}} - \boldsymbol{\lambda}_b^0\right)$ is the minimum of the function,

$$\mathcal{S}_T(s) \triangleq \int_{\mathbf{U}_T} l(s-u) \frac{p_T\left(\boldsymbol{\lambda}_b^0 + T^{-1}u\right)\pi\left(\boldsymbol{\lambda}_b^0 + T^{-1}u\right)}{\int_{\mathbf{U}_T} p_T\left(\boldsymbol{\lambda}_b^0 + T^{-1}w\right)\pi\left(\boldsymbol{\lambda}_b^0 + T^{-1}w\right)dw} du,$$

where the optimization is over $\mathbf{U}_T$. As in the proof of Lemma A.8, we exploit the following relationship,

$$\mathbb{P}\left[\left|\widehat{\boldsymbol{\lambda}}_b^{\mathrm{GL}} - \boldsymbol{\lambda}_b^0\right| > B\right] \leq \mathbb{P}\left[\inf_{|s|>TB} \mathcal{S}_T(s) \leq \mathcal{S}_T(0)\right]. \tag{A.60}$$

Thus, we need to show that the random function $\mathcal{S}_T(s)$ is strictly larger than $\mathcal{S}_T(0)$ on $\{|s| > TB\}$ with high probability as $T \to \infty$. The same steps as in Lemma A.8 lead to,

$$\begin{aligned} &\mathcal{S}_T(0) - \inf_{|s|>TB} \mathcal{S}_T(s) \qquad\qquad (\text{A.61}) \\ &\leq -\varpi \int_{\Gamma_{1,T}} p_T(u)\,du + \int_{\mathbf{U}_T \cap \left(|u|>(TB/2)^{\vartheta}\right)} l_T(u)\,p_T(u)\,du. \end{aligned}$$

We can use the relationship (A.59) in place of (A.15) in Lemma A.8 to show that the second term above converges to zero. The first term is negative using the same argument as in Lemma A.8. Thus, $\mathcal{S}_T(0) - \inf_{|s|>TB} \mathcal{S}_T(s) < 0$. This gives a contradiction to the fact that $\widehat{\boldsymbol{\lambda}}_b^{\mathrm{GL}}$ minimizes $\int_{\Gamma^0} l(s-u)\,p_{1,T}(u)\,\pi(u)\,du$. Hence, each change-point is consistently estimated. □

Lemma A.35. *Under Assumption 5.1-5.2, for $l \in \boldsymbol{L}$ and any $B > 0$ and $\varepsilon > 0$, we have for all large T, $\mathbb{P}\left[\left|\widehat{\lambda}_i^{\mathrm{GL}} - \lambda_i^0\right| > B\right] < \varepsilon$ for each i.*

Proof. The structure of the proof is similar to that of Lemma A.34. The difference consists on the fact that now $T^{-1}\sum_{t=1}^T d_t^2 \overset{\mathbb{P}}{\to} 0$ even when a break is not consistently estimated. However, Bai and Perron (1998) showed that $T^{-1}\sum_{t=1}^T d_t^2 > 2T^{-1}\sum_{t=1}^T e_t d_t$ and thus one can proceed as in the aforementioned proof to complete the proof. □

Lemma A.36. *Under Assumption 5.1-5.2, for $l \in \boldsymbol{L}$ and for every $\varepsilon > 0$ there exists a $B < \infty$ such that for all large T, $\mathbb{P}\left[Tv_T^2\left|\widehat{\lambda}_i^{\mathrm{GL}} - \lambda_i^0\right| > B\right] < \varepsilon$ for each i.*

