Econometric Theory, **38**, 2022, 35–65.
doi:10.1017/S0266466621000013

GENERALIZED LAPLACE INFERENCE IN MULTIPLE CHANGE-POINTS MODELS*

ALESSANDRO CASINI
University of Rome Tor Vergata

PIERRE PERRON
Boston University

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, *Econometrica* 66, 47–78). The GL estimator is defined by an integration rather than optimization-based method and relies on the LS 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 good finite-sample performance.

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 transformations of a least-squares (LS) 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

*This paper is based on the fourth chapter of the first author's doctoral dissertation at Boston University. We thank the Editor and a Co-Editor for guiding the review, and three anonymous referees for constructive comments. We also thank Zhongjun Qu for useful comments. Address Correspondence to Alessandro Casini, Department of Economics and Finance, University of Rome Tor Vergata, Via Columbia 2, Rome 00133, Italy; e-mail: alessandro.casini@uniroma2.it.

© The Author(s), 2021. Published by Cambridge University Press

https://doi.org/10.1017/S0266466621000013 Published online by Cambridge University Press

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 LS criterion. In a first step, we use the approach of Bai and Perron (1998) to evaluate the LS 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 (2019) and references therein.

One of our goals is to develop GL estimates with better small-sample properties compared to LS estimates, namely lower mean absolute error (MAE) and root-mean squared error (RMSE), and confidence sets with accurate coverage probabilities and relatively short lengths for a wide range of break sizes, whether small or large; existing methods 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) LS 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 LS criterion function. We derive two distinct limiting distributions corresponding to different smoothing sequences of the criterion function (cf. Jun et al., 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 LS estimator derived in Bai and Perron (1998) (see also Hinkley, 1971; Picard, 1985; 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 (HDRs) introduced by Casini and Perron (2020a) for structural change problems, which best summarizes the properties of the probability distribution of interest. The HDRs are

https://doi.org/10.1017/S0266466621000013 Published online by Cambridge University Press

common in Bayesian analysis where they are applied to a posterior distribution (see, e.g., Box and Tiao, 1973). Kendall and Stuart (1961) discussed the difference between frequentist confidence intervals and Bayesian approaches in relation to the existence of a sufficient statistic. Our 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 LS criterion function. As noted in the literature on likelihood-based inference in some classes of nongranular problems (see, e.g., Chernozhukov and Hong, 2003; Ghosal et al., 1995; Hirano and Porter, 2003; 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 nonregular problems (see, e.g., Buehler, 1959).

https://doi.org/10.1017/S0266466621000013 Published online by Cambridge University Press

The GL inference developed in this paper has been applied by Casini and Perron (2020b) to achieve finite-sample improvements under the continuous record asymptotic framework of Casini and Perron (2020a). The latter proposed an alternative asymptotic framework to explain the nonstandard features of the finite-sample distribution of the LS 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 the Supplementary Material.

2. THE MODEL AND THE ASSUMPTIONS

This section introduces the structural change model with a single break, reviews the LS 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 componentwise. We use $\lfloor \cdot \rfloor$ to denote the largest smaller integer function. We use $\xrightarrow{\mathbb{P}}$ and $\xrightarrow{d}$ 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 specified 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

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

where y_t is a 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 occurs at date T_b^0. It is useful to reparameterize the model. Letting $x_t \triangleq \left(w_t', z_t'\right)'$ and $\beta^0 \triangleq \left(\left(\phi^0\right)', \left(\delta_1^0\right)'\right)'$, we have

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

https://doi.org/10.1017/S0266466621000013 Published online by Cambridge University Press

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 $D = \left(0_{q\times p}, I_{q\times q}\right)'$. In order to facilitate the derivations, we reformulate model (2.2) in matrix format. Let $Y = (y_1, \ldots, y_T)', X = (x_1, \ldots, x_T)', e = (e_1, \ldots, e_T)', X_1 = \left(x_1, \ldots, x_{T_b}, 0, \ldots, 0\right)'$, $X_2 = (0, \ldots, 0, x_{T_b+1}, \ldots, x_T)'$, 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: $Z_1 = X_1 D$, $Z_2 = X_2 D$, and $Z_0 = X_0 D$. 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 $\theta^0 \triangleq \left(\left(\phi^0\right)', \left(\delta_1^0\right)', \left(\delta^0\right)'\right)'$ denote the true value of the parameter vector $\theta \triangleq (\phi, \delta_1, \delta)$. The break date 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 $T_b = T\lambda_b$, i.e., $\widehat{T}_b^{\mathrm{LS}} = \arg\min_{1\le T_b\le T} S_T(\widehat{\theta}^{\mathrm{LS}}(T_b), T_b)$, where $\widehat{\theta}^{\mathrm{LS}}(T_b) = \arg\min_\theta S_T(\theta, T_b)$ with $S_T(\theta, T_b) = \sum_{t=1}^{T_b}\left(y_t - \phi' w_t - \delta_1' z_t\right)^2 + \sum_{t=T_b+1}^{T}\left(y_t - \phi' w_t - \delta' z_t\right)^2$. Also,

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

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

Assumption 2.1. $T_b^0 = \left\lfloor T\lambda_b^0 \right\rfloor$, 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, $\{z_t e_t, \mathscr{F}_t\}$ forms an L^r-mixingale sequence with $r = 2+\nu$, for some $\nu > 0$. That is, there exist nonnegative constants $\left\{\varrho_{1,t}\right\}_{t\ge1}$ and $\left\{\varrho_{2,j}\right\}_{j\ge0}$ such that $\varrho_{2,j} \to 0$ as $j \to \infty$, and for all $t \ge 1$, $j \ge 0$ and $r \ge 1$, (i) $\left\|\mathbb{E}\left(z_t e_t | \mathscr{F}_{t-j}\right)\right\|_r \le \varrho_{1,t}\varrho_{2,j}$, (ii) $\left\|z_t e_t - \mathbb{E}\left(z_t e_t | \mathscr{F}_{t+j}\right)\right\|_r \le \varrho_{1,t}\varrho_{2,j+1}$. 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 $l \ge p+q$ 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.

