Continuous Record Asymptotics for Change-Point Models*

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

23rd October 2022

Abstract

In the context of a linear regression model with a single break point, we develop a continuous record asymptotic framework to build inference methods for the break date. We have T observations with a sampling frequency h over a fixed time horizon $[0, N]$, and let $T \to \infty$ with $h \downarrow 0$ while keeping the time span N fixed. We consider the least-squares estimate of the break date and establish consistency and convergence rate. We provide a limit theory for shrinking magnitudes of shifts and locally increasing variances. The asymptotic distribution corresponds to the location of the extremum of a function of the quadratic variation of the regressors and of a Gaussian centered martingale process over a certain time interval. We can account for the asymmetric informational content provided by the pre- and post-break regimes and show how the location of the break and shift magnitude are key ingredients in shaping the distribution. We consider a feasible version based on plug-in estimates, which provides a very good approximation to the finite sample distribution. We use the concept of Highest Density Region to construct confidence sets. Overall, our method is reliable and delivers accurate coverage probabilities and relatively short average length of the confidence sets. Importantly, it does so irrespective of the size of the break.

JEL Classification: C10, C12, C22
Keywords: Asymptotic distribution, break date, change-point, highest density region, semimartingale.

*We wish to thank Davide Pirino, Mark Podolskij, an Associate Editor and two anonymous referees for helpful suggestions and comments. We also thank Christian Gouriéroux, Hashem Pesaran, Myung Hwan Seo and Viktor Todorov for related discussions on some aspects of this project. We are grateful to Iván Fernández-Val, Hiro Kaido, Zhongjun Qu as well as seminar participants at Boston University and participants at 2018 NBER-NSF Time Series Conference and 11th Annual SoFiE Conference for valuable comments and suggestions. Seong Yeon Chang, Andres Sagner and Yohei Yamamoto have provided generous help with computer programming. We thank Yunjong Eo and James Morley for sharing their programs.

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

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

1 Introduction

In the context of a linear regression model with a single break point, we develop a continuous record asymptotic framework and inference methods for the break date. Our model is specified in continuous time but estimated with discrete-time observations using a least-squares method. We have T observations with a sampling frequency h over a fixed time horizon $[0, N]$, where $N = Th$ denotes the time span of the data. We consider a continuous record (or infill) asymptotic framework whereby T increases by shrinking the time interval h to zero while keeping the time span N fixed. We impose very mild conditions on an underlying continuous-time model assumed to generate the data, basically continuous Itô semimartingales.

An extensive amount of research addressed change-point problems under the classical large-N asymptotics. Early contributions are Hinkley (1971), Bhattacharya (1987), and Yao (1987), who adopted a Maximum Likelihood (ML) approach, and for linear regression models, Bai (1997) and Bai and Perron (1998). See the reviews of Perron (2006), Aue and Hórvath (2013), Casini and Perron (2019) and references therein. In this literature, the resulting large-N limit theory for the estimate of the break date depends on the exact distributions of the regressors and disturbances. Therefore, a so-called shrinkage asymptotic theory was adopted whereby the magnitude of the shift, say δ_T, converges to zero which leads to a pivotal limit distribution.

We study a general change-point problem under a continuous record asymptotic framework and develop inference procedures based on the derived asymptotic distribution. We establish consistency at rate-T convergence for the least-squares estimate of the break date, assumed to occur at time N_b^0. Given the fast rate of convergence, we introduce a limit theory with shrinking magnitudes of shifts and increasing variance of the residual process local to the change-point. The asymptotic distribution corresponds to the location of the extremum of a function of the (quadratic) variation of the regressors and of a Gaussian centered martingale process over some time interval. It is characterized by some notable aspects. With the time horizon $[0, N]$ fixed, we can account for the asymmetric informational content provided by the pre- and post-break observations, i.e., the time span and the position of the break date N_b^0 convey useful information about the finite-sample distribution. In contrast, this is not achievable under the large-N shrinkage asymptotic framework because both pre- and post-break segments expand proportionately as T increases and, given the mixing assumptions imposed, only the neighborhood around the break date remains relevant. Further, the domain of the extremum depends on the position of the break N_b^0 relative to N, or total span, and thus the distribution is asymmetric, in general. The degree of asymmetry increases as the true break point moves away from mid-sample. This holds unless the magnitude of the break is large, in which case the density is symmetric irrespective of the location of the break. This accords with simulation evidence which documents that the break point estimate is less precise and the coverage rates of the confidence intervals less reliable when the break is not at mid-sample [see, e.g., Chang and Perron (2018)]. When the shift magnitude is small, the density displays three modes. As the shift magnitude increases, this tri-modality vanishes. We show via simulations that all of these features are shared by the finite-sample distribution of the least-squares estimator of the break date. Hence, the continuous record asymptotics theory provides an accurate approximation to the finite-sample distribution of the break date estimator. In contrast, we show that the large-N shrinkage asymptotic distribution of Yao (1987) and Bai (1997) provides a poor approximation to the finite-sample distribution and does not share any of those features.

Our asymptotics can be seen as intermediate between the shrinkage asymptotics and more recent approaches relying on weak identification [see e.g., Elliott and Müller (2007) and Elliott, Müller and Watson (2015)]. On the one hand, using the usual shrinking condition of Yao (1987) and Bai (1997) for which the break magnitude δ_T goes to zero at a rate slower than $O(T^{-1/2})$ leads to underestimation of the uncertainty about the break date. On the other hand, the weak identification condition of Elliott and Müller (2007) for which δ_T goes to zero at a fast rate (i.e., $\delta_T = O(T^{-1/2})$, so that the change-point cannot be consistently estimated) leads to overstating the uncertainty. This has opposite consequences for the confidence intervals of the break date. Confidence sets have poor coverage probabilities when the break is small under Bai's framework while they can be too wide under that of Elliott and Müller (2007). In this paper, the key is not to focus our asymptotic experiment on shrinking condition on δ_T but to make assumptions on the signal-to-noise ratio δ_T/σ_t instead, where σ_t is the volatility of the errors. We require δ_T to go to zero at a slower rate than that of Elliott and Müller (2007)—to guarantee strong identification—and require σ_t to increase without bound when t approaches the break date T_b^0. This offers a new characterization of the uncertainty without compromising strong identification and consistency of the model parameters needed to conduct inference.

Despite the effort devoted to the construction of confidence intervals for the break date [see e.g., Bai and Perron (1998), Elliott and Müller (2007) and Eo and Morley (2015)], what is still missing is a method that, for both large and small breaks, achieves both accurate coverage rates and satisfactory average lengths. The most popular method is that of Bai and Perron (1998) which yields confidence intervals that are relatively short but have good coverage only when the magnitude of the break is not small. However, both small and large breaks are relevant for empirical work; breaks that are statistically small can still be practically relevant.

Given the peculiar properties (e.g., multi-modality and asymmetry) of the continuous record asymptotic distribution, we propose a non-standard inference procedure related to Bayesian analyses. We use the concept of Highest Density Region to construct confidence sets for the break date. Our method has good coverage and length across all break magnitudes. This has important implications for empirical work because the user can be confident that our confidence interval includes the true value across all break sizes. For small breaks, the length of the confidence intervals from any method can be quite large for some models. However, our confidence interval is still informative because it reveals that there is high uncertainty about the change-point. The same information cannot be provided by existing methods either because they do not have good coverage unless the break is not small [e.g., Bai and Perron (1998)] or because they have a large length even when the break is not small [e.g., Elliott and Müller (2007)].

We use the continuous record asymptotics to provide an alternative approximation to the finite-sample distribution of the least-squares estimator of the break date based on discrete-time data. This creates no contradiction since asymptotic theory is intended as a thought experiment used to obtain approximations to the distribution of estimators or test statistics. The continuous record asymptotics has proven to be useful in other discrete-time settings such as in the context of unit roots [cf. Phillips (1987) and Perron (1991)] and nonparametric regression [cf. Brown and Low (1996)]. Brown and Low (1996) showed the asymptotic equivalence between a nonparametric regression problem and a white noise with drift problem.

Our empirical applications of interest involve regression models using time series data sampled at, say, annual, quarterly, monthly and daily frequency (i.e., the same empirical settings considered in the long-span structural break literature). For ultra high-frequency data (i.e., intra-daily data) one would need to account for features such as jumps, market microstructure noise, etc. This would require some extension/modification of our method. For example, one should apply some techniques to remove the jumps and consider coefficient estimators that are built on spot volatility estimators rather than on integrated volatility. To show the empirical usefulness of our approach, we apply our inference methods to construct the confidence sets about the break dates in monthly U.S. industrial production growth. The results show that the proposed confidence set works well and have better properties than existing methods.

Recent work in change-point analysis has focused on estimation when the number of change-points is allowed to increase with the sample size [e.g., Fryzlewicz (2014)] and when the change-point is allowed to approach the start and end sample point. A growing literature has also considered change-points in a high-dimensional setting [e.g., Lee, Seo and Shin (2016), Leonardi and Bühlmann (2016), Wang, Lin and Willett (2019a) and Wang, Yu, Rinaldo and Willett (2019b)]. This work is mainly concerned with consistent estimation of the change-point dates and development of corresponding computational algorithms. Our focus is on asymptotic theory and inference within the classical change-point model with a single break. Our results can also have useful implications for the growing literature on inference in high-dimensional change-point analysis and for the literature on threshold regression [see, e.g., Hansen (2000) and Hidalgo, Lee and Seo (2019)].

This paper relates to other work by the authors. Casini and Perron (2021b) used the asymptotic results developed in this paper and proposed a new Generalized Laplace estimator of the break date under a continuous record asymptotic framework. Casini and Perron (2020b) analyzed the Generalized Laplace method under classical asymptotics and focused on the theoretical relationship between the asymptotic distribution of frequentist and Bayesian estimators of the break point. Finally, Chambers and Taylor (2019) considered both deterministic one-time change and continuous stochastic parameter change in a continuous-time autoregressive model while Casini (2018) introduced continuous-time asymptotics to test for forecast failure. Casini and Perron (2021a) considered testing for and estimating change-points in a locally stationary process using frequency-domain methods.

The paper is organized as follows. Section 2 introduces the model and the estimation method. Section 3 contains results about the consistency and rate of convergence for fixed shifts. Section 4 develops the asymptotic theory. We compare our limit theory with the finite-sample distribution in Section 5. Section 6 describes how to construct the confidence sets, with simulation results reported in Section 7. An empirical application is presented in Section 8. Section 9 provides brief concluding remarks. The Supplement [Casini and Perron (2021c)] contains the proofs as well as additional material.

2 Model and Assumptions

We denote the transpose of a matrix A by A' and the $(i,\, j)$ elements of A by $A^{(i,j)}$. 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 $x \in \mathbb{R}^p$. We use $\lfloor \cdot \rfloor$ to denote the largest smaller integer function. A sequence $\{u_{kh}\}_{k=1}^{T}$ is *i.i.d.* (resp., *i.n.d*) if the u_{kh} are independent and identically (resp., non-identically) distributed. We use $\xrightarrow{P}$, $\Rightarrow$, and $\overset{\mathcal{L}\text{-s}}{\Rightarrow}$ to denote convergence in probability, weak convergence and stable convergence in law, respectively. For semimartingales $\{S_t\}_{t\geq 0}$ and $\{R_t\}_{t\geq 0}$, we denote their covariation process by $[S,\, R]_t$ and their predictable counterpart by $\langle S,\, R\rangle_t$. The symbol "$\triangleq$" denotes definitional equivalence.