Proof. Let $S_T(\delta(\boldsymbol{\lambda}_b), \boldsymbol{\lambda}_b) \triangleq Q_T(\delta(\boldsymbol{\lambda}_b), \boldsymbol{\lambda}_b) - Q_T\left(\delta\left(\boldsymbol{\lambda}_b^0\right), \boldsymbol{\lambda}_b^0\right)$. Without loss of generality, we assume there are only three change-points and provide an explicit proof only for λ_2^0. We use the same notation as in Bai and Perron (1998), pp. 69-70. Note that their results concerning the estimates of the regression parameters can be used in our context because once we have the consistency of the fractional change-points the estimates of the regression parameters are asymptotically equivalent to the corresponding least-squares estimates. For each $\epsilon > 0$, let $V_\epsilon = \left\{(T_1, T_2, T_3);\left|\widehat{T}_i - T_i^0\right| \leq \epsilon T,\ i = 1 \leq i \leq 3\right\}$. By the consistency result, for each $\epsilon > 0$ and T large, we have $\left|\widehat{T}_i - T_i^0\right| \leq \epsilon T$, where $\widehat{T}_i = \widehat{T}_i^{\mathrm{GL}} = T\widehat{\lambda}_i^{\mathrm{GL}}$. Hence, $\mathbb{P}\left(\left\{\widehat{T}_1, \widehat{T}_2, \widehat{T}_3\right\} \in V_\epsilon\right) \to 1$ with high probability. Therefore we only need to examine the behavior of $S_T(\delta(\boldsymbol{\lambda}_b), \boldsymbol{\lambda}_b)$ for those T_i that are close to the true break dates such that $\left|T_i - T_i^0\right| < \epsilon T$ for all i. By symmetry, we can, without loss of generality, consider the case $T_2 < T_2^0$. For $C > 0$, define

$$V_\epsilon^*(C) = \left\{(T_1, T_2, T_3);\left|\widehat{T}_i - T_i^0\right| < \epsilon T,\ 1 \leq i \leq 3,\ T_2 - T_2^0 < -C/v_T^2\right\}.$$

Define the sum of squared residuals evaluated at (T_1, T_2, T_3) by $S_T(T_1, T_2, T_3)$. Let $SSR_1 = S_T(T_1, T_2, T_3)$, $SSR_2 = S_T(T_1, T_2^0, T_3)$ and $SSR_3 = S_T(T_1, T_2, T_2^0, T_3)$. We have omitted the dependence on δ. With this notation, we have $S_T(\delta(\boldsymbol{\lambda}_b), \boldsymbol{\lambda}_b) = S_T(T_1, T_2, T_3) - S_T(T_1^0, T_2^0, T_3^0)$ which can be decomposed as

$$S_T(\delta(\boldsymbol{\lambda}_b), \boldsymbol{\lambda}_b) = [(SSR_1 - SSR_3) - (SSR_2 - SSR_3)] + \left(SSR_2 - S_T\left(T_1^0, T_2^0, T_3^0\right)\right). \tag{A.62}$$

In their Proposition 4-(ii), Bai and Perron (1998) showed that the first term on the right-hand side above satisfies the following: for every $\varepsilon > 0$, there exists $B > 0$ and $\epsilon > 0$ such that for large T,

$$\mathbb{P}\left[\min\left\{\left[S_T(T_1, T_2, T_3) - S_T\left(T_1, T_2^0, T_3\right)\right] / \left(T_2^0 - T_2\right)\right\} \leq 0\right] < \varepsilon,$$

where the minimum is taken over $V_\epsilon^*(C)$. The second term of (A.62) divided by $T_2^0 - T_2$ can be shown to be negligible for $\{T_1, T_2, T_3\} \in V_\epsilon^*(C)$ and C large enough because on $V_\epsilon^*(C)$ the consistency result guarantees that $\widehat{\lambda}_i$ can be made arbitrary close to λ_i^0. This leads to a result similar to (A.59) where T is replaced by v_T^{-2}. Then one can continue with the same argument used in the second part of the proof of Lemma A.34. □

A.6 Proofs of Section 6

A.6.1 Proof of Proposition 6.1

Let

$$p_{1,T}\left(y|\,\lambda_b^0 + \psi_T^{-1}u\right) \triangleq \exp\left(\left(\widetilde{G}_{T,0}(u, 0) + Q_{T,0}(u)\right)/2\right),$$