https://doi.org/10.1017/S0266466621000013 Published online by Cambridge University Press

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, 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 independent and identically distributed (i.i.d.) Gaussian process with a mean change). 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 LS criterion function $Q_T(\delta(\lambda_b), \lambda_b)$, with the parameter of interest being $\lambda_b^0 = T_b^0/T$. The Quasi-posterior $p_T(\lambda_b)$ is defined by the exponential transformation

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

where $\pi(\cdot)$ is a density 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 $m = 1$), and (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

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

https://doi.org/10.1017/S0266466621000013 Published online by Cambridge University Press

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., $\widehat{\lambda}_b^{\mathrm{GL}} \triangleq \arg\min_{s\in\Gamma^0}\left[\mathcal{R}_{l,T}(s)\right]$. 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 (2020a). 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 nonregular 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 $r = 0$, *and* (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 $\pi^0 \triangleq \pi\left(\lambda_b^0\right) > 0$, 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 apply some basic smoothing techniques to make it differentiable locally (e.g., mean smoothing, Gaussian smoothing, and Savitzky–Golay filter). We did not find any particular difference in the empirical results, and so we used the mean smoothing. The assumption on the differentiability of the kernel can be relaxed at the expense of one more step in the proof. Chernozhukov and Hong (2003) assumed differentiability of the prior; we also keep the same assumption and applied the smoothing. 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.

https://doi.org/10.1017/S0266466621000013 Published online by Cambridge University Press

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 $\widehat{\delta}_T^{\mathrm{LS}} - \delta_T = o_{\mathbb{P}}(1)$.

3.2. Discussion About the GL Approach

We use Figures 1 and 2 to illustrate the main idea behind the usefulness of the GL method. They present plots of the density of the distribution of $\widehat{T}_b^{\mathrm{LS}}$ and $\widehat{T}_b^{\mathrm{GL}}$ for the simple model $y_t = \phi^0 + z_t\left(\delta_1^0 + \delta^0 \mathbf{1}\left\{t > T_b^0\right\}\right) + e_t$, where $\{z_t\}$ follows an ARMA(1,1) process and $e_t \sim i.i.d. \mathcal{N}(0, 1)$. The distributions presented are the exact finite-sample distributions of the LS and GL estimators, Bai's (1997)

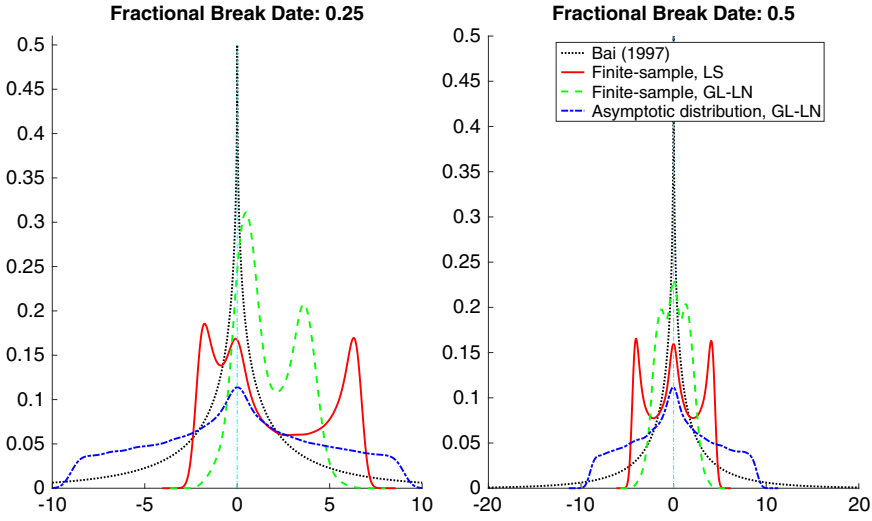


Figure 1. The probability density of the LS estimator for the model $y_t = \mu^0 + Z_t\delta_1^0 + Z_t\delta_2^0 \mathbf{1}\{t > \lfloor T\lambda_0 \rfloor\} + e_t$, $Z_t = 0.3Z_{t-1} + u_t - 0.1u_{t-1}$, $u_t \sim$ i.i.d. $\mathcal{N}(0, 1)$, $e_t \sim$ i.i.d. $\mathcal{N}(0, 1)$, $\{u_t\}$ independent from $\{e_t\}$, $T = 100$ with $\delta^0 = 0.3$, and $\lambda_0 = 0.25$ and 0.5 (the left and right panels, respectively). The black dotted line is the density of the asymptotic distribution from Bai's (1997), the red broken line break is the density of the finite-sample distribution of the LS estimator, the green broken line is the density of the finite-sample distribution of the GL estimator, and the blue broken line is the density of the asymptotic distribution of the GL estimator.

https://doi.org/10.1017/S0266466621000013 Published online by Cambridge University Press

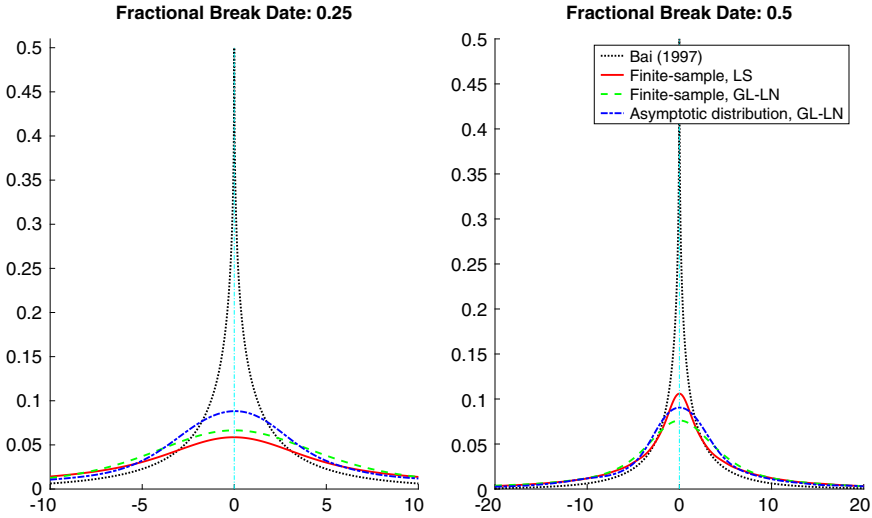


FIGURE 2. The descriptions and comments given in Figure 1 apply but with a break magnitude $\delta^0 = 1.5$.