Consider a classical partial change-point model with a single break point:

$$\begin{aligned} Y_t &= D_t'\nu^0 + Z_t'\delta^0_{Z,1} + e_t, \quad (t = 0,\, 1, \ldots,\, T_b^0) \\ Y_t &= D_t'\nu^0 + Z_t'\delta^0_{Z,2} + e_t, \quad (t = T_b^0 + 1, \ldots,\, T), \end{aligned} \tag{2.1}$$

where Y_t is the dependent variable, D_t and Z_t are, respectively, $q \times 1$ and $p \times 1$ vectors of regressors and e_t is an unobservable disturbance. The vector-valued parameters ν^0, $\delta^0_{Z,1}$ and $\delta^0_{Z,2}$ are unknown with $\delta^0_{Z,1} \neq \delta^0_{Z,2}$. Our main purpose is to develop inference methods for the unknown change-point date T_b^0 when $T+1$ observations on $(Y_t,\, D_t,\, Z_t)$ are available. Before moving to the re-parametrization of the model, we discuss the underlying continuous-time model assumed to generate the data. The discrete-time variables are assumed to be generated from the continuous-time processes $\{D_s,\, Z_s,\, e_s\}_{s\geq 0}$ defined on a filtered probability space $(\Omega,\, \mathscr{F},\, (\mathscr{F}_s)_{s\geq 0},\, P)$.

The sampling occurs at regularly spaced time intervals of length h within a fixed time horizon $[0,\, N]$ where N denotes the span of the data. We observe $\{{}_hY_{kh},\, {}_hD_{kh},\, {}_hZ_{kh};\, k = 0,\, 1, \ldots,\, T = N/h\}$, with ${}_hD_{kh} \in \mathbb{R}^q$ and ${}_hZ_{kh} \in \mathbb{R}^p$ are random vector step functions which jump only at times

$0,\, h, \ldots, Th$. We shall allow ${}_hD_{kh}$ and ${}_hZ_{kh}$ to include both predictable processes and locally-integrable semimartingles, though the case with predictable regressors is more delicate and discussed in the supplement. The discretized processes ${}_hD_{kh}$ and ${}_hZ_{kh}$ are adapted to $\{\mathscr{F}_t\}_{t\geq 0}$. As a matter of notation, hereafter we use $t,\, s \in [0,\, N]$ as indices of continuous time while we use $k = 0,\, 1, \ldots, T = N/h$ as the index of discrete-time. For any process X we denote its "increments" by $\Delta_h X_k = X_{kh} - X_{(k-1)h}$. For $k = 1, \ldots, T$, let $\Delta_h D_k \triangleq \mu_{D,k} h + \Delta_h M_{D,k}$ and $\Delta_h Z_k \triangleq \mu_{Z,k} h + \Delta_h M_{Z,k}$ where the "drifts" $\mu_{D,t} \in \mathbb{R}^q$, $\mu_{Z,t} \in \mathbb{R}^p$ are $\mathscr{F}_{t-h}$-measurable, and $M_{D,k} \in \mathbb{R}^q$, $M_{Z,k} \in \mathbb{R}^p$ are continuous local martingales with finite conditional covariance matrix P-a.s., $\mathbb{E}(\Delta_h M_{D,t} \Delta_h M'_{D,t} |\, \mathscr{F}_{t-h}) = \Sigma_{D,t-h} \Delta t$ and $\mathbb{E}(\Delta_h M_{Z,t} \Delta_h M'_{Z,t} |\, \mathscr{F}_{t-h}) = \Sigma_{Z,t-h} \Delta t$ (Δt and h are used interchangeably). Let $\lambda_0 \in (0,\, 1)$ denote the fractional break date (i.e., $T_b^0 = \lfloor T\lambda_0 \rfloor$). Via the Doob-Meyer Decomposition, we can write the model as

$$\Delta_h Y_k \triangleq \begin{cases} (\Delta_h D_k)'\, \nu^0 + (\Delta_h Z_k)'\, \delta^0_{Z,1} + \Delta_h e^*_k, & (k = 1, \ldots, \lfloor T\lambda_0 \rfloor)\,, \\ (\Delta_h D_k)'\, \nu^0 + (\Delta_h Z_k)'\, \delta^0_{Z,2} + \Delta_h e^*_k, & (k = \lfloor T\lambda_0 \rfloor + 1, \ldots, T)\,, \end{cases} \tag{2.2}$$

where the error process $\{\Delta_h e^*_t,\, \mathscr{F}_t\}$ is a continuous local martingale difference sequence with conditional variance $\mathbb{E}[(\Delta_h e^*_t)^2 \,|\, \mathscr{F}_{t-h}] = \sigma^2_{e,t-h} \Delta t$ P-a.s. finite. The underlying continuous-time data-generating process can thus be represented (up to P-null sets) in integral equation form as

$$D_t = D_0 + \int_0^t \mu_{D,s} ds + \int_0^t \sigma_{D,s} dW_{D,s}, \quad Z_t = Z_0 + \int_0^t \mu_{Z,s} ds + \int_0^t \sigma_{Z,s} dW_{Z,s}, \tag{2.3}$$

where $\sigma_{D,t}$ and $\sigma_{Z,t}$ are the instantaneous covariance processes taking values in $\mathcal{M}_q^{\text{càdlàg}}$ and $\mathcal{M}_p^{\text{càdlàg}}$ [the space of $p \times p$ positive definite real-valued matrices whose elements are càdlàg]; W_D (resp., W_Z) is a q (resp., p)-dimensional standard Wiener process; $e^* = \{e^*_t\}_{t\geq 0}$ is a continuous local martingale which is orthogonal (in a martingale sense) to $\{D_t\}_{t\geq 0}$ and $\{Z_t\}_{t\geq 0}$; and D_0 and Z_0 are $\mathscr{F}_0$-measurable random vectors. In (2.3), $\int_0^t \mu_{D,s} ds$ is a continuous adapted process with finite variation paths and $\int_0^t \sigma_{D,s} dW_{D,s}$ corresponds to a continuous local martingale. Thus, (2.3) implies that D_t and Z_t are continuous semimartingales.

Assumption 2.1. *(i) $\mu_{D,t}$, $\mu_{Z,t}$, $\sigma_{D,t}$ and $\sigma_{Z,t}$ satisfy P-a.s., $\sup_{\omega\in\Omega,\, 0<t\leq\tau_T} \|\mu_{D,t}(\omega)\| < \infty$, $\sup_{\omega\in\Omega,\, 0<t\leq\tau_T} \|\mu_{Z,t}(\omega)\| < \infty$, $\sup_{\omega\in\Omega,\, 0<t\leq\tau_T} \|\sigma_{D,t}(\omega)\| < \infty$ and $\sup_{\omega\in\Omega,\, 0<t\leq\tau_T} \|\sigma_{Z,t}(\omega)\| < \infty$ for some localizing sequence $\{\tau_T\}$ of stopping times. Also, $\sigma_{D,s}$ and $\sigma_{Z,s}$ are càdlàg; (ii) $\int_0^t \mu_{D,s} ds$ and $\int_0^t \mu_{Z,s} ds$ belong to the class of continuous adapted finite variation processes; (iii) $\int_0^t \sigma_{D,s} dW_{D,s}$ and $\int_0^t \sigma_{Z,s} dW_{Z,s}$ are continuous local martingales with P-a.s. finite positive definite conditional variances (or spot covariances) defined by $\Sigma_{D,t} = \sigma_{D,t}\sigma'_{D,t}$ and $\Sigma_{Z,t}, = \sigma_{Z,t}\sigma'_{Z,t}$, which for all $t < \infty$ satisfy $\int_0^t \Sigma^{(j,j)}_{D,s} ds < \infty$ $(j = 1, \ldots, q)$ and $\int_0^t \Sigma^{(j,j)}_{Z,s} ds < \infty$ $(j = 1, \ldots, p)$. Furthermore, for every $j = 1, \ldots, q$, $r = 1, \ldots, p$, and $k = 1, \ldots, T$, $h^{-1} \int_{(k-1)h}^{kh} \Sigma^{(j,j)}_{D,s} ds$ and $h^{-1} \int_{(k-1)h}^{kh} \Sigma^{(r,r)}_{Z,s} ds$ are bounded away from zero and infinity, uniformly in k and h.*

Part (i) restricts the processes to be locally bounded and part (ii) requires the drifts to be adapted finite variation processes. These are standard regularity conditions in the high-frequency statistics literature [cf. Barndorff-Nielsen and Shephard (2004)]. Part (iii) requires the regressors to have finite integrated covariance.

Assumption 2.2. *$e_t^* \triangleq \int_0^t \sigma_{e,s} dW_{e,s}$ with $0 < \sigma_{e,t}^2 < \infty$, where W_e is a one-dimensional standard Wiener process. Furthermore, $\langle e, D \rangle_t = \langle e, Z \rangle_t = 0$ identically for all $t \geq 0$.*

Assumption 2.2 presents the continuous time representation of the error process and states that the error process is uncorrelated with the regressors.

We rule out jump processes; hence, our results are not expected to provide good approximations for high-frequency data but for data sampled at lower frequencies. The use of ultra high frequency data involves a host of issues that we cannot handle (e.g., market-microstructure, bid-ask spread, volatility jumps, non-continuous sampling, etc.). However, our methods and results could be extended. Although the raw least-squares method is not robust to jumps, it can be augmented with threshold-holding techniques that detect the jumps. Further, since ultra high-frequency data imply larger sample sizes one should modify the raw least-squares method for the regression coefficients so that the estimation is based on the integrated spot covariation (constructed over blocks of ultra high-frequency observations) rather than just the full-sample covariation. The integrated spot estimator was considered by Li, Todorov and Tauchen (2017) who proposed an optimal weighted average of local nonparametric co-volatility estimates. Our main theoretical results should continue to hold under this more general setting, though the proofs would be substantially more involved. We leave these extensions for future research.

Assumption 2.3. *$\Sigma^0 \triangleq \{\Sigma_{\cdot,t}, \sigma_{e,t}\}_{t\geq 0}$ has P-a.s. continuous sample paths.*

An interesting issue is whether the theoretical results to be derived for model (2.2) are applicable to classical structural change models for which an increasing span of data is assumed. This requires establishing a connection between the assumptions imposed on the stochastic processes in both settings. Roughly, the classical long-span setting uses approximation results valid for weakly dependent data; e.g., ergodic and mixing processess. Such assumptions are not needed under our fixed-span asymptotics. Nonetheless, we can impose restrictions on the probabilistic properties of the latent volatility processes in our model and thereby guarantee that ergodic and mixing properties are inherited by the corresponding observed processes. This follows from Theorem 3.1 in Genon-Catalot, Jeantheau and Laredo (2000) together with Proposition 4 in Carrasco and Chen (2002). For example, these results imply that the observations $\{Z_{kh}\}_{k\geq 1}$ (with fixed h) can be viewed (under certain conditions) as a hidden Markov model which inherits the ergodic and mixing properties of $\{\sigma_{Z,t}\}_{t\geq 0}$. Hence, our model encompasses those considered in the structural change literature that uses a long-span asymptotic setting.

In the context of predictive regression analyzed under long-span asymptotics, recent work by Georgiev, Harvey, Leybourne and Taylor (2020) and Andersen and Varneskov (2021, 2022) proposed tests for structural breaks and parameter instability that are robust to (near) unit roots or nonstationary fractional integration. The new tests are constructed by modifying popular structural break tests [e.g., the sup-Wald test of Andrews (1993)] to allow for such features. In contrast, in our work we consider the original least-squares estimator of the break point but we analyze its limiting distribution under a continuous record. We extend model (2.2) to allow for predictable processes (e.g., a constant and/or lagged dependent variable) in the supplement.

Assumption 2.4. $N_b^0 = N\lambda_0$ *for some* $\lambda_0 \in (0,\, 1)$.

It is useful to re-parametrize model (2.2). Let $y_{kh} = \Delta_h Y_k$, $x_{kh} = (\Delta_h D_k',\, \Delta_h Z_k')'$, $z_{kh} = \Delta_h Z_k$, $e_{kh} = \Delta_h e_k^*$, $\beta^0 = ((\pi^0),\, (\delta_{Z,1}^0)')'$ and $\delta_Z^0 = \delta_{Z,2}^0 - \delta_{Z,1}^0$. (2.2) can be expressed as:

$$\begin{aligned} y_{kh} &= x_{kh}'\beta^0 + e_{kh}, && (k = 1, \ldots,\, T_b^0), \\ y_{kh} &= x_{kh}'\beta^0 + z_{kh}'\delta_Z^0 + e_{kh}, && (k = T_b^0 + 1, \ldots,\, T), \end{aligned} \tag{2.4}$$

where the true parameter $\theta^0 = ((\beta^0)',\, (\delta_Z^0)')'$ takes value in a compact space $\Theta \subset \mathbb{R}^{\dim(\theta)}$. Also, define $z_{kh} = R'x_{kh}$, where R is a $(q+p) \times p$ known matrix with full column rank. We consider a partial structural change model for which $R = (0,\, I)'$ with I an identity matrix.