where $\widetilde{G}_{T,0}(u, 0)$ and $Q_{T,0}(u)$ were defined in equation (3.7). Let $p_1(y|\,\lambda_b) \triangleq \exp\left(\left(L^2(\lambda_b) - L^2(\lambda_0)\right)/2\right)$ where $L(\lambda_b) = (T_b(T - T_b))^{1/2}\left(\overline{Y}_{T_b}^* - \overline{Y}_{T_b}\right)$ with $\overline{Y}_{T_b} = T_b^{-1}\sum_{t=1}^{T_b} y_t$ and $\overline{Y}_{T_b}^* = (T - T_b)^{-1}\sum_{t=T_b+1}^{T} y_t$. Following Bhattacharya (1994) we use a prior $\breve{\pi}(\cdot)$ on the random variable $\overline{\lambda}_b$. The posterior distribution of $\overline{\lambda}_b = \lambda_b$ is given by $p(\lambda_b|\,y) = h(\lambda_b) / \int_0^1 h(s)\,ds$ where $h(\lambda_b) = p_1(y|\,\lambda_b)\,\breve{\pi}(\lambda_b)$. The total variation distance between two probability measures ν_1 and ν_2 defined on some probability space $S \in \mathbb{R}$ is denoted as $|\nu_1 - \nu_2|_{\text{TV}} \triangleq \int_S |\nu_1(u) - \nu_2(u)|\,du$. Given the local parameter $\lambda_b = \lambda_b^0 + \left(Tv_T^2\right)^{-1}u$ with $u \in [-M, M]$ for a given $M > 0$, the posterior for u is equal to $p^*(u|\,y) = \left(Tv_T^2\right)^{-1} p\left(\left(Tv_T^2\right)^{-1}u + \lambda_b^0|\,y\right)$ while the quasi-posterior is given by $p_T^*(u|\,y) = \left(Tv_T^2\right)^{-1} p_T\left(\left(Tv_T^2\right)^{-1}u + \lambda_b^0|\,y\right)$.

Lemma A.37. *Let Assumption 3.2-3.3 and 3.6-(i) hold and $\breve{\pi}(\cdot)$ satisfy Assumption 3.2. Then,*

$$\left|p_T^*\left(Tv_T^2\left(\overline{\lambda}_b - \lambda_b^0\right)|\,y\right) - p^*\left(Tv_T^2\left(\overline{\lambda}_b - \lambda_b^0\right)|\,y\right)\right|_{\text{TV}} \xrightarrow{\mathbb{P}} 0.$$

Proof. By assumption 3.2, $\pi(\cdot)$ and $\breve{\pi}(\cdot)$ are bounded, and

$$\sup_{|u|\leq M}\left|\pi\left(\left(Tv_T^2\right)^{-1}u + \lambda_b^0\right) - \pi\left(\lambda_b^0\right)\right| \xrightarrow{\mathbb{P}} 0,$$
$$\sup_{|u|\leq M}\left|\breve{\pi}\left(\left(Tv_T^2\right)^{-1}u + \lambda_b^0\right) - \breve{\pi}\left(\lambda_b^0\right)\right| \xrightarrow{\mathbb{P}} 0.$$

Since $\pi(\cdot)$ $[\breve{\pi}(\cdot)]$ appears in both the numerator and denominator of $p_T^*(\cdot|\,y)$ $[p^*(\cdot|\,y)]$, it cancels from that expression asymptotically. Turning to the Laplace estimator, the results of Section 3 (see Lemma A.2 and A.4) imply that for $u \leq 0$, using $Q(\delta(\lambda_b), \lambda_b)/2$ in place of $Q(\delta(\lambda_b), \lambda_b)$,

$$\exp\left(\left(\widetilde{G}_{T,0}(u, 0) + Q_{T,0}(u)\right)/2\right) \tag{A.63}$$

$$= \exp\left(\delta_T \sum_{t=0}^{v_T^{-2}|u|} e_{T_b^0-t} - |u|\,\delta_0^2/2\right)(1+A_T),$$

where $A_T = o_{\mathbb{P}}(1)$ is uniform in the region $u \le \eta T v_T^2$ for small $\eta > 0$. By symmetry, the case $u > 0$ results in the same relationship as (A.63) with $e_{T_b^0-t}$ replaced by $e_{T_b^0+t}$. The results in the proof of Theorem 1 in Bai (1994) combined with the arguments referenced for the derivation of (A.63) suggest that for $u \le 0$,