classical large-N limit distribution, and the asymptotic distribution of the GL estimator. Noteworthy are the nonstandard features of the finite-sample distribution of the LS estimator when the break magnitude is small, which include multimodality, fat tails, and asymmetry. 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. Some of these features have been found in other works (see, e.g., Perron and Zhu, 2005; Deng and Perron, 2006; Jiang et al., 2018, 2020; Casini and Perron, 2020a). Turning to the densities of the GL estimators, some of the nonstandard features appear also for the GL estimator although to a much lesser extent. In particular, the densities of the GL estimators are less spread out than the corresponding densities for the LS estimator. For small breaks, the finite-sample distributions of the LS and GL estimators are quite different, which suggests that standard measures of accuracy (e.g., MAE and RMSE) can be expected to differ substantially. For $\lambda_0 = 0.5$, the GL estimator exhibits much less variability and more precision. The figures also show that the asymptotic distribution of the GL estimator provides an accurate approximation for large breaks, while for small breaks, the approximation is less accurate. However, it captures the fat-tails of the finite-sample distribution, which suggests that it does not underestimate uncertainty about the break location unlike Bai's (1997) distribution.

The GL method is useful, because it weights the information from the LS criterion function with the information from the prior density—which, here, is

https://doi.org/10.1017/S0266466621000013 Published online by Cambridge University Press

the density of the asymptotic distribution of $\widehat{T}_b^{\mathrm{LS}}$. Note that the LS objective function is quite flat when the magnitude of the break is small, and so $\widehat{T}_b^{\mathrm{LS}}$ is imprecise. The resulting Quasi-posterior, or, e.g., its median, is likely to 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 a smoothing 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 $\widetilde{Q}\left(\theta^0, \lambda_b\right) = \mathbb{E}\left[Q_T\left(\theta^0, \lambda_b\right)\right]$ and that $\theta^0 \triangleq \left(\left(\phi^0\right)', \left(\delta_1^0\right)', \left(\delta^0\right)'\right)' \in \mathbf{S} \subset \mathbb{R}^p \times \mathbb{R}^q \times \mathbb{R}^q$. 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 $\overline{Q}_T(\theta, \lambda_b) \triangleq Q_T(\theta, \lambda_b) - Q_T\left(\theta, \lambda_b^0\right)$, $Q_T^0(\theta, \lambda_b) \triangleq \mathbb{E}\left[Q_T(\theta, \lambda_b) - Q_T\left(\theta, \lambda_b^0\right) | X\right]$, 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

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

and

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

They are derived in Section A.2 of the Supplementary Material. For the purpose of developing the asymptotic theory, the GL estimator $\widehat{\lambda}_b^{\mathrm{GL}}(\theta)$ is defined as the minimizer of a normalized version of $\mathcal{R}_{l,T}(s)$:

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

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.

https://doi.org/10.1017/S0266466621000013 Published online by Cambridge University Press

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 $\gamma_T/\left(T\left\|\delta_T\right\|^2\right)$ 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

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

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 Jurečová (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. 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 $\Psi_{l,T}(s;v,\widetilde{v})$ to the marginal distributions of the random function

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

where

$$\mathscr{V}(s) \triangleq \mathscr{W}(s) - \Lambda^0(s) \triangleq \begin{cases} 2\left(\left(\delta^0\right)'\Sigma_1\delta^0\right)^{1/2} W_1(-s) - |s|\left(\delta^0\right)'V_1\delta^0, & \text{if } s\le 0 \\ 2\left(\left(\delta^0\right)'\Sigma_2\delta^0\right)^{1/2} W_2(s) - s\left(\delta^0\right)'V_2\delta^0, & \text{if } s>0, \end{cases} \tag{3.7}$$

and W_1, W_2 are independent standard Wiener processes defined on $[0,\infty)$. The limit process $\Psi_l^0(s)$ does not depend on v nor $\widetilde{v}$. Next, we show that the family of probability measures in $\mathbb{C}_b(\mathbf{K})$, with $\mathbf{K}\triangleq\{s\in\mathbb{R}:\,|s|\le K \text{ and } K<\infty\}$, generated by the contractions of $\Psi_{l,T}(s;\widetilde{v},v)$ on $\mathbf{K}$ is dense uniformly in $(v,\widetilde{v})$. Finally, we examine the oscillations of the minimizers of the sample criterion $\Psi_{l,T}(s;v,\widetilde{v})$.

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 $\Psi_{l,T}(s) = \Psi_{l,T}(s;0,0)$.

https://doi.org/10.1017/S0266466621000013 Published online by Cambridge University Press

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) LS projections and would not apply, for example, to the least absolute deviation (LAD) estimator of the break date (cf. Bai, 1995) for which $\Psi_{l,T}(s;\widetilde{\nu},\nu)$ 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 et al., 2010; Perron and Yamamoto, 2014, 2015). We establish theoretical results under this more general setting, since they may be useful for future work.