Finally, we write the model in matrix format which will be useful for the derivations. Let $Y = (y_h,\, \ldots,\, y_{Th})'$, $X = (x_h,\, \ldots,\, x_{Th})'$, $e = (e_h,\, \ldots,\, e_{Th})'$, $X_1 = (x_h,\, \ldots,\, x_{T_b h},\, 0,\, \ldots,\, 0)'$, $X_2 = (0,\, \ldots,\, 0,\, x_{(T_b+1)h}, \ldots,\, x_{Th})'$ and $X_0 = (0,\, \ldots,\, 0,\, x_{\left(T_b^0+1\right)h}, \ldots,\, x_{Th})'$. Note that the difference between X_0 and X_2 is that the latter uses T_b rather than T_b^0. Define $Z_1 = X_1 R$, $Z_2 = X_2 R$ and $Z_0 = XR$. (2.4) in matrix format is: $Y = X\beta^0 + Z_0\delta_Z^0 + e$. We consider the least-squares estimator of T_b, i.e., the minimizer of $S_T(T_b)$, the sum of squared residuals when regressing Y on X and Z_2 over all possible partitions, namely: $\widehat{T}_b^{\mathrm{LS}} = \mathrm{argmin}_{p+q \leq T_b \leq T} S_T(T_b)$. It is straightforward to show that $\widehat{T}_b^{\mathrm{LS}} = \mathrm{argmin}_{p+q \leq T_b \leq T} Q_T(T_b)$ where $Q_T(T_b) \triangleq \widehat{\delta}_{T_b}'\,(Z_2' M Z_2)\,\widehat{\delta}_{T_b}$, $\widehat{\delta}_{T_b}$ is the least-squares estimator of δ_Z^0 when regressing Y on X and Z_2, and $M = I - X\,(X'X)^{-1}\,X'$. For brevity, we will write $\widehat{T}_b$ for $\widehat{T}_b^{\mathrm{LS}}$ with the understanding that $\widehat{T}_b$ is a sequence indexed by T. Let $\widehat{\delta} = \widehat{\delta}_{\widehat{T}_b}$. The estimate of the break fraction is then $\widehat{\lambda}_b = \widehat{T}_b/T$. Both in practice and for theoretical analyses, a trimming parameter $\pi \in (0,\, 1/2)$ is applied to restrict the minimization over the interval $[T\pi,\, (1-\pi)\,T]$.

3 Consistency and Convergence Rate under Fixed Shifts

We now establish the consistency and convergence rate of the least-squares estimator under fixed shifts. Under the classical large-N asymptotics, related results have been established by Bai (1997) and Bai and Perron (1998). Early important results for a mean-shift appeared in Yao (1987) and

Bhattacharya (1987) for an $i.i.d.$ series, Bai (1994) for linear processes and Picard (1985) for a Gaussian autoregressive model.

Assumption 3.1. *There exists an* l_0 *such that for all* $l > l_0$*, the matrices* $(lh)^{-1}\sum_{k=1}^{l} x_{kh}x'_{kh}$, $(lh)^{-1}\sum_{k=T-l+1}^{T} x_{kh}x'_{kh}$, $(lh)^{-1}\sum_{k=T_b^0-l+1}^{T_b^0} x_{kh}x'_{kh}$, *and* $(lh)^{-1}\sum_{k=T_b^0+1}^{T_b^0+l} x_{kh}x'_{kh}$, *have minimum eigenvalues bounded away from zero in probability uniformly in* h.

Assumption 3.2. *Let* $Q_0(T_b, \theta^0) \triangleq \mathbb{E}[Q_T(T_b, \theta^0) - Q_T(T_b^0, \theta^0)]$. *There exists a* T_b^0 *such that* $Q_0(T_b^0, \theta^0) > \sup_{(T_b, \theta^0)\notin \mathbf{B}} Q_0(T_b, \theta^0)$, *for every open set* $\mathbf{B}$ *that contains* (T_b^0, θ^0).

Assumption 3.1 is similar to A2 in Bai and Perron (1998) and requires enough variation around the break point and at the beginning and end of the sample. The factor h^{-1} normalizes the observations so that the assumption is implied by a weak law of large numbers. Assumption 3.2 is a standard uniqueness identification condition. We then have the following results.

Proposition 3.1. *Under Assumption 2.1-2.4 and 3.1-3.2,* $\widehat{\lambda}_b \xrightarrow{P} \lambda_0$.

Proposition 3.2. *Under Assumption 2.1-2.4 and 3.1-3.2 for any* $\varepsilon > 0$*, there exists a* $K > 0$ *such that for all large* T*,* $P(T\left|\widehat{\lambda}_b - \lambda_0\right| > K) < \varepsilon$.

We have the same T-convergence rate as under large-N asymptotics. Let $\theta^0 = ((\beta^0)', (\delta_1^0)', (\delta_2^0)')'$. The fast T-rate of convergence implies that the least-squares estimate of θ^0 is the same as when λ_0 is known. A natural estimator for θ^0 is $\mathrm{argmin}_{\beta\in\mathbb{R}^{p+q},\delta\in\mathbb{R}^p}||Y - X\beta - \widehat{Z}_2\delta||^2$, where we use $\widehat{T}_b$ instead of T_b in the construction of $\widehat{Z}_2$. Then we have the following result, akin to an extension of the results in Barndorff-Nielsen and Shephard (2004) and Li et al. (2017).

Proposition 3.3. *Under Assumption 2.1-2.4 and 3.1-3.2, we have as* $T \to \infty$ *(*N *fixed),* $(\sqrt{T/N}(\widehat{\beta} - \beta^0), \sqrt{T/N}(\widehat{\delta} - \delta_Z^0))' \xrightarrow{d} \mathscr{MN}(0, V)$ *where* $\mathscr{MN}$ *denotes a mixed Gaussian distribution and* V *is a positive definite* $\mathscr{F}$*-conditional asymptotic variance. If* $\Sigma_{\cdot,t}$ *is independent of* $\sigma_{e,t}$*, then*

$$V = \overline{V}^{-1}\begin{bmatrix} \int_0^N \Sigma_{X,s}\sigma_{e,s}^2 ds & \int_{N_b^0}^N \Sigma_{XZ,s}\sigma_{e,s}^2 ds \\ \int_{N_b^0}^N \Sigma_{ZX,s}\sigma_{e,s}^2 ds & \int_{N_b^0}^N \Sigma_{Z,s}\sigma_{e,s}^2 ds \end{bmatrix}\overline{V}^{-1},$$

and

$$\overline{V} = \begin{bmatrix} \int_0^N \Sigma_{X,s} ds & \int_{N_b^0}^N \Sigma_{XZ,s} ds \\ \int_{N_b^0}^N \Sigma_{ZX,s} ds & \int_{N_b^0}^N \Sigma_{Z,s} ds \end{bmatrix}, \qquad \text{with } \Sigma_{XZ,s} = \sigma_{X,s}\sigma'_{Z,s}.$$

V can be random because $\Sigma_{\cdot,t}$ and $\sigma_{e,t}$ can be stochastic. Under fixed shifts, Proposition 3.1-3.3 shows the asymptotic equivalence of discrete and continuous-time regression models with a change-point, a result corresponding to Brown and Low (1996) for nonparametric regression.

4 Asymptotic Distribution under a Continuous Record

We now present results about the limiting distribution of the least-squares estimate of the break date under a continuous record framework. As in the classical large-N asymptotics, it depends on the exact distribution of the data and the errors for fixed break sizes [c.f., Hinkley (1971)]. This has forced researchers to consider a shrinkage asymptotic theory where the size of the shift is made local to zero as T increases, an approach developed by Picard (1985) and Yao (1987). We continue with this avenue. Given the consistency result, we know that there exists some h^* such that for all $h < h^*$ with high probability $\eta Th \leq \widehat{N}_b \leq (1-\eta)\, Th$, for $\eta > 0$ such that $\lambda_0 \in (\eta, 1-\eta)$. By Proposition 3.2, $\widehat{N}_b - N_b^0 = O_p(T^{-1})$, i.e., $\widehat{N}_b$ is in a shrinking neighborhood of N_b^0. With a certain rescaling of the objective function one can first obtain the shrinkage asymptotic distribution of Bai (1997). However, this is unsatisfactory for two reasons. First, as we show below [see also Casini and Perron (2020b; 2021b)], the shrinkage asymptotic distribution provides a poor approximation to the finite-sample distribution of the least-squares estimator. Second, the latter point also explains the poor coverage properties of the confidence intervals derived from the shrinkage asymptotic distribution when the magnitude of the break is not large. Some related results were obtained with Jiang et al. (2018) for a simple location model. Their approach is, however, quite restrictive and no feasible inference procedure suggested. See the supplement for a more complete discussion.

We begin with the following assumption which specifies that i) we use a shrinking condition on δ_Z^0; ii) we introduce a locally increasing variance condition on the residual process. The first is similarly used under classical large-N asymptotics, while the second is new and useful in our context in order to accurately capture the relevant uncertainty in the change-point problem. We do not impose restrictions only on δ_Z^0 but also on the ratio δ_Z^0/σ_t when t is close to T_b^0. We refer to δ_Z^0/σ_t as the signal-to-noise ratio. Controlling this ratio rather than just δ_Z^0 allows for an alternative characterization of the uncertainty about the change-point date in order to obtain an asymptotic distribution which provides a better approximation of the finite-sample distribution of the estimator. To emphasize that δ_Z^0 depends on the sample-size we denote it by δ_h.

Assumption 4.1. *Let $\delta_h = \delta^0 h^{1/4}$, $\delta^0 \in \mathbb{R}^p$ and assume that for all $t \in (N_b^0 - \epsilon,\, N_b^0 + \epsilon)$, with $\epsilon \downarrow 0$ and $T^{1-\kappa}\epsilon \to B < \infty$, $0 < \kappa < 1/2$, $\mathbb{E}[(\Delta_h e_t^*)^2 \,|\, \mathscr{F}_{t-h}] = \sigma_{h,t-h}^2 \Delta t$ P-a.s., where $\sigma_{h,t} \triangleq \sigma_h \sigma_{e,t}$, $\sigma_h \triangleq \overline{\sigma} h^{-1/4}$ and $0 < \overline{\sigma} < \infty$.*

Note that the localization parameter δ^0 in the definition of δ_h is different from the fixed parameter δ_Z^0 since $h \to 0$. The rate $1/4$ in the conditions $\delta_h = O(h^{1/4})$ and $\sigma_h = O(h^{-1/4})$ is for tractability. One can show that consistency also holds for a rate faster than $1/4$, though slower than κ. However, for the derivation of the limiting distribution one needs $\delta_h/\sigma_h = O(h^{1/2})$ and $O(\delta_h) = O(\sigma_h^{-1})$ with $\kappa < 1/2$. The increasing local volatility condition in Assumption 4.1 is similar to that used by Christensen, Oomen and Renò (2022). They named it volatility burst [cf.

eq. (7)] and they used it along with a drift bust condition to model the existence of short-lived locally explosive trends in stock prices (i.e., flash crashes). Although that is a different context from ours, it is the deviation from the standard diffusion setting that allows the authors to account for flash crashes and that allows us to provide a more accurate asymptotic theory for change-point estimation.

The vector of scaled true parameters is $\theta_h \triangleq ((\beta^0)', \delta_h')'$. Under Assumption 4.1 the error process has the following representation.

Assumption 4.2. $e_t^* = \int_0^t \sigma_{e,s,\Delta} dW_{e,s}$ *where* W_e *is a one-dimensional standard Wiener process,*

$$\sigma_{e,s,\Delta} = \begin{cases} \sigma_{e,s} & s \notin (N_b^0 - \epsilon,\, N_b^0 + \epsilon) \\ \overline{\sigma} h^{-1/4} \sigma_{e,s} & s \in (N_b^0 - \epsilon,\, N_b^0 + \epsilon) \end{cases}.$$

with $0 < \sigma_{e,s}^2 < \infty$. *Furthermore,* $\langle e,\, D\rangle_t = \langle e,\, Z\rangle_t = 0$ *identically for all* $t \geq 0$.