$$\exp\left(\left(L^2\left(\left(Tv_T^2\right)^{-1} u + \lambda_b^0\right) - L^2\left(\lambda_b^0\right)\right)/2\right) \tag{A.64}$$
$$= \exp\left(\delta_T \sum_{t=0}^{v_T^{-2}|u|} e_{T_b^0-t} - |u|\,\delta_0^2/2\right)(1+B_T),$$

where $B_T = o_{\mathbb{P}}(1)$ is uniform in the region $u \le \eta T v_T^2$ for small $\eta > 0$. By symmetry, the case $u > 0$ results in the same relationship as (A.64) with $e_{T_b^0-t}$ replaced by $e_{T_b^0+t}$. By Lemma A.6 and the results in Bai (1994), $p_T(u|\,y)$ and $p(u|\,y)$ are negligible uniformly in u for $u > \eta T v_T^2$ for every η. Thus, (A.63)-(A.64) yield,

$$\left|p_T^*\left(Tv_T^2\left(\overline{\lambda}_b - \lambda_b^0\right),\, y\right) - p^*\left(Tv_T^2\left(\overline{\lambda}_b - \lambda_b^0\right),\, y\right)\right|_{\mathrm{TV}} \le |A_T| + |B_T| \xrightarrow{\mathbb{P}} 0.$$

□

Continuing with the proof of Proposition 6.1, we begin with part (i). Note that $\varphi(\lambda_b,\, y)$ is defined by

$$\int (1-\varphi(\lambda_b,\, y))\, p_T(y|\,\lambda_b)\, d\mathit{\Pi}(\lambda_b) \ge 1-\alpha$$

for all y, where $\mathit{\Pi}(\cdot)$ is a probability measure on $\mathit{\Gamma}^0$ such that $\mathit{\Pi}(\lambda_b) = \pi(\lambda_b)\, d\lambda_b$. The fact that $|1-\varphi(\lambda_b,\, y)| \le 1$ and Lemma A.37 lead to,

$$\int (1-\varphi(\lambda_b,\, y))\, p_T(y|\,\lambda_b)\, d\mathit{\Pi}(\lambda_b) \tag{A.65}$$
$$= \int (1-\varphi(\lambda_b,\, y))\, p(y|\,\lambda_b)\, d\mathit{\Pi}(\lambda_b) + o_{\mathbb{P}}(1).$$

Given that Definition 4.1 of the GL confidence interval involves an inequality that explicitly allows for conservativeness, (A.65) implies the following relationship,

$$\int \varphi(\lambda_b,\, y)\, p_T(y|\,\lambda_b)\, d\mathit{\Pi}(\lambda_b) = \int \varphi(\lambda_b,\, y)\, p(y|\,\lambda_b)\, d\mathit{\Pi}(\lambda_b) + \varepsilon_T$$
$$\le \alpha \int p(y|\,\lambda_b)\, d\mathit{\Pi}(\lambda_b),$$

where $\varepsilon_T = \int \varphi(\lambda_b,\, y)\,(p_T(y|\,\lambda_b) - p(y|\,\lambda_b))\, d\mathit{\Pi}(\lambda_b)$. Rearranging, we have,

$$\int (\alpha - \varphi(\lambda_b,\, y))\, p(y|\,\lambda_b)\, d\mathit{\Pi}(\lambda_b) - \varepsilon_T \ge 0,$$

for all y. Now multiply both sides by $\widetilde{b}(y) \geq 0$ and integrating with respect to $\zeta(y)$ yields,

$$\int\int (\alpha - \varphi(\lambda_b, y)) \widetilde{b}(y) p(y|\lambda_b) d\zeta(y) d\Pi(\lambda_b) - \varepsilon_T \int \widetilde{b}(y) d\zeta(y) \geq 0,$$

or

$$(1-\alpha) \int L_\alpha\left(\varphi, \widetilde{b}, \lambda_b\right) d\Pi(\lambda_b) - \varepsilon_T \int \widetilde{b}(y) d\zeta(y) \geq 0.$$

Taking the limit as $T \to \infty$,

$$(1-\alpha) \int L_\alpha\left(\varphi, \widetilde{b}, \lambda_b\right) d\Pi(\lambda_b) \geq 0.$$

The latter implies that $L_\alpha\left(\varphi, \widetilde{b}, \lambda_b\right) \geq 0$ for some λ_b. Thus, φ is bet-proof at level $1-\alpha$.