Let $\lambda^0_{b,T}(\nu) = \lambda^0_{b,T}\left(\theta^0+\nu/r_T\right)$. Introduce the local parameter $u = \psi_T\left(\lambda_b-\lambda^0_{b,T}(\nu)\right)$ and let $\pi_{T,\nu}(u) \triangleq \pi\left(\lambda^0_{b,T}(\nu)+u/\psi_T\right)$, $Q_{T,\nu}(u) \triangleq Q^0_T(\theta^0+\nu/r_T, \lambda^0_{b,T}(\nu)+u/\psi_T)$, and $\widetilde{G}_{T,\nu}(u,\widetilde{\nu}) \triangleq G_T\left(\theta^0+\widetilde{\nu}/r_T, \lambda^0_{b,T}(\nu)+u/\psi_T\right)$, where the sequence $\{\psi_T\}$ depends on the results on consistency and rate of convergence of $\widehat{\lambda}^{\mathrm{GL}}_b$ in Proposition 3.1. Apply a simple substitution in (3.6) to yield

$$\Psi_{l,T}(s;\widetilde{\nu},\nu) = \int_{\Gamma_T} l(s-u)\frac{\exp\left(\left(\gamma_T/T\|\delta_T\|^2\right)\left(\widetilde{G}_{T,\nu}(u,\widetilde{\nu})+Q_{T,\nu}(u)\right)\right)\pi_{T,\nu}(u)\,du}{\int_{\Gamma_T}\exp\left(\left(\gamma_T/T\|\delta_T\|^2\right)\left(\widetilde{G}_{T,\nu}(w,\widetilde{\nu})+Q_{T,\nu}(w)\right)\right)\pi_{T,\nu}(w)\,dw}, \tag{3.8}$$

where $\Gamma_T \triangleq \left\{u\in\mathbb{R}:\ \lambda^0_b+u/\psi_T\in\Gamma^0\right\}$.

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

Assumption 3.5. For $r\in[0,1]$, $\left(T^0_b\right)^{-1/2}\sum_{t=1}^{\left\lfloor rT^0_b\right\rfloor} z_te_t \Rightarrow \mathscr{G}_1(r)$ and $\left(T-T^0_b\right)^{-1/2}\sum_{t=T^0_b+1}^{T^0_b+\left\lfloor r\left(T-T^0_b\right)\right\rfloor} z_te_t \Rightarrow \mathscr{G}_2(r)$, 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 $\Sigma_1 \triangleq \lim_{T\to\infty}\mathbb{E}\left[\left(T^0_b\right)^{-1/2}\sum_{t=1}^{T^0_b} z_te_t\right]^2$ and $\Sigma_2 \triangleq \lim_{T\to\infty}\mathbb{E}\left[\left(T-T^0_b\right)^{-1/2}\sum_{t=T^0_b+1}^{T} z_te_t\right]^2$. Furthermore, for any $0<r_0<1$ with $r_0<\lambda_0$, $T^{-1}\sum_{t=\lfloor r_0T\rfloor+1}^{\lfloor\lambda_0T\rfloor} z_tz_t' \xrightarrow{\mathbb{P}} (\lambda_0-r_0)V_1$, and with $\lambda_0<r_0$, $T^{-1}\sum_{t=\lfloor\lambda_0T\rfloor+1}^{\lfloor r_0T\rfloor} z_tz_t' \xrightarrow{\mathbb{P}} (r_0-\lambda_0)V_2$ so that λ_- and λ_+ (the minimum and maximum of the eigenvalues of the last two matrices) satisfy $0<\lambda_-\le\lambda_+<\infty$.

Assumptions 3.4 and 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 $u\le 0$, let $g(\zeta_t;u) \triangleq \left(\delta^0\right)'\sum_{t=T^0_b+\left\lfloor u/v_T^2\right\rfloor}^{T^0_b}\zeta_t$ and $\widetilde{g}(\zeta_t;u,\widetilde{\nu},\nu;\psi_T,r_T) \triangleq$

https://doi.org/10.1017/S0266466621000013 Published online by Cambridge University Press

$\sqrt{\psi_T}\left(\delta^0+\widetilde{v}/r_T\right)'\sum_{t=T\lambda_b^0\left(\theta^0+v/r_T\right)+\lfloor u/\psi_T\rfloor}^{T\lambda_b^0\left(\theta^0+v/r_T\right)}\zeta_t$. 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 $\lambda_b\neq\lambda_b^0$, $\widetilde{Q}\left(\theta^0,\lambda_b\right)<\widetilde{Q}\left(\theta^0,\lambda_b^0\right)$; (ii) for any $v,\widetilde{v}_1,\widetilde{v}_2\in\mathbf{V}$ and $u,s\in\mathbb{R}$,

$$\Sigma(u,s)\triangleq\lim_{T\to\infty}\mathbb{E}\left[\widetilde{g}(\zeta_t;u,\widetilde{v}_1,v;\psi_T,r_T)\widetilde{g}(\zeta_t;s,\widetilde{v}_2,v;\psi_T,r_T)'\right]$$

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 $\Sigma(\cdot,\cdot)$ is strictly positive and that

$$\forall u,s\in\mathbb{R}:\begin{cases}\forall c>0:\ \Sigma(cu,cs)=c\Sigma(u,s),\\ \Sigma(u,u)+\Sigma(s,s)-2\Sigma(u,s)=\Sigma(u-s,u-s),\end{cases}\tag{3.9}$$

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. $\xi_l^0\triangleq\xi\left(\lambda_b^0\right)$ is uniquely defined by

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

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 $\widehat{\lambda}_b^{\mathrm{GL}}=\int_{I^0}\lambda_b p_T(\lambda_b)\,d\lambda_b$, 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 LS 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.

https://doi.org/10.1017/S0266466621000013 Published online by Cambridge University Press

PROPOSITION 3.1. *Under Assumptions 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, $\widehat{\lambda}_b^{\mathrm{GL}}(\widehat{\theta}) \triangleq \widehat{\lambda}_b^{\mathrm{GL},*}(\widetilde{\nu}, \nu)$, where