Define

$$\Delta_h \widetilde{e}_t \triangleq \begin{cases} \Delta_h e_t^*, & t \notin (N_b^0 - \epsilon,\, N_b^0 + \epsilon) \\ h^{1/4} \Delta_h e_t^*, & t \in (N_b^0 - \epsilon,\, N_b^0 + \epsilon) \end{cases}. \tag{4.1}$$

We shall refer to $\{\Delta_h \widetilde{e}_t,\, \mathscr{F}_t\}$ as the normalized residual process. Under this framework, the rate of convergence of $\widehat{N}_b$ is now $T^{1-\kappa}$ with $0 < \kappa < 1/2$. Due to the fast rate of convergence of the change-point estimator, the objective function oscillates too rapidly as $h \downarrow 0$. By scaling up the volatility of the errors around the change-point, we make the objective function behave as if it were a function of a standard diffusion process. The neighborhood in which the errors have relatively higher variance is shrinking at rate $1/T^{1-\kappa}$, the rate of convergence of $\widehat{N}_b$. Hence, in a neighborhood of N_b^0 in which we study the limiting behavior of the break point estimator, the rescaled criterion function is regular enough so that a feasible limit theory can be developed. The rate of convergence $T^{1-\kappa}$ is still sufficiently fast to guarantee a $\sqrt{T}$-consistent estimation of the slope parameters, as stated in the following proposition. Let $\langle Z_\Delta,\, Z_\Delta\rangle (v)$ be the predictable quadratic variation process of Z_Δ. The process $\mathscr{W}(v)$ is, conditionally on $\mathscr{F}$, a two-sided centered Gaussian martingale with independent increments and variances given in Section S.B of the supplement.

Proposition 4.1. *Under Assumption 2.1, 2.3-2.4, 3.1-3.2 and 4.1-4.2, (i)* $\widehat{\lambda}_b \xrightarrow{P} \lambda_0$; *(ii) for every* $\varepsilon > 0$ *there exists a* $K > 0$ *such that for all large* T, $P(T^{1-\kappa} \left|\widehat{\lambda}_b - \lambda_0\right| > K||\delta^0||^{-2}\overline{\sigma}^2) < \varepsilon$; *and (iii) for* $\kappa \in (0,\, 1/4]$, $(\sqrt{T/N}(\widehat{\beta} - \beta^0),\, \sqrt{T/N}(\widehat{\delta} - \delta_h))' \xrightarrow{d} \mathscr{MN}(0,\, V)$ *as* $T \to \infty$, *with* V *given in Proposition 3.3.*

We first present a general result which shows that under Assumption 4.1 one can obtain a shrinkage asymptotic distribution similar to Bai (1997). The latter exploits the consistency of

$\widehat{\lambda}_b$ and the fact that mixing conditions implies that the regimes before and after λ_0 are asymptotically independent. Let $Z_\Delta \triangleq (0, \ldots, 0, z_{(T_b+1)h}, \ldots, z_{T_b^0 h}, 0, \ldots, 0)$ if $T_b < T_b^0$ and $Z_\Delta \triangleq (0, \ldots, 0, z_{(T_b^0+1)h}, \ldots, z_{T_b h}, 0, \ldots, 0)$ if $T_b > T_b^0$.

Proposition 4.2. *Under Assumption 2.1, 2.3-2.4, 3.1-3.2 and 4.1-4.2,*

$$T^{1-\kappa}(\widehat{\lambda}_b - \lambda_0) \overset{\mathcal{L}-\mathrm{s}}{\Rightarrow} \underset{v \in (-\infty, \infty)}{\operatorname{argmax}} \, 2(\delta^0)' \mathscr{W}(v) \,. \tag{4.2}$$

The distribution in Proposition 4.2 is different from Bai (1997). One can show that his distribution can be obtained under a continuous record if Assumption 4.1 is modified as follows: $\delta_h = \delta^0 h^{\kappa/2}$, $T^{1-\kappa}\epsilon \to B < \infty$, $0 < \kappa \leq 1/2$ and $\sigma_h \triangleq \overline{\sigma} h^{-\kappa/2}$. This would result in,

$$\begin{aligned} & T^{1-\kappa}\left(\widehat{\lambda}_b - \lambda_0\right) \overset{\mathcal{L}-\mathrm{s}}{\Rightarrow} \\ & \quad \underset{v \in (-\infty, \infty)}{\operatorname{argmax}} \left\{ -\left(\delta^0\right)' \langle Z_\Delta, Z_\Delta \rangle (v) \, \delta^0 + 2\left(\delta^0\right)' \mathscr{W}(v) \right\}. \end{aligned} \tag{4.3}$$

The difference between (4.2) and (4.3) is the presence of the drift (or deterministic) part $(\delta^0)'$ $\langle Z_\Delta, Z_\Delta \rangle (v) \, \delta^0$. Without relating the magnitude of the break to the local variance condition, the order of the stochastic part dominates that of the deterministic part and so the latter vanishes asymptotically. The distributions in (4.2)-(4.3) share the same issues as Bai's and so they do not add any particular insight. We therefore move to discuss how to obtain a more useful continuous record asymptotic distribution.

Consider the set $\mathcal{D}(C) \triangleq \{N_b : N_b \in \{N_b^0 + Ch^{1-\kappa}\}, |C| < \infty\}$, on the original time scale. Let $\psi_h \triangleq h^{1-k}$. Here we use the same device as in Foster and Nelson (1994; 1996). Different scaling factors applied to an objective function can lead to different asymptotic distributions. We normalize $Q_T(T_b)$ by ψ_h, where ψ_h corresponds to the rate of convergence in Proposition 4.1. The rate of convergence implicitly describes the order of the terms in the expansion of $Q_T(T_b) - Q_T(T_b^0)$.

Lemma 4.1. *Under Assumption 2.1, 2.3-2.4, 3.1-3.2 and 4.1-4.2,*

$$\begin{aligned} & \left(Q_T(T_b) - Q_T\left(T_b^0\right)\right)/\psi_h \\ & \quad = -\delta_h' \left(Z_\Delta' Z_\Delta / \psi_h\right) \delta_h + 2\delta_h' \left(Z_\Delta' e / \psi_h\right) \operatorname{sgn}\left(T_b^0 - T_b\right) + o_p\left(h^{1/2}\right). \end{aligned} \tag{4.4}$$

For brevity, we use the notation $\pm$ in place of $\operatorname{sgn}(T_b^0 - T_b)$, henceforth. The conditional first moment of the centered criterion function $Q_T(T_b) - Q_T(T_b^0)$ is of order $O(h^{1-\kappa})$, i.e., it "oscillates" rapidly as $h \downarrow 0$. Hence, in order to approximate the behavior of $\{\widehat{T}_b - T_b^0\}$, we proceed as in Section 3 in Nelson and Foster (1994) and rescale "time". For any $C > 0$, let $L_C \triangleq N_b^0 - Ch^{1-\kappa}$ and $R_C \triangleq N_b^0 + Ch^{1-\kappa}$, where L_C and R_C are the left and right boundary points of $\mathcal{D}(C)$, respectively. We then have $|R_C - L_C| = O(Ch^{1-\kappa})$. Now, take the vanishingly small interval

$[L_C,\, R_C]$ on the original time scale, and stretch it into a time interval $[T^{1-\kappa}L_C,\, T^{1-\kappa}R_C]$ on a new "fast time scale". Changing time scale simply means that we rescale the objective function in such a way that it is of higher order as $h \downarrow 0$, i.e., it fluctuates less. This leads to an asymptotic distribution that accounts for higher uncertainty. Yet, under our framework it is still possible to consistently estimate the break fraction and the regression coefficients so that inference is feasible.

Since the criterion function is scaled by ψ_h^{-1}, all scaled processes are $O_p(1)$. Now, let $N_b(v) = N_b^0 - vh^{1-\kappa}$, $v \in [-C,\, C]$. Using Lemma 4.1 and Assumption 4.1 (see the appendix),

$$\begin{aligned}&\psi_h^{-1}\left(Q_T(T_b(v)) - Q_T\left(T_b^0\right)\right) = \\ &\quad -\delta_h'\left(\sum_{k=T_b(v)+1}^{T_b^0}\frac{z_{kh}}{\sqrt{\psi_h}}\frac{z_{kh}'}{\sqrt{\psi_h}}\right)\delta_h \pm 2\left(\delta^0\right)'\sum_{k=T_b(v)+1}^{T_b^0}\frac{z_{kh}}{\sqrt{\psi_h}}\frac{\widetilde{e}_{kh}}{\sqrt{\psi_h}} + o_p\left(h^{1/2}\right).\end{aligned}$$

where $\widetilde{e}_{kh} \triangleq h^{1/4}e_{kh}$. In addition, in view of (2.3), we let $dZ_{\psi,s} = \psi_h^{-1/2}\sigma_{Z,s}dW_{Z,s}$ for $s \in [N_b^0 - vh^{1-\kappa},\, N_b^0 + vh^{1-\kappa}]$. Applying the time scale change $s \to t \triangleq \psi_h^{-1}s$ to all processes including Σ^0, we have $dZ_{\psi,t} = \sigma_{Z,t}dW_{Z,t}$ with $t \in \mathcal{T}(C)$, where $\mathcal{T}(C) \triangleq \{t:\, t \in \left[N_b^0 + v\left\|\delta^0\right\|^2/\overline{\sigma}^2\right],\, |v| \leq C\}$. Therefore,

$$\begin{aligned}&\psi_h^{-1}\left(Q_T(T_b(v)) - Q_T\left(T_b^0\right)\right) \\ &\quad = -\delta_h'\left(\sum_{k=T_b(v)+1}^{T_b^0} z_{\psi,kh}z_{\psi,kh}'\right)\delta_h \pm 2\left(\delta^0\right)'\sum_{k=T_b(v)+1}^{T_b^0} z_{\psi,kh}\widetilde{e}_{\psi,kh} + o_p\left(h^{1/2}\right),\end{aligned}$$

with $NT_b(v)/T = N_b(v) = N_b^0 + v$, where $z_{\psi,kh} \triangleq z_{kh}/\sqrt{\psi_h}$ and $\widetilde{e}_{\psi,kh} \triangleq \widetilde{e}_{kh}/\sqrt{\psi_h}$. Because of the change of time scale, all processes in the last display are scaled up to be $O_p(1)$ and thus behave as diffusion-like processes. On this new "fast time scale", we have $T^{1-\kappa}R_C - T^{1-\kappa}L_C = O(1)$ and $Q_T(T_b(v)) - Q_T(T_b^0)$ is restored to be $O_p(1)$. Observe that changing the time scale does not affect any statistic which depends on observations from $k = 1$ to $k = \lfloor L_C/h \rfloor$ or from $k = \lfloor R_C/h \rfloor$ to $k = T$ (since these involve a positive fraction of data). However, it does affect quantities which include observations that fall in $[T_bh,\, T_b^0h]$ (assuming $T_b < T_b^0$). In particular, on the original time scale, the processes $\{D_t\}$, $\{Z_t\}$ and $\{e_t\}$ are well-defined and scaled to be $O_p(1)$ while $Q_T(T_b) - Q_T(T_b^0)$ (asymptotically) oscillates more rapidly than a simple diffusion-type process. On the new "fast time scale", $\{D_t\}$, $\{Z_t\}$ and $\{e_t\}$ are not affected since they have the same order in $[T^{1-\kappa}L_C,\, T^{1-\kappa}R_C]$ as $h \downarrow 0$. That is, the first conditional moments are $O(h)$ while the corresponding moments for $Q_T(T_b) - Q_T(T_b^0)$ on $\mathcal{T}(C)$ are restored to be $O(h)$. As the continuous-time limit is approached, the rescaled criterion function $(Q_T(T_b(v)) - Q_T(T_b^0))/h^{1/2}$ operates on a "fast time scale" on $\mathcal{T}(C)$.