We now prove part (ii). We use a proof by contradiction. If $\int \varphi'(\lambda_b, y) d\lambda_b \geq \int \varphi(\lambda_b, y) d\lambda_b$ for all $y \in \mathcal{Y}$ and $\int \varphi'(\lambda_b, y) d\lambda_b > \int \varphi(\lambda_b, y) d\lambda_b$ for all $y \in \mathcal{Y}_0$ with $\zeta(\mathcal{Y}_0) > 0$, then we show that $\int \varphi'(\lambda_b, y) p(y|\lambda_b) d\zeta(y) > \alpha$ for some $\lambda_b \in \Gamma^0$. By Lemma A.37 and (6.1) holding with equality,

$$\begin{aligned}\int \varphi(\lambda_b, y) p_T(y|\lambda_b) d\Pi(\lambda_b) &= \alpha \int p_T(y|\lambda_b) d\Pi(\lambda_b) \\ &= \alpha \int p(y|\lambda_b) d\Pi(\lambda_b) + o_{\mathbb{P}}(1).\end{aligned}$$

Integrating both sides with respect to $\zeta(y)$ yields,

$$\int \left(\int \varphi(\lambda_b, y) p(y|\lambda_b) d\zeta(y)\right) d\Pi(\lambda_b) = \alpha + o_{\mathbb{P}}(1). \tag{A.66}$$

By Assumption 3.2, $\pi(\lambda_b) > 0$ for all $\lambda_b \in \Gamma^0$. Taking the limit as $T \to \infty$ of both sides of (A.66) yields $\int \left(\int \varphi(\lambda_b, y) p(y|\lambda_b) d\zeta(y)\right) d\Pi(\lambda_b) = \alpha$. The latter holds only if $\int \varphi(\lambda_b, y) p(y|\lambda_b) d\zeta(y) = \alpha$ for all $\lambda_b \in \Gamma^0$. This means that φ is similar. The definition of HPD confidence set $\varphi(\lambda_b, y)$ implies that for ζ-almost all y, if $\int \varphi(\lambda_b, y) d\lambda_b = \int \varphi'(\lambda_b, y) d\lambda_b$ then $\int \varphi(\lambda_b, y) p_T(\lambda_b|y) d\lambda_b \leq \int \varphi'(\lambda_b, y) p_T(\lambda_b|y) d\lambda_b$. The latter relationship and Lemma A.37 imply that,

$$\int \varphi(\lambda_b, y) p(y|\lambda_b) d\Pi(\lambda_b) \leq \int \varphi'(\lambda_b, y) p(y|\lambda_b) d\Pi(\lambda_b),$$

for all $y \in \mathcal{Y}$ and

$$\int \varphi(\lambda_b, y) p(y|\lambda_b) d\Pi(\lambda_b) < \int \varphi'(\lambda_b, y) p(y|\lambda_b) d\Pi(\lambda_b),$$

for all $y \in \mathcal{Y}_0$. Integrating both sides with respect to ζ yields

$$\begin{aligned}&\int \left(\int \varphi(\lambda_b, y) p(y|\lambda_b) d\zeta(y)\right) d\Pi(\lambda_b) \\ &\quad < \int \left(\int \varphi'(\lambda_b, y) p(y|\lambda_b) d\zeta(y)\right) d\Pi(\lambda_b),\end{aligned}$$

or

$$\int \left(\int (\varphi(\lambda_b, y) - \varphi'(\lambda_b, y)) p(y|\lambda_b) d\zeta(y)\right) d\Pi(\lambda_b) < 0.$$

Since $\varphi(\lambda_b, y)$ is similar, there exists a λ_b such that $\int \varphi'(\lambda_b, y)\, p(y|\, \lambda_b)\, d\zeta(y) > \alpha$. Thus, φ' is not of level $1 - \alpha$. □