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

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

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

where again we have used the notation $\pi_{T,\nu}(u) = \pi\left(\lambda_{b,T}^0(\nu) + u/\psi_T\right)$, $Q_{T,\nu}(u) = Q_T^0\left(\theta^0 + \nu/r_T\ \ \lambda_{b,T}^0(\nu) + u/\psi_T\right)$, and $\widetilde{G}_{T,\nu}(u, \widetilde{\nu}) = G_T\left(\theta^0 + \widetilde{\nu}/r_T, \lambda_{b,T}^0(\nu) + u/\psi_T\right)$. The limit of the GL estimator depends on the limit of the process $\left(\gamma_T/\left(T\|\delta_T\|^2\right)\right)\left(\widetilde{G}_{T,\nu}(u, \widetilde{\nu}) + Q_{T,\nu}(u)\right)$. As part of the proof of Theorem 3.1, we show that the sequence of processes $\left\{\widetilde{G}_{T,\nu}(u, \widetilde{\nu}), T \geq 1\right\}$ converges weakly in $\mathbb{D}_b(\mathbb{R} \times \mathbf{V})$ to a Gaussian process $\mathscr{W}$ not varying with ν, whereas $Q_{T,\nu}(\cdot)$ is approximated by a (deterministic) drift process taking negative values, and is monotonic in ν and flat in $\widetilde{\nu}$. We show that this implies that $\widehat{\lambda}_b^{\mathrm{GL},*}(\widetilde{\nu}, \nu) - \lambda_{b,T}^0(\nu)$ is monotonic in ν, which then leads to uniform convergence in ν following the argument of Jurečová (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 [cï¿œdlï¿œg] functions $\mathbb{D}_b$. Let $\mathbf{V} \subset \mathbb{R}^{p+2q}$ be a compact set. Let $W_T(u, \widetilde{\nu}, \nu)$ denote an arbitrary sample process with bounded [cï¿œdlï¿œg] paths evaluated at the local parameters $u \in \mathbb{R}$, and $\nu, \widetilde{\nu} \in \mathbf{V}$. For each fixed $\nu \in \mathbf{V}$, we shall write $W_T(u, \widetilde{\nu}, \nu) \Rightarrow \mathscr{W}(u, \widetilde{\nu}, \nu)$ in $\mathbb{D}_b(\mathbb{R} \times \mathbf{V})$ whenever the process $W_T(\cdot, \cdot, \nu)$ converges weakly to $\mathscr{W}(\cdot, \cdot, \nu)$, where $\mathscr{W}(\cdot, \cdot, \nu)$ also belongs to $\mathbb{D}_b(\mathbb{R} \times \mathbf{V})$. As a shorthand, we shall omit the argument u ($\widetilde{\nu}$) if the limit process does not depend on u ($\widetilde{\nu}$). The same notational

https://doi.org/10.1017/S0266466621000013 Published online by Cambridge University Press

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 Assumptions 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 (3.7).

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 equation (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 (and exactly so in the i.i.d. case with stationary regimes) 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 Assumptions 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 $\mathbb{D}_b$, with $\mathscr{V}(\cdot)$ defined in (3.7).*

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

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

Corollary 3.1 and Proposition 3.2 show that with enough smoothing applied, the GL estimator is (first-order) asymptotically equivalent to the LS or MLE (cf. Bai, 1997; Yao, 1987, respectively). The intuition is that when the criterion function is sufficiently smoothed, the Quasi-posterior probability density converges to the

https://doi.org/10.1017/S0266466621000013 Published online by Cambridge University Press

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 $T\|\delta_T\|^2 \left(\widehat{\lambda}_b^{\mathrm{GL}} - \lambda_b^0\right)$ is (first-order) asymptotically equivalent to ξ_l^0 defined by

$$\Psi_l\left(\xi_l^0\right) \triangleq \inf_r \Psi_l(r) = \inf_{r \in \mathbb{R}} \left\{ \int_{\mathbb{R}} l(r-u) p_0^*(u) du \right\}. \tag{3.13}$$

THEOREM 3.2. *Under Assumptions 2.1–2.4 and 3.1–3.7, and Condition 1, for* $l \in \boldsymbol{L}$, $T\|\delta_T\|^2 (\widehat{\lambda}_b^{\mathrm{GL}} - \lambda_b^0) \Rightarrow \xi_l^0$ *as defined by (3.13).*

The existence and uniqueness of ξ_l^0 follow from Assumption 3.7. If one interprets $p_0^*(u)$ as a true posterior density function, then ξ_l^0 would naturally be viewed as a Bayesian estimator for the loss function $l_T(\cdot)$. 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 LAD loss, the GL estimator converges in distribution to the median of $p_0^*(u)$. We shall use the results in Theorems 3.1 and 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 $\arg\max_{s \in \mathbb{R}} \mathscr{V}(s)$, which underestimates uncertainty as shown by confidence intervals 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 work 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.

https://doi.org/10.1017/S0266466621000013 Published online by Cambridge University Press

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; 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 $\widehat{\lambda}_{b,T} = \lambda_0 + o_{\mathbb{P}}(1)$, $\widehat{\delta}_T = \delta_T + o_{\mathbb{P}}(1)$, $\widehat{\Xi}_{Z,T} = \Xi_Z + o_{\mathbb{P}}(1)$, and $\widehat{\Xi}_{e,T} = \Xi_e + o_{\mathbb{P}}(1)$. Furthermore, for all $u, s \in \mathbb{R}$ and any $c > 0$, there exist covariation processes $\widehat{\Sigma}_{i,T}(\cdot)$ $(i = 1, 2)$ 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) $\widehat{\Sigma}_{i,T}(cu, cu) = c\widehat{\Sigma}_{i,T}(u, u)$, and (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). Parts (ii) and (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) and (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 analog 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. The finite-sample distribution of the LS and GL estimate of the break date displays significant nonstandard 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 (2020a), we use the concept of highest Quasi-posterior density (HQPD) regions, defined analogously to the HDR (cf. Hyndman, 1996). See also Samworth and Wand (2010) and Mason and Polonik (2008, 2009) for more recent developments. For an illustrative example on the properties of the HDR, see the discussion of Figure 11 in Casini and Perron (2020a).