Our analysis is local; we examine the limiting behavior of the centered and rescaled criterion function process in a neighborhood $\mathcal{T}(C)$ of the the true break date N_b^0 defined on a new time scale. We first obtain the weak convergence results for the statistic $\left(Q_T\left(T_b\left(v\right)\right)-Q_T\left(T_b^0\right)\right)/h^{1/2}$ and then apply a continuous mapping theorem for the argmax functional. However, it is convenient to work with a re-parametrized objective function. Proposition 4.1 allows us to use

$$\overline{Q}_T\left(\theta^*\right)=\left(Q_T\left(\theta_h,\,T_b\left(v\right)\right)-Q_T\left(\theta^0,\,T_b^0\right)\right)/h^{1/2},$$

where $\theta^* \triangleq (\theta_h',\,v)'$ with $T_b\left(v\right)\triangleq T_b^0+\lfloor v/h\rfloor$ and $T_b\left(v\right)$ is the time index on the "fast time scale". The normalizing factor $\psi_h h^{1/2}$ allows us to change the time scale and obtain an alternative asymptotic distribution. When v varies, $T_b\left(v\right)$ potentially visits all integers between 1 and T. Thus, on the new time scale, we need to introduce the trimming parameter $\pi\in(0,\,1)$ which determines the region where $T_b\left(v\right)$ can vary. We have the normalizations $T_b\left(v\right)=T\pi$ if $T_b\left(v\right)\leq T\pi$ and $T_b\left(v\right)=T\left(1-\pi\right)$ if $T_b\left(v\right)\geq T\left(1-\pi\right)$. On the old time scale, $N_b\left(u\right)=N_b^0+u$ with $v\rightarrow\psi_h^{-1}u$, so that $N_b\left(u\right)$ is in a vanishing neighborhood of N_b^0. On $\mathcal{T}\left(C\right)$, we index the process $Q_T\left(\theta_h,\,T_b\left(v\right)\right)-Q_T\left(\theta^0,\,T_b^0\right)$ by two time subscripts: one referring to the time T_b on the original time scale and one referring to the time elapsed since $T_b h$ on the "fast time scale". For simplicity, we omit the former; since the limiting distribution of the least-squares estimator will now depend on the trimming we use the notation $\widehat{T}_{b,\pi}=T\widehat{\lambda}_{b,\pi}$ where $\widehat{\lambda}_{b,\pi}$ is the least-squares estimator of the fractional break date associated to the fast time scale (i.e., associated to the factor $\psi_h h^{1/2}$).

The optimization problem is not affected by the change of time scale. In fact, by Proposition 4.1, $u=Th(\widehat{\lambda}_b-\lambda_0)=KO_p\left(h^{1-\kappa}\right)$ on the old time scale; whereas on the new "fast time scale", $v=Th(\widehat{\lambda}_{b,\pi}-\lambda_0)=O_p\left(1\right)$. The maximization problem is not changed because v/h can take any value in $\mathbb{R}$. The process $Q_T\left(\theta_h,\,T_b\left(v\right)\right)-Q_T\left(\theta^0,\,T_b^0\right)$ is thus analyzed on a fixed horizon since v now varies over $[(N\pi-N_b^0)/(||\delta^0||^{-2}\overline{\sigma}^2),\,(N\left(1-\pi\right)-N_b^0)/(\|\delta^0\|^{-2}\,\overline{\sigma}^2)]$. Define the modification to the set $\mathcal{D}\left(C\right)$ applicable to the new time scale by

$$\mathcal{D}^*\left(C\right)=\left\{\left(\beta^0,\,\delta_h,\,v\right):\,\left\|\theta^0\right\|\leq C;\,T_b\left(v\right)=T_b^0+vN^{-1}\left\|\delta^0\right\|^{-2}\overline{\sigma}^2;\right.$$
$$\left.\frac{(N\pi-N_b^0)}{\|\delta^0\|^{-2}\,\overline{\sigma}^2}\leq v\leq\frac{N\left(1-\pi\right)-N_b^0}{\|\delta^0\|^{-2}\,\overline{\sigma}^2}\right\}.$$

Let $\mathbb{D}\left(\mathcal{D}^*\left(C\right),\,\mathbb{R}\right)$ denote the space of all *càdlàg* functions from $\mathcal{D}^*\left(C\right)$ into $\mathbb{R}$. Endow this space with the Skorokhod topology. Under a continuous record, we can apply limit theorems for statistics involving (co)variation between regressors and errors. This enables us to deduce the limiting process for $\overline{Q}_T\left(\theta^*\right)$, relying upon the work of Jacod (1994; 1997) and Jacod and Protter (1998).

To guide intuition, note that under the new re-parametrization, the limit law of $\overline{Q}_T\left(\theta^*\right)$ is, according to Lemma 4.1, the same as the limit law of

$$-h^{-1/2}\delta_h'\left(Z_\Delta'Z_\Delta\right)\delta_h\pm2h^{-1/2}\delta_h'\left(Z_\Delta'e\right)$$

$$\overset{d}{\equiv} -\left(\delta^0\right)'\left(Z_\Delta' Z_\Delta\right)\delta^0 \pm 2h^{-1/2}\left(\delta^0\right)' h^{1/4}\left(Z_\Delta' h^{-1/4}\widetilde{e}\right),$$

where $\overset{d}{\equiv}$ denotes (first order) equivalence in law, and since (approximately) $e_{kh} \sim i.n.d.\,\mathscr{N}(0, \sigma_{h,k-1}^2 h)$, $\sigma_{h,k} = \sigma_h \sigma_{e,k}$ then $\widetilde{e}_{kh} \sim i.n.d.\,\mathscr{N}\left(0,\, \sigma_{e,k-1}^2 h\right)$. Hence, the limit law of $\overline{Q}_T\left(\theta^*\right)$ is, to first-order, equivalent to the law of

$$-\left(\delta^0\right)'\left(Z_\Delta' Z_\Delta\right)\delta^0 \pm 2\left(\delta^0\right)'\left(h^{-1/2} Z_\Delta' \widetilde{e}\right). \tag{4.5}$$

We apply a law of large numbers to the first term and a stable convergence in law under the Skorokhod topology to the second. Assumption 4.1 combined with the normalizing factor $h^{-1/2}$ in $\overline{Q}_T\left(\theta^*\right)$ account for the discrepancy between the deterministic and stochastic component in (4.5).

Having outlined the main steps in the arguments used to derive the continuous records limit distribution of the break date estimate, we now state the main result of this section. The limiting process is realized on a extension of the original probability space and we relegate this description to Section S.B in the supplement.

Theorem 4.1. *Under Assumption 2.1, 2.3-2.4, 3.1-3.2 and 4.1-4.2,*

$$N\left(\widehat{\lambda}_{b,\pi} - \lambda_0\right) \overset{\mathcal{L}-\mathrm{s}}{\Rightarrow} \underset{v\in\mathcal{A}}{\operatorname{argmax}}\left\{-\left(\delta^0\right)'\left\langle Z_\Delta,\, Z_\Delta\right\rangle(v)\,\delta^0 + 2\left(\delta^0\right)'\mathscr{W}(v)\right\}, \tag{4.6}$$

where

$$\mathcal{A} \triangleq \left[\frac{N\pi - N_b^0}{\left\|\delta^0\right\|^{-2}\overline{\sigma}^2},\, \frac{N(1-\pi) - N_b^0}{\left\|\delta^0\right\|^{-2}\overline{\sigma}^2}\right].$$

Note the differences between the results in Theorem 4.1 and in Proposition 4.2. First, on the fast time scale, $\widehat{\lambda}_{b,\pi}$ behaves as an inconsistent estimator for λ_0 for N fixed, but it is consistent as $N\rightarrow\infty$. On the original time scale $\widehat{\lambda}_b$ is not only consistent for λ_0 but it also enjoys a similar asymptotic distribution as in Bai (1997). Second, the asymptotic distribution of $\widehat{\lambda}_{b,\pi}$ depends on the span of the data and consequently on the trimming π. Proposition 4.2, in contrast, suggests that the span, the trimming and the location of the break are irrelevant for the limiting behavior of the estimator. This intuitively follows from the fact that under the original time scale the break date estimator is consistent. We will show that indeed the span of the data and the location of the break influence the finite-sample properties of the least-squares estimator, and that Theorem 4.1 provides a more useful approximation. An important implication of Theorem 4.1 is that the precision of the estimator depends more on the span N than to the number of observations T.

Unlike Bai's distribution, the distribution in Theorem 4.1 involves the location of the maximum of a function of the (quadratic) variation of the regressors and of a two-sided centered Gaussian martingale process over the interval $[(N\pi - N_b^0)/(\left\|\delta^0\right\|^{-2}\overline{\sigma}^2),\, (N(1-\pi) - N_b^0)/(\|\delta^0\|^{-2}\overline{\sigma}^2)]$.

Notably, this domain depends on the true value of N_b^0 and therefore the limit distribution is asymmetric, in general. The degree of asymmetry increases as the true break point moves away from mid-sample. This holds even when the distributions of the errors and regressors are the same in the pre- and post-break regimes. The presence of the trimming confirms that the span of the (trimmed) data affects the limit distribution. It is well-known that the least-squares estimator of the break date can be sensitive to trimming [see Bai and Perron (2003) for some recommendations on the trimming choice]. Our asymptotic theory accommodates this property of the least-squares estimator while others do not.

Additional relevant remarks follow; more details are provided in the supplement. The magnitude of the break plays a key role in determining the density of the asymptotic distribution. More precisely, the density displays interesting properties which change when the signal-to-noise ratio as well as other parameters of the model change. Moreover, the distribution in Theorem 4.1 is able to reproduce important features of the small-sample results obtained via simulations [e.g., Bai and Perron (2006)]. First, the second moments of the regressors impact the asymptotic mean as well as the second-order behavior of the break point estimator (e.g., the persistence of the regressors influences the finite-sample performance of the estimator). Second, the continuous record setting manages to preserve information about the time span N of the data, a clear advantage since the location of the true break point matters for the small-sample distribution of the estimator. It has been shown via simulations that in small-samples the break point estimator tends to be imprecise if the break size is small, and some bias arises if the break point is not at mid-sample. In our framework, the (trimmed) time horizon $[N\pi,\, N(1-\pi)]$ is fixed and thus we can distinguish between the statistical content of the segments $[N\pi,\, N_b^0]$ and $[N_b^0,\, N(1-\pi)]$. In contrast, this is not feasible under the classical shrinkage large-N asymptotics because both the pre- and post-break segments increase proportionately and mixing conditions are imposed so that the only relevant information is a neighborhood around the true break date. Details on how to simulate the limiting distribution in Theorem 4.1 are given in Section S.A of the supplement.

We further characterize the asymptotic distribution by exploiting the ($\mathscr{F}$-conditionally) Gaussian property of the limit process. The analysis also holds unconditionally if we assume that the volatility processes are non-stochastic. Thus, as in the classical setting, we begin with a second-order stationarity assumption within each regime. The following assumption guarantees that the results below remain valid without the need to condition on $\mathscr{F}$.

Assumption 4.3. *The process Σ^0 is (possibly time-varying) deterministic; $\{z_{kh},\, e_{kh}\}$ is second-order stationary within each regime. For $k = 1, \ldots, T_b^0$, $\mathbb{E}(z_{kh}z'_{kh}|\,\mathscr{F}_{(k-1)h}) = \Sigma_{Z,1}h$, $\mathbb{E}(\tilde{e}^2_{kh}|\,\mathscr{F}_{(k-1)h}) = \sigma^2_{e,1}h$ and $\mathbb{E}(z_{kh}z'_{kh}\tilde{e}^2_{kh}|\,\mathscr{F}_{(k-1)h}) = \Omega_{\mathscr{W},1}h^2$ while for $k = T_b^0+1, \ldots, T$, $\mathbb{E}(z_{kh}z'_{kh}|\,\mathscr{F}_{(k-1)h}) = \Sigma_{Z,2}h$, $\mathbb{E}(\tilde{e}^2_{kh}|\,\mathscr{F}_{(k-1)h}) = \sigma^2_{e,2}h$ and $\mathbb{E}(z_{kh}z'_{kh}\tilde{e}^2_{kh}|\,\mathscr{F}_{(k-1)h}) = \Omega_{\mathscr{W},2}h^2$.*

Let W_i^*, $i = 1, 2$, be two independent standard Wiener processes defined on $[0, \infty)$, starting at the origin when $s=0$. Let

$$\mathscr{V}(s)=\begin{cases}-\frac{|s|}{2}+W_1^*(s), & \text{if } s<0\\ -\frac{(\delta^0)'\Sigma_{Z,2}\delta^0}{(\delta^0)'\Sigma_{Z,1}\delta^0}\frac{|s|}{2}+\left(\frac{(\delta^0)'\Omega_{\mathscr{W},2}(\delta^0)}{(\delta^0)'\Omega_{\mathscr{W},1}(\delta^0)}\right)^{1/2}W_2^*(s), & \text{if } s\geq 0.\end{cases}$$

Theorem 4.2. *Under Assumption 2.1, 2.3-2.4, 3.1-3.2 and 4.1-4.3,*

$$\frac{\left((\delta^0)'\langle Z,Z\rangle_1\delta^0\right)^2}{(\delta^0)'\Omega_{\mathscr{W},1}\delta^0}N\left(\widehat{\lambda}_{b,\pi}-\lambda_0\right)\Rightarrow\underset{s\in\mathcal{A}^*}{\operatorname{argmax}}\mathscr{V}(s), \tag{4.7}$$

where

$$\mathcal{A}^*\triangleq\left[\frac{N\pi-N_b^0}{\|\delta^0\|^{-2}\overline{\sigma}^2}\frac{\left((\delta^0)'\langle Z,Z\rangle_1\delta^0\right)^2}{(\delta^0)'\Omega_{\mathscr{W},1}(\delta^0)},\ \frac{N(1-\pi)-N_b^0}{\|\delta^0\|^{-2}\overline{\sigma}^2}\frac{\left((\delta^0)'\langle Z,Z\rangle_1\delta^0\right)^2}{(\delta^0)'\Omega_{\mathscr{W},1}\delta^0}\right].$$

Unlike the asymptotic distribution derived under classical large-N asymptotics, the probability density in (4.7) is not available in closed form. Furthermore, the limiting distribution depends on unknown quantities. In the next section we explain how one can derive a feasible counterpart. This will be useful to characterize the main features of interest that will guide us in devising methods to construct confidence sets for T_b^0.

5 Feasible Approximations to the Finite-Sample Distributions

In Section 5.1 we propose a feasible version of our limit theory and compare it with the finite-sample distribution. In Section 5.2 we discuss some differences between our approach and others. Let

$$\rho=\frac{\left((\delta^0)'\langle Z,Z\rangle_1\delta^0\right)^2}{\left((\delta^0)'\Omega_{\mathscr{W},1}\delta^0\right)},\quad \xi_1=\frac{(\delta^0)'\langle Z,Z\rangle_2\delta^0}{(\delta^0)'\langle Z,Z\rangle_1\delta^0},\quad \xi_2=\frac{(\delta^0)'\Omega_{\mathscr{W},2}\delta^0}{(\delta^0)'\Omega_{\mathscr{W},1}\delta^0}.$$

5.1 A Feasible Version of the Limit Distribution

In order to use the continuous record asymptotic distribution in practice one needs consistent estimates of the unknown quantities. In this section, we compare the finite-sample distribution of the least-squares estimator of the change-point date with a feasible version of the continuous record asymptotic distribution obtained with plug-in estimates. We obtain the finite-sample distribution of $\rho(\widehat{T}_{b,\pi}-T_b^0)$ based on 100,000 simulations from the following model:

$$Y_t = D_t'\nu^0 + Z_t'\beta^0 + Z_t'\delta_Z^0 \mathbf{1}_{\left\{t>T_b^0\right\}} + e_t, \qquad t=1,\ldots,T, \tag{5.1}$$

where $Z_t = 0.5Z_{t-1} + u_t$ with $u_t \sim i.i.d.\, \mathscr{N}(0,\,1)$ independent of $e_t \sim i.i.d.\, \mathscr{N}(0,\,\sigma_e^2)$, $\sigma_e^2=1$, $\nu^0=1$, $Z_0=0$, $D_t=1$ for all t, and $T=100$. We set $\pi=0.05$, $T_b^0=\lfloor T\lambda_0 \rfloor$ with $\lambda_0 = 0.3,\, 0.5,\, 0.7$ and consider different break sizes $\delta_Z^0 = 0.2,\, 0.3,\, 0.5,\, 1$. The infeasible continuous record asymptotic distribution is computed assuming knowledge of the data generating process (DGP) as well as of the model parameters, i.e., using Theorem 4.2 where we set N_b^0 equal to its true value, $||\delta^0||^{-2}\overline{\sigma}^2 = ||\delta_Z^0||^{-2}\sigma_e^2$, and ξ_1, ξ_2 and ρ equal to their true values, respectively, with δ_Z^0 in place of δ^0. Note that the scaling $h^{\kappa/2}$ and $h^{-1/4}$ in the definition of δ_h and σ_h respectively, cancel using the fact that they appear in both numerator and denominator and applying a change in variables. The feasible counterparts are constructed with plug-in estimates of ξ_1, ξ_2, ρ and $(N_b^0 \left\|\delta^0\right\|^2/\overline{\sigma}^2)\rho$. In practice we need to use a normalization for N. A common choice is $N=1$. Then $\widehat{\lambda}_b = \widehat{T}_b/T$ is a natural estimate of λ_0, using the consistency result of $\widehat{\lambda}_b$ that holds in the setting of Theorem 4.1 which can also be rationalized for large N under the conditions of Theorem 4.1. In practice this means that we approximate the distribution of the estimator $\widehat{\lambda}_{b,\pi}$ where π is chosen by the researcher and we plug-in the estimator $\widehat{\lambda}_b$ which can be based on any trimming because of the consistency property. Here, we set $\widehat{\lambda}_b$ equal to the least-squares estimator based on a trimming 0.15, which is also used for the other plug-in estimates. The estimates of ξ_1 and ξ_2 are given, respectively, by

$$\widehat{\xi}_1 = \frac{\widehat{\delta}'\left(T-\widehat{T}_b\right)^{-1}\sum_{k=\widehat{T}_b+1}^{T} z_{kh}z_{kh}'\widehat{\delta}}{\widehat{\delta}'\left(\widehat{T}_b\right)^{-1}\sum_{k=1}^{\widehat{T}_b} z_{kh}z_{kh}'\widehat{\delta}}, \qquad \widehat{\xi}_2 = \frac{\widehat{\delta}'\left(T-\widehat{T}_b\right)^{-1}\sum_{k=\widehat{T}_b+1}^{T} \widehat{e}_{kh}^2 z_{kh}z_{kh}'\widehat{\delta}}{\widehat{\delta}'\left(\widehat{T}_b\right)^{-1}\sum_{k=1}^{\widehat{T}_b} \widehat{e}_{kh}^2 z_{kh}z_{kh}'\widehat{\delta}},$$

where $\widehat{\delta}$ is the least-squares estimator of δ_h and $\widehat{e}_{kh}$ are the least-squares residuals. Note that in $\widehat{\xi}_1$ and $\widehat{\xi}_2$, the estimate $\widehat{\delta}$ appears in both numerator and denominator so that the scaling $h^{\kappa/2}$ in the definition of δ_h cancels. Use is made of the fact that $\langle Z,\, Z\rangle_1$ is consistently estimated by $\sum_{k=1}^{\widehat{T}_b} z_{kh}z_{kh}'/\widehat{\lambda}_b$ while $\Omega_{\mathscr{W},1}$ is consistently estimated by $T\sum_{k=1}^{\widehat{T}_b} \widehat{e}_{kh}^2 z_{kh}z_{kh}'/\widehat{\lambda}_b$. The method to estimate $\lambda_0 \left\|\delta^0\right\|^2 \overline{\sigma}^{-2}\rho$ is less immediate because it involves manipulating the scaling of each of the three estimates. Let $\vartheta = \left\|\delta^0\right\|^2 \overline{\sigma}^{-2}\rho$. We use the following estimates for ϑ and ρ, respectively,

$$\widehat{\vartheta} = \widehat{\rho}\left\|\widehat{\delta}\right\|^2 \left(T^{-1}\sum_{k=1}^{T}\widehat{e}_{kh}^2\right)^{-1}, \qquad \widehat{\rho} = \frac{\left(\widehat{\delta}'\left(\widehat{T}_b\right)^{-1}\sum_{k=1}^{\widehat{T}_b} z_{kh}z_{kh}'\widehat{\delta}\right)^2}{\widehat{\delta}'\left(\widehat{T}_b\right)^{-1}\sum_{k=1}^{\widehat{T}_b} \widehat{e}_{kh}^2 z_{kh}z_{kh}'\widehat{\delta}},$$

Whereas we have $\widehat{\xi}_i \xrightarrow{p} \xi_i$ $(i=1,\,2)$, the corresponding approximations for $\widehat{\rho}$ and $\widehat{\vartheta}$ are given by $\widehat{\rho}/h^{\kappa} \xrightarrow{p} \rho$ and $\widehat{\vartheta}/h^{2\kappa} \xrightarrow{p} \vartheta$. However, before letting $T\to\infty$ we can apply a change in variable using the fact that $\widehat{\lambda}_b - \lambda_b^0 = O\left(h^{1-\kappa}\right)$ which result in the extra factor $h^{2\kappa}$ canceling.

Proposition 5.1. *Under the conditions of Theorem 4.2, (4.7) holds when using $\widehat{\xi}_1$, $\widehat{\xi}_2$, $\widehat{\rho}$ and $\widehat{\vartheta}$ in place of ξ_1, ξ_2, ρ and ϑ, respectively.*

The proposition implies that the limiting distribution can be simulated using plug-in estimates. This allows feasible inference about the break date. The results are presented in Figure 1-4 which also plot the asymptotic distribution from Bai (1997) and the infeasible distribution from Theorem 4.2. Here by signal-to-noise ratio we mean δ_Z^0/σ_e which, given $\sigma_e^2 = 1$, equals the break size δ_Z^0.

Several interesting observations appear at the outset. The density of the large-N shrinkage asymptotic distribution does not depend on the location of the break, and thus it is always unimodal and symmetric about the origin. None of these features are shared by the density derived under a continuous record. When the true break is at mid-sample ($\lambda_0 = 0.5$), the density function is symmetric and centered at zero. However, when the signal-to-noise ratio is low, the density features three modes. This tri-modality vanishes as the signal-to-noise ratio increases. When δ_Z^0 is low and the break is not at mid-sample the density is asymmetric; for values of λ_0 less (larger) than 0.5, the density is right (left) skewed. When the signal is low and λ_0 is less (larger) than 0.5, the density has highest mode at some value near $\widehat{\lambda}_b$ being close to the starting (end) sample point than centered at λ_0. However, as in the case of $\lambda_0 = 0.5$, when the signal-to-noise ratio increases the highest mode is centered at a value which corresponds to $\widehat{\lambda}_b$ being close to λ_0. Asymmetry and multi-modality of the finite-sample distribution of the break point estimator were also found by Perron and Zhu (2005) and Deng and Perron (2006) in models with a trend.

The interpretation of these features are straightforward. For example, asymmetry reflects the fact that the span of the data and the actual location of the break play a crucial role on the behavior of the estimator. If the break occurs early in the sample there is a tendency to overestimate the break date and vice-versa if the break occurs late in the sample. The marked changes in the shape of the density as we raise δ_Z^0 confirms that the magnitude of the shift matters a great deal as well. The tri-modality of the density when the shift size is small reflects the uncertainty in the data as to whether a structural change is present at all; i.e., the least-squares estimator finds it easier to locate the break at either the beginning or the end of the sample. Unlike the large-N shrinkage asymptotic distribution, the density of the feasible version of the continuous record distribution provides a remarkably good approximation to the infeasible one and thus also to the finite-sample distribution. The extended working paper Casini and Perron (2020a) shows that the quality of the approximation is good for a variety of models.