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 $(1-\alpha)100\%$ HDR 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 $\mathbb{P}_Y(Y \in \mathbf{S}(\kappa_\alpha)) \geq 1-\alpha$.*

https://doi.org/10.1017/S0266466621000013 Published online by Cambridge University Press

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 LS 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 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 $\left(T_1^0, \ldots, T_m^0\right)$ 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 $\theta \triangleq \left(\phi', \delta_1', \Delta_1' \ldots, \Delta_m'\right)'$, where $\Delta_i = \delta_{i+1} - \delta_i$, $i = 1, \ldots, m$. The class $\mathscr{L}(\theta, T_i; 1 \le i \le m)$ of GL estimators in multiple change-points models relies on the LS criterion function $Q_T(\delta(\boldsymbol{\lambda}_b), \boldsymbol{\lambda}_b) = \sum_{i=1}^{m+1}\sum_{t=T_i-1}^{T_i}\left(y_t - w_t'\phi - z_t'\delta_i\right)^2$, with $\boldsymbol{\lambda}_b \triangleq (\lambda_i; 1 \le i \le m)$. In order to state the large-sample properties, we need to introduce the shrinkage theoretical framework of Bai and Perron (1998).

https://doi.org/10.1017/S0266466621000013 Published online by Cambridge University Press

Assumption 5.1. (i) Let $x_t = \left(w_t', z_t'\right)'$, $X = (x_1, \ldots x_T)'$, and $\overline{X}_0 = \text{diag}\left(X_1^0, \ldots, X_{m+1}^0\right)$ be the diagonal partition of X at $\left(T_1^0, \ldots, T_m^0\right)$. For each $i = 1, \ldots, m+1$, $\left(X_1^0\right)' X_1^0 / \left(T_i^0 - T_{i-1}^0\right)$ converges to a nonrandom 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) $T_i^0 = \left\lfloor T\lambda_i^0 \right\rfloor$, 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 $\mathbb{E}\|z_t\|^2 < C$ and $\mathbb{E}\|e_t\|^{2/\vartheta} < C$, 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) $\left(\Delta T_i^0\right)^{-1} \sum_{t=T_{i-1}^0+1}^{T_{i-1}^0+\left\lfloor s\Delta T_i^0 \right\rfloor} z_t z_t' \xrightarrow{\mathbb{P}} s V_i$, $\left(\Delta T_i^0\right)^{-1} \sum_{t=T_{i-1}^0+1}^{T_{i-1}^0+\left\lfloor s\Delta T_i^0 \right\rfloor} e_t^2 \xrightarrow{\mathbb{P}} s\sigma_i^2$, and

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

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

Next, for $i = 1, \ldots, m$, define $\Xi_{Z,i} = \left(\Delta_i^0\right)' V_{i+1} \Delta_i^0 / \left(\Delta_i^0\right)' V_i \Delta_i^0$ and $\Xi_{e,i}^2 = \left(\Delta_i^0\right)' \Sigma_{i+1} \Delta_i^0 / \left(\Delta_i^0\right)' \Sigma_i \Delta_i^0$, 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

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

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)), $\widehat{\boldsymbol{\lambda}}_b^{\text{GL}} = \arg\min_{s \in \Gamma^0}\left[\mathcal{R}_{l,T}(s)\right]$, 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. Assumptions 3.1 and 3.2 hold with obvious modifications to allow for the multidimensional parameter $\boldsymbol{\lambda}_b \in \Gamma^0$. 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 $\xi_{l,i}^0 \triangleq \xi\left(\lambda_i^0\right)$ 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

https://doi.org/10.1017/S0266466621000013 Published online by Cambridge University Press

estimator is defined as the minimizer of

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

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 $\lambda_{i,T}^0(v) = \lambda_{i,T}^0\left(\theta^0 + v/r_T\right)$.

Theorems 5.1 and 5.2 extend corresponding results from Theorems 3.1 and 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 are 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 Assumptions 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 $T\|\delta_T\|^2 \left(\widehat{\lambda}_i^{\mathrm{GL}} - \lambda_i^0\right)$ 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.

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 nonregular problems. The literature concluded that

https://doi.org/10.1017/S0266466621000013 Published online by Cambridge University Press

frequentist confidence sets may exhibit undesirable properties in nonregular problems (e.g., Buehler, 1959; Cornfield, 1969; Cox, 1958; Müller and Norest, 2016; Pierce, 1973; Robinson, 1977; 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 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 $y_t = \delta_T \mathbf{1}\left\{t > T_b^0\right\} + e_t$, where $e_t \sim i.i.d.\mathcal{N}(0,1)$. The result below can also be shown to hold 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. *Under Assumptions 2.1–2.4 and 3.1–3.7, and Condition 1, 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

https://doi.org/10.1017/S0266466621000013 Published online by Cambridge University Press

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 data generating processes (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 $e_t \sim i.i.d.\mathcal{N}(0, 1)$; M2 is similar to M1 but with $e_t = 0.3e_{t-1} + u_t$, $u_t \sim \mathcal{N}(0, 1)$; M3 is a dynamic model with $D_t = y_{t-1}$, $Z_t = 1$, $e_t \sim i.i.d.\mathcal{N}(0, 0.5)$, and $\alpha^0 = 0.6$. We set $\beta^0 = 1$ in M1 and M2 and $\beta^0 = 0$ in M3. We consider $\lambda_0 = 0.3$ and 0.5, and break magnitudes $\delta^0 = 0.3$, 0.4, 0.6, and 1. Additional simulations are presented in the Supplementary Material.