5.2 Comparison with Other Approaches

The figures reported above have shown that there is a high degree of uncertainty when the break magnitude is not large. The classical shrinkage asymptotics of Bai (1997) with δ_T required to convergence to zero at a rate slower than $O(T^{-1/2})$ clearly underestimates that degree of uncertainty and, as the figures show, provides a poor approximation to the finite-sample behavior of the least-squares estimator. In Section 7 we show that this issue is responsible for the poor coverage probabilities of the confidence intervals introduced in Bai (1997) when the break magnitude is small. On the other hand, Elliott and Müller (2007) and Elliott et al. (2015) require δ_T to go to zero at the fast rate $O(T^{-1/2})$ leading to weak identification. The latter implies that the relevant quantities in the model become inconsistent. This can be problematic for inference and indeed, their inference often suffers from the opposite problem in that confidence intervals for $\widehat{T}_b$ can be too large [Casini and Perron (2019; 2021b) and Chang and Perron (2018)].

We impose conditions on the signal-to-noise ratio δ/σ rather than just on δ. Consider a simple location model with a change δ in the mean and independent errors. What describes the uncertainty about the break in this model is the ratio δ/σ where σ is the volatility of the errors. We let δ go to zero at a not too fast rate while letting σ increase to infinity in a neighborhood of T_b^0. That is $(\delta_T/\sigma_t) \to 0$ at rate $O(T^{-1/2})$ in a neighborhood of T_b^0. Interestingly, this is the same rate Elliott and Müller used for $\delta_T \to 0$. Away from T_b^0, we require $(\delta_T/\sigma_t) \to 0$ at slower rate—similar to Yao (1987) and Bai (1997). The difference now is that we do not lose identification and all the parameters in the model remain consistent. Under continuous-time, the variance of the processes is proportional to the sampling interval. This allows us to trade-off the rate of convergence at which $\widehat{\lambda}_b$ approaches λ_0 with the variance of the errors in a neighborhood of T_b^0 by letting σ_t become large when t is close to T_b^0 [i.e., a change of time scale as in Foster and Nelson (1994, 1996)]. This offers a new characterization of higher uncertainty without losing identification.

6 Highest Density Region-based Confidence Sets

The features of the limit and finite-sample distributions suggest that standard methods to construct confidence intervals may be inappropriate; e.g., two-sided intervals around the estimated break date based on the standard deviations of the estimate. Our suggested approach is rather non-standard and relates to Bayesian methods. In our context, the Highest Density Region (HDR) seems the most appropriate in light of the asymmetry and, especially, the multi-modality of the distribution for small break sizes. All that is needed to implement the procedure is an estimate of the density function, using plug-in estimates as explained in Section 5. Choose some significance level $0 < \alpha < 1$ and let $\widehat{P}_{T_b}$ denote the empirical counterpart of the probability distribution of $\rho N(\widehat{\lambda}_{b,\pi} - \lambda_b^0)$ as defined in Theorem 4.2. Further, let $\widehat{p}_{T_b}$ denote the empirical density function defined by the Radon-Nikodym equation $\widehat{p}_{T_b} = d\widehat{P}_{T_b}/d\lambda_{\mathrm{L}}$, where λ_{L} denotes the Lebesgue measure.

Definition 6.1. Highest Density Region: Assume that the density function $f_Y(y)$ of some random variable Y defined on a probability space $(\Omega_Y, \mathscr{F}_Y, \mathbb{P}_Y)$ and taking values on the measurable space

$(\mathcal{Y}, \mathscr{Y})$ is continuous and bounded. Then the $(1-\alpha)\,100\%$ Highest Density Region is a subset $\mathbf{S}(\kappa_\alpha)$ of $\mathcal{Y}$ defined as $\mathbf{S}(\kappa_\alpha) = \{y : f_Y(y) > \kappa_\alpha\}$ where κ_α the largest constant that satisfies $\mathbb{P}_Y(Y \in \mathbf{S}(\kappa_\alpha)) \geq 1-\alpha$.

The concept of HDR and of its estimation has an established literature in statistics. The definition reported here is from Hyndman (1996); see also Samworth and Wand (2010) and Mason and Polonik (2008; 2009) for more recent developments.

Definition 6.2. Confidence Sets for T_b^0 under a Continuous Record: Under Assumption 2.1, 2.3-2.4, 3.1-3.2 and 4.1-4.3, a $(1-\alpha)\,100\%$ confidence set for T_b^0 is a subset of $\{1,\ldots,T\}$ given by $C(\mathrm{cv}_\alpha) = \{T_b \in \{1,\ldots,T\} : T_b \in \mathbf{S}(\mathrm{cv}_\alpha)\}$, where $\mathbf{S}(\mathrm{cv}_\alpha) = \{T_b : \widehat{p}_{T_b} > \mathrm{cv}_\alpha\}$ and cv_α satisfies $\sup_{\mathrm{cv}_\alpha \in \mathbb{R}_+} \widehat{P}_{T_b}(T_b \in \mathbf{S}(\mathrm{cv}_\alpha)) \geq 1-\alpha$.

The confidence set $C(\mathrm{cv}_\alpha)$ has a frequentist interpretation even though the concept of HDR is often encountered in Bayesian analyses since it associates naturally to the derived posterior distribution, especially when the latter is multi-modal. A feature of the confidence set $C(\mathrm{cv}_\alpha)$ under our context is that, at least when the size of the shift is small, it consists of the union of several disjoint intervals. The appeal of using HDR is that one can directly deal with such features. As the break size increases and the distribution becomes unimodal, the HDR becomes equivalent to the standard way of constructing confidence sets. In practice, one can proceed as follows.

Algorithm 1. Confidence sets for T_b^0: *1) Estimate by least-squares the break point and the regression coefficients from model (2.4); 2) Replace quantities appearing in (4.7) by consistent estimators as explained in Section 5; 3) Simulate the limiting distribution $\widehat{P}_{T_b}$ from Theorem 4.2; 4) Compute the HDR of the empirical distribution $\widehat{P}_{T_b}$ and include the point T_b in the level $1-\alpha$ confidence set $C(cv_\alpha)$ if T_b satisfies the conditions in Definition 6.2.*

This procedure will not deliver contiguous confidence sets when the size of the break is small. Indeed, we find that in such cases, the overall confidence set for T_b^0 consists in general of the union of disjoint intervals if $\widehat{T}_b$ is not near the tails of the sample. One is located around the estimate of the break date, while the others are in the pre- and post-break regimes. To provide an illustration, we consider a simple example involving a single draw from a simulation experiment. Figure 5 reports the HDR of the feasible limiting distribution of $\rho(\widehat{T}_{b,\pi} - T_b^0)$ for a random draw from the model in (5.1) with parameters $\nu^0 = 1$, $\beta^0 = 0$, unit variance and autoregressive coefficient 0.6 for Z_t and $\sigma_e^2 = 1.2$. We set $\lambda_0 = 0.35$, 0.5 and $\delta_Z^0 = 0.3$, 0.8, 1.5. We use a trimming 0.15 for the plug-in estimator $\widehat{T}_b$ and $\pi = 0.05$ for $\widehat{T}_{b,\pi}$. As explained in Section 5.1, we could use any other trimming in place of 0.15. The results remain unchanged. We set $T = 100$ and the significance level is $\alpha = 0.05$. Note that the origin is at the estimated break date. The point on the horizontal axis corresponds to the true break date. The black intervals on the horizontal axis correspond to regions of high density. The resulting confidence set is their union. Once a confidence region for $\rho(\widehat{T}_{b,\pi} - T_b^0)$ is computed, it is straightforward to derive a 95% confidence set for T_b^0. The top panel (left plot) reports results for the case $\delta_Z^0 = 0.3$ and $\lambda_0 = 0.35$ and shows that the HDR is composed of two disjoint intervals. The estimated break date is $\widehat{T}_b = 70$ and the implied 95% confidence set for T_b^0 is given by $C(\text{cv}_{0.05}) = \{1, \ldots, 12\} \cup \{18, \ldots 100\}$. This includes T_b^0 and the overall length is 95 observations. Table 1 reports for various methods whether T_b^0 is covered or not and the length of the confidence sets for this example. The length of Bai's (1997) confidence interval is 55 but does not include T_b^0. Elliott and Müller's (2007) confidence set, denoted by $\widehat{U}_T$.eq in Table 1, also does not include the true break date at the 90% confidence level, but does so at the 95% and its length is 95. Our method covers T_b^0 and has a relatively shorter length across different δ_Z^0.

7 Small-Sample Properties of the HDR Confidence Sets

We now assess via simulations the finite-sample performance of the method proposed to construct confidence sets for the break date. We also make comparisons with alternative methods in the literature: Bai's (1997) approach based on the large-N shrinkage asymptotics; Elliott and Müller's (2007), hereafter EM, method on inverting Nyblom's (1989) statistic; the Inverted Likelihood Ratio (ILR) approach of Eo and Morley (2015). We omit the technical details of these methods and refer to the original sources or Chang and Perron (2018) for a review and comparisons. We consider two DGPs: M1 is $y_t = \beta^0 + \delta_Z^0 \mathbf{1}_{\{t > T_b^0\}} + e_t$ with $\beta^0 = 1$ and $e_t \sim i.i.d.\,\mathscr{N}(0, 1)$; M2 is $y_t = \delta_Z^0 (1 - \nu^0) \mathbf{1}_{\{t > T_b^0\}} + \nu^0 y_{t-1} + e_t$ with $\nu^0 = 0.8$ and $e_t \sim i.i.d.\,\mathscr{N}(0, 0.04)$. Our companion paper Casini and Perron (2020a) includes extensive simulation results. We set the significance level at $\alpha = 0.05$, and the break occurs at date $\lfloor T\lambda_0 \rfloor$, where $\lambda_0 = 0.2$, 0.35, 0.5 and $T = 200$ for M1 and $T = 100$ for M2. The results are presented in Table 2-3. The last row in each table includes the rejection probability of a 5%-level sup-Wald test using the asymptotic critical value in Andrews (1993), which provides a measure of the magnitude of the break relative to the noise. For models with predictable processes we use the two-step procedure described in Section S.C.2.

Overall, the simulation results 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. Overall, our HDR method and that of EM show accurate empirical coverage rates for all DGP considered. However, EM's method almost always displays confidence sets which are larger than those from the other approaches.[1] Over all DGPs considered, the average length of the HDR confidence sets are 40% to 70% shorter than those obtained with EM's approach when the size of the shift is moderate to high. The results for M2, a change in mean with a lagged dependent

[1]This problem is more severe when the errors are serially correlated or the model includes lagged dependent variables (see also the supplement). Regarding the former, this in part may be due to issues with Newey and West HAC-type estimators when there are breaks [see, e.g., Casini (2022) and Casini, Deng and Perron (2022)].

variable and strong correlation, are quite revealing. EM's method yields confidence intervals that are very wide, increasing with the size of the break and for large breaks covering nearly the entire sample. This does not occur with the other methods. For instance, when $\lambda_0 = 0.5$ and $\delta_Z^0 = 2$, the average length from the HDR method is 8.34 compared to 93.71 with EM's. This concurs with the results in Chang and Perron (2018).

In summary, the small-sample simulation results suggest that our continuous record HDR-based inference provides accurate coverage probabilities close to the nominal level and average lengths of the confidence sets shorter relative to existing methods. It is also valid and reliable under a wider range of DGPs including long-memory processes. Specifically noteworthy is the fact that it performs well for all break sizes, whether small or large.