7.1. Precision of the Change-Point Estimate

We consider the following estimators of T_b^0: the LS estimator (ordinary least squares (OLS)), the GL estimator under a least-absolute loss function (GL-LN), and the GL estimator under a least-absolute loss function with a uniform prior (GL-Uni). We compare the MAE, standard deviation (Std), RMSE, and the 25% and 75% quantiles. We set the trimming parameter ϵ equal to 0.05. As explained in Casini and Perron (2020b), the trimming ϵ should not be chosen too high, because otherwise the estimate might tend to overestimate (resp. underestimate) the break date if it is in the first (resp. second) half of the sample. They found that $\epsilon = 0.05$ performs well for different locations of the break date and this is also confirmed in the simulations in this section. See Sections 1 and 5 in Casini and Perron (2020b) for more discussion.

Tables 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

https://doi.org/10.1017/S0266466621000013 Published online by Cambridge University Press

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

Notes: The model is $y_t = \delta^0 \mathbf{1}\{t > \lfloor T\lambda_0 \rfloor\} + e_t$, $e_t \sim i.i.d. \mathcal{N}(0, 1)$, and $T = 100$. 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 LS 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.

and Perron, 2020a). 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. We have not reported the bias. However, here is a summary of its behavior, which can also be learned from Figures 1 and 2. When $\lambda_b^0 = 0.5$, the bias is small and close to zero, because the finite-sample distributions of the estimators are symmetric. When $\lambda_b^0 < 0.5$, the bias is positive, which means that the break date estimators tend to be on the right of λ_b^0. The opposite hold for $\lambda_b^0 > 0.5$.

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 onred inverting a sequence of locally best invariant tests using Nyblom's (1989) statistic, the inverted

https://doi.org/10.1017/S0266466621000013 Published online by Cambridge University Press

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

Notes: The model is $y_t = \delta^0 \mathbf{1}\{t > \lfloor T\lambda_0 \rfloor\} + e_t$, $e_t = 0.3e_{t-1} + u_t$, $u_t \sim i.i.d.\mathcal{N}(0, 1)$, and $T = 100$. 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

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

https://doi.org/10.1017/S0266466621000013 Published online by Cambridge University Press

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 (2020a) based on continuous record asymptotics, labelled OLS-CR. These methods have been discussed in detail in Casini and Perron (2020a) 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 breaks down 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 LS estimation method is employed with a trimming parameter, $\epsilon = 0.15$ and we use the required degrees of freedom adjustment for the statistic $\widehat{U}_T$ of Elliott and Müller (2007). To construct the OLS-CR method, we follow the steps outlined in Casini and Perron (2020a). 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 Corollary 3.1. For model M2, the estimate of the long-run variance is the prewhitened HAC estimator of Andrews and Monahan (1992). We consider the version $\widehat{U}_T$ 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 Tables 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

[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 (1987) heteroskedasticity and autocorrelation (HAC)-type estimators when there are breaks (see Casini, 2018, 2020 ; Casini et al., 2020; Casini and Perron, 2019; 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, 2021; Vogelsang, 1999).

https://doi.org/10.1017/S0266466621000013 Published online by Cambridge University Press

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	

Notes: The model is $y_t = \delta^0 \mathbf{1}_{\{t > \lfloor T\lambda_0 \rfloor\}} + e_t$, $e_t \sim i.i.d. \mathcal{N}(0, 1)$, and $T = 100$. 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.

coverage rates. However, the 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. This is also consistent with Figures 1 and 2, which show that the asymptotic distribution of the GL estimator does not underestimate uncertainty about the break location even when the break magnitude is small, thereby yielding good coverage rates also in this case. In model M1, the GL method leads to shorter lengths than Bai's even for large breaks. This is not in contradiction with Figure 2, because model M1 is a simpler model than that reported in the figure, which shows that the density of the asymptotic distribution of the GL estimator is more spread out than that from Bai (1997).

Nonreported simulations show that the GL method is robust to heteroskedastic errors $e_t = |z_t| u_t$ and nonnormal errors. The case of multiple breaks is not

https://doi.org/10.1017/S0266466621000013 Published online by Cambridge University Press

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	0.895	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	

Notes: The model is $y_t = \delta^0 \mathbf{1}\{t > \lfloor T\lambda_0 \rfloor\} + e_t$, $e_t = 0.3e_{t-1} + u_t$, $u_t \sim i.i.d.\mathcal{N}(0, 1)$, and $T = 100$. 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	

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

https://doi.org/10.1017/S0266466621000013 Published online by Cambridge University Press

considered ,since they are expected to be similar as in the single break case by virtue of the assumption that the break dates are sufficiently separated. Finally, in the Supplementary Material, we compare the GL method above with its continuous record counterparts developed in Casini and Perron (2020b). Overall, we find that both estimation and confidence intervals based on GL-LN perform well relative to the continuous record counterparts, where significant gains appear to occur when there is high serial correlation in the errors. See the additional results reported in the Supplementary Material.

8. CONCLUSIONS

We developed large-sample results for a class of GL estimators in multiple change-points models where popular methods face some challenges due to the nonregularities of the problem. The GL method exploits the insight of Laplace, who proposed to generate a density from taking an exponential transformation of a LS 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. Since the issues about the finite-sample performance of OLS especially for small breaks continue to hold in more complex structural change models, we believe that our method can be usefully extended to those models. For example, the GL approach can be immediately applied to nonlinear models (e.g., instrumental variable models, linear model with restrictions, nonlinear regression models, etc.) even though particular attention to the appropriate choice of the prior should be given in each context. We believe that our approach can also be relevant for high-dimensional regression with structural changes, although this would require a careful consideration of additional aspects related to the growing number of regressors.

SUPPLEMENTARY MATERIAL

To view supplementary material for this article, please visit: http://doi.org/10.1017/S0266466621000013

REFERENCES

Andrews, D.W.K. (1993) Tests for parameter instability and structural change with unknown change-point. *Econometrica* 61(4), 821–856.

Andrews, D.W.K. & J.C. Monahan (1992) An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica* 60(4), 953–966.

Andrews, D.W.K. & W. Ploberger (1994) Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica* 62(6), 1383–1414.

https://doi.org/10.1017/S0266466621000013 Published online by Cambridge University Press

Bai, J. (1995) Least absolute deviation estimation of a shift. *Econometric Theory* 11(3), 403–436.
Bai, J. (1997) Estimation of a change-point in multiple regression models. *The Review of Economics and Statistics* 79(4), 551–563.
Bai, J. & P. Perron (1998) Estimating and testing linear models with multiple structural changes. *Econometrica* 66(1), 47–78.
Bai, J. & P. Perron (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.
Bickel, P.J. & J.A. Yahav (1969) Some contributions to the asymptotic theory of Bayes solutions. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 11(4), 257–276.
Box, G.E.P. & G.C. Tiao (1973) *Bayesian Inference in Bayesian Analysis*. Addison-Wesley.
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.
Casini, A. (2020) 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., T. Deng, & P. Perron (2020) Theory of Low-Frequency Contamination from Unaccounted Nonstationarity: Consequences for HAR Inference. Unpublished manuscript, Department of Economics and Finance, University of Rome Tor Vergata.
Casini, A. & P. Perron (2019) *Structural Breaks in Time Series*. Oxford Research Encyclopedia of Economics and Finance, Oxford University Press.
Casini, A. & P. Perron (2020a) Continuous record asymptotics for structural change models. arXiv preprint arXiv:1803.10881.
Casini, A. & P. Perron (2020b) Continuous record Laplace-based inference about the break date in structural change models. *Journal of Econometrics*, forthcoming.
Chang, S.Y. & 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. & H. Hong (2003) An MCMC approach to classical estimation. *Journal of Econometrics* 115(2), 293–346.
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. & 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. & L. Horváth (1997) *Limit Theorems in Change-Point Analysis*. John Wiley and Sons.
Deng, A. & 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. & 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. & J. Morley (2015) Likelihood-ratio-based confidence sets for the timing of structural breaks. *Quantitative Economics* 6(2), 463–497.
Forneron, J.J. & 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, & T. Samanta (1995) On convergence of posterior distributions. *Annals of Statistics* 23(6), 2145–2152.

https://doi.org/10.1017/S0266466621000013 Published online by Cambridge University Press

Hall, A., S. Han, & O. Boldea (2010) Inference regarding multiple structural changes in linear models with endogenous regressors. *Journal of Econometrics* 170(2), 281–302.

Hawkins, D.M. (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. & 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. *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. & R.Z. Has'minskiĭ (1981) *Statistical Estimation: Asymptotic Theory*. Springer-Verlag.

Jiang, L., X. Wang, & J. Yu (2018) New distribution theory for the estimation of structural break point in mean. *Journal of Econometrics* 205(1), 156–176.

Jiang, L., X. Wang, & J. Yu (2020) In-fill asymptotic theory for structural break point in autoregression: A unified theory. *Econometric Reviews*, forthcoming.

Juhl, T. & Z. Xiao (2009) Testing for changing mean with monotonic power. *Journal of Econometrics* 148(1), 14–24.

Jun, S.J., J. Pinkse, & 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.

Kendall, M.G. & A. Stuart (1961) *The Advanced Theory of Statistics: Vol. 2–Inference and Relationship*. Hafner Publishing Company.

Kim, D. & 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. & D. Pollard (1990) Cube root asymptotics. *Annals of Statistics* 18(1), 191–219.

Kim, H.J. & 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* 6, 621–646 (English translation by S. M. Stigler 1986. Statist. Sci., 1(19):364–378).

Martins, L. & 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. & W. Polonik (2008) Asymptotic Normality of Plug-in Level Set Estimates. Extended version.

Mason, D.M. & W. Polonik (2009) Asymptotic normality of plug-in level set estimates. *Annals of Applied Probability* 19(3), 1108–1142.

Müller, U.K. & A. Norest (2016) Credibility of confidence sets in nonstandard econometric problems. *Econometrica* 84(6), 2183–2213.

Newey, W.K. & K.D. West (1987) A simple positive semidefinite, heteroskedastic and autocorrelation consistent covariance matrix. *Econometrica* 55(3), 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. & Z. Qu (2010) Estimating structural changes in regression quantiles. *Journal of Econometrics* 162(2), 248–267.

Perron, P. (2006) Dealing with structural breaks. In K. Patterson & T. Mills (eds.), *Palgrave Handbook of Econometrics, Volume 1: Econometric Theory*, pp. 278–352. Palgrave Macmillan.

Perron, P. & Z. Qu (2006) Estimating restricted structural change model. *Journal of Econometrics* 134(2), 373–399.

https://doi.org/10.1017/S0266466621000013 Published online by Cambridge University Press

Perron, P. & 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.
Perron, P. & Y. Yamamoto (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. & Y. Yamamoto (2021) Testing for changes in forecast performance. *Journal of Business and Economic Statistics* 39(1), 148–165.
Perron, P. & X. Zhu (2005) Structural breaks with deterministic and stochastic trends. *Journal of Econometrics* 129(1–2), 65–119.
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. & P. Perron (2007) Estimating and testing structural changes in multivariate regressions. *Econometrica* 75(2), 459–502.
Robert, C. & G. Casella (2004) *Monte Carlo Statistical Methods*, *2nd Edition*, Springer Texts in Statistics. Springer-Verlag.
Robinson, G.K. (1977) Conservative statistical inference. *Journal of the Royal Statistical Society Series B* 39(3), 381–386.
Samworth, R.J. & M.P. Wand (2010) Asymptotics and optimal bandwidth selection for highest density region estimation. *Annals of Statistics* 38(3), 1767–1792.
Vogelsang, T.J. (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.

https://doi.org/10.1017/S0266466621000013 Published online by Cambridge University Press