8 Empirical Application

We apply the HDR method of constructing confidence sets about the break dates in monthly U.S. industrial production growth. The data were obtained from the Federal Reserve Economic Data website for the period 1990:01–2020:01. Annualized monthly growth rates are calculated as $1,200 \times \Delta y_t$ where Δy_t is the first difference of the natural logarithms of the levels data. We assume that the log industrial production follows a trend plus a potentially autocorrelated noise,

$$\Delta y_t = \beta_t + u_t, \qquad t = 1, \ldots, T,$$

where β_t has a finite number of breaks and $\{u_t\}$ is a zero-mean potentially serially correlated series with a bounded spectral density at frequency zero. Figure 6 plots Δy_t. The series resembles the realization of a low-order autoregressive process with low persistence where the mean may change over time. During the time period corresponding to the 2008-2009 financial crisis the series exhibits a large drop of several standard deviations. Applying Bai and Perron's (1998) testing procedure with 15% trimming at the beginning and the end of the sample period and between break dates except for the period 2008-2009 where a trimming of 2.5% is used, we find evidence of five breaks. The estimates of the break dates and of the β in each regime are reported in Table 4. The results from the testing procedure (not reported) suggest that all breaks are significant at 1% significant level except the first break for which the statistical evidence is weaker. Indeed, the sup-F test for no break versus two breaks in the sub-sample [1990:01, 2008:05] (i.e., [1, $\widehat{T}_3$] using the time index) rejects the null at 1% significant level. In this sub-sample the least-squares method first detects $\widehat{T}_2 =$ 2000:04 and then $\widehat{T}_1 =$ 1992:01. However, in the sub-sample [1990:01, 2000:04] (i.e., [1, $\widehat{T}_2$]) the evidence for an additional break is weak. Thus, while the break date estimates $\widehat{T}_2$, $\widehat{T}_3$, $\widehat{T}_4$, $\widehat{T}_5$ are associated to breaks of large magnitudes, the estimate of the first break date $\widehat{T}_1$ is associated to a small break. The first and the fourth regimes (resp., [1990:01, 1992:01] and [2008:06, 2009:03]) are short-lasting. Further, the first regime is close to the start of the sample. The first break suggests a higher growth rate from 1992 to early 2000 relative to the period 1990-1992. This period also corresponds to the well-known above-average growth in labor productivity in the late 1990s. The fourth regime lasts for ten months and corresponds to the 2008-2009 financial crisis. To detect this regime a trimming of 2.5% is needed.

We now move to the confidence sets about the break dates. Even though our theoretical results are established for a single break date, we can apply it to each individual estimated break date.[2] We consider the HDR method, Bai's (1997) and Elliott and Müller's (2007) method. We do not consider the ILR method since it is based on a testing procedure that detects a different number of breaks from the Bai and Perron's (1998) procedure and so this affects the number of confidence sets and their length making the comparison hard. For the Elliott and Müller's (2007) method we report only the length of the confidence set.[3] We begin with considering the confidence set for the first break date. The length of Bai's (1997) confidence set is shorter than that from the HDR confidence set. However, the tests for an additional break in [1990:01, 2000:04] provides weak evidence against the null of no break. In this case, a larger length of a confidence set should provide a better representation of the uncertainty about the break. Thus, the short length of Bai's (1997) method raises concerns that this confidence interval may underestimate the uncertainty about the break and following the simulation analysis of Section 7 it should leave us with concerns about its finite-sample coverage properties. The HDR method yields a larger confidence set and in addition we note that the start date of the confidence set corresponds to the start date of the sample. This is consistent with the weak evidence about this break.

The HDR method results in shorter confidence sets than Bai's (1997) for the remaining break dates except for the fourth break where the two methods yields the same length. However, for the latter case Bai's (1997) confidence set constructed using 5% trimming does not include the break date estimate obtained with 2.5% trimming (i.e., $\widehat{T}_4 =$ 2009:03) whereas the HDR confidence set does. This feature suggests that the HDR confidence set is less sensitive to the trimming choice and that it likely has better coverage properties. This is consistent with our theoretical results. The HDR confidence set for the second break date is less than a half that from Bai's. Elliott and Müller's (2007) confidence set is in general much larger than those from the other methods. This often holds even when the break magnitude is large. Thus, Elliott and Müller's (2007) confidence sets are often less informative than the other confidence sets, consistent with the simulation results discussed in Section 7. Overall, these results suggest that the improvement in inferences using the HDR method can be substantial.

[2]The HDR method and the corresponding theoretical results carry over to the multiple break case as long as the break fractions satisfy $\lambda_1^0 < \ldots < \lambda_m^0$ with $m > 1$ a finite integer, so that the break points $T_1^0, \ldots, T_m^0$ are asymptotically distinct. This assumption is ubiquitous in the literature.

[3]This method is based on test inversion and so one would need to report each date that enters the confidence set but this takes too much space in our table. They are available upon request.

9 Conclusions

We examined a change-point model under a continuous record asymptotics. With the time horizon $[0,\ N]$ fixed, we can account for the asymmetric informational content provided by the pre- and post-break samples. We derived a feasible counterpart of the continuous record asymptotic distribution of the change-point estimator using consistent plug-in estimates and showed that it provides accurate approximations to the finite-sample distributions. We used our limit theory to construct confidence sets for the change-point date based on the concept of Highest Density Region. Overall, it delivers accurate coverage probabilities and relatively short average lengths of the confidence sets. Importantly, it does so irrespective of the magnitude of the break, whether large or small, a notoriously difficult problem in the literature.

References

Andersen, T.G., Varneskov, R.T., 2021. Testing for parameter instability and structural change in persistent predictive regressions. Journal of Econometrics, forthcoming.

Andersen, T.G., Varneskov, R.T., 2022. Consistent local spectrum (LCM) inference for predictive return regressions. Econometric Theory, forthcoming.

Andrews, D.W.K., 1993. Tests for parameter instability and structural change with unknown change-point. Econometrica 61, 821–56.

Aue, A., Hórvath, L., 2013. Structural breaks in time series. Journal of Time Series Analysis 34, 1–16.

Bai, J., 1994. Least squares estimation of a shift in linear processes. Journal of Time Series Analysis 15, 453–472.

Bai, J., 1997. Estimation of a change-point in multiple regression models. The Review of Economics and Statistics 79, 551–563.

Bai, J., Perron, P., 1998. Estimating and testing linear models with multiple structural changes. Econometrica 66, 47–78.

Bai, J., Perron, P., 2003. Computation and analysis of multiple structural changes. Journal of Applied Econometrics 18, 1–22.

Bai, J., Perron, P., 2006. Multiple structural change models: a simulation analysis, in: D. Corbea, S.D., Hansen, B.E. (Eds.), Econometric Theory and Practice: Frontiers of Analysis and Applied Research. Cambridge University Press, pp. 212–237.

Barndorff-Nielsen, O.E., Shephard, N., 2004. Econometric analysis of realised covariation: high frequency based covariance, regression and correlation in financial economics. Econometrica 72, 885–925.

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, 183–208.

Brown, L.D., Low, M.G., 1996. Asymptotic equivalence of nonparametric regression and white noise. Annals of Statistics 24, 2384–2398.

Carrasco, M., Chen, X., 2002. Mixing and moment properties of various GARCH and stochastic volatility models. Econometric Theory 18, 17–39.

Casini, A., 2018. Tests for forecast instability and forecast failure under a continuous record asymptotic framework. arXiv preprint arXiv:1803.10883.

Casini, A., 2022. Theory of evolutionary spectra for heteroskedasticity and autocorrelation robust inference in possibly misspecified and nonstationary models. Journal of Econometrics, forthcoming.

Casini, A., Deng, T., Perron, P., 2022. Theory of low frequency contamination from nonstationarity and misspecification: consequences for HAR inference. arXiv preprint arXiv:2103.01604.

Casini, A., Perron, P., 2019. Structural breaks in time series. Oxford Research Encyclopedia of Economics and Finance, Oxford University Press.

Casini, A., Perron, P., 2020a. Continuous record asymptotics for structural change models. Extended working paper, arXiv preprint arXiv:1803.10881.

Casini, A., Perron, P., 2020b. Generalized Laplace inference in multiple change-points models. Econometric Theory 38, 35–65.

Casini, A., Perron, P., 2021a. Change-point analysis of time series with evolutionary spectra. arXiv preprint arXiv 2106.02031.

Casini, A., Perron, P., 2021b. Continuous record Laplace-based inference about the break date in structural change models. Journal of Econometrics 224, 3–21.

Casini, A., Perron, P., 2021c. Supplement to "Continuous record asymptotics for change-point models". Unpublished Manuscript, Department of Economics and Finance, University of Rome Tor Vergata.

Chambers, M.J., Taylor, A.M.R., 2019. Deterministic parameter change models in continuous and discrete time. Journal of Time Series Analysis 41, 134–145.

Chang, S.Y., Perron, P., 2018. A comparison of alternative methods to construct confidence intervals for the estimate of a break date in linear regression models. Econometric Reviews 37, 577–601.

Christensen, K., Oomen, R., Renò, R., 2022. The drift burst hypothesis. Journal of Econometrics 227, 461–497.

Deng, A., Perron, P., 2006. A comparison of alternative asymptotic frameworks to analyse a structural change in a linear time trend. Econometrics Journal 9, 423–447.

Elliott, G., Müller, U.K., 2007. Confidence sets for the date of a single break in linear time series regressions. Journal of Econometrics 141, 1196–1218.

Elliott, G., Müller, U.K., Watson, M.W., 2015. Nearly optimal tests when a nuisance parameter is present under the null hypothesis. Econometrica 83, 771–811.

Eo, Y., Morley, J., 2015. Likelihood-ratio-based confidence sets for the timing of structural breaks. Quantitative Economics 6, 463–497.

Foster, D.P., Nelson, D.B., 1996. Continuous record asymptotics for rolling sample variance estimators. Econometrica 62, 1–41.

Fryzlewicz, P., 2014. Wild binary segmentation for multiple change-point detection. Annals of Statistics 42, 2243–2281.

Genon-Catalot, V., Jeantheau, T., Laredo, C., 2000. Stochastic volatility models as hidden Markov models and statistical applications. Bernoulli 6, 1051–1079.

Georgiev, I., Harvey, D.I., Leybourne, S.J., Taylor, A.M.R., 2020. Testing for parameter instability in predictive regression models. Journal of Econometrics 204, 101–118.

Hansen, B., 2000. Sample splitting and threshold estimation. Econometrica 68, 575–603.

Hidalgo, J., Lee, J., Seo, M.H., 2019. Robust inference for threshold regression models. Journal of Econometrics 210, 291–309.

Hinkley, D.V., 1971. Inference about the change-point from cumulative sum tests. Biometrika 58, 509–523.

Hyndman, R.J., 1996. Computing and graphing highest density regions. The American Statistician 50, 120–126.

Jacod, J., 1994. Limit of random measures associated with the increments of a Brownian semimartingale. Discussion Paper, Université de Paris VI .

Jacod, J., 1997. On continuous conditional Gaussian martingales and stable convergence in law, in: Seminaire de Probabilitiés de Strasbourg, pp. 232–246.

Jacod, J., Protter, P., 1998. Asymptotic error distributions for the Euler method for stochastic differential equations. Annals of Probability 26, 267–307.

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

Lee, S., Seo, M., Shin, Y., 2016. The LASSO for high dimensional regression with a possible change-point. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 78, 193–210.

Leonardi, F., Bühlmann, P., 2016. Computationally efficient change-point detection for high-dimensional regression. arXiv preprint arXiv:1601.03704 .

Li, J., Todorov, V., Tauchen, G., 2017. Adaptive estimation of continuous-time regression models using high-frequency data. Journal of Econometrics 200, 36–47.

Mason, D.M., Polonik, W., 2008. Asymptotic normality of plug-in level set estimates. Extended version.

Mason, D.M., Polonik, W., 2009. Asymptotic normality of plug-in level set stimates. Annals of Applied Probability 19, 1108–1142.

Nelson, D.B., Foster, D.P., 1994. Asymptotic filtering theory for univariate ARCH models. Econometrica 62, 1–41.

Newey, W.K., West, K.D., 1987. A simple positive semidefinite, heteroskedastic and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.

Nyblom, J., 1989. Testing for the constancy of parameters over time. Journal of the American Statistical Association 89, 223–230.

Perron, P., 1991. A continuous time approximation to the unstable first-order autoregressive model: the case without an intercept. Econometrica 59, 211–236.

Perron, P., 2006. Dealing with structural breaks, in: Patterson, K., Mills, T. (Eds.), Palgrave Handbook of Econometrics. Palgrave Macmillan. volume 1: Econometric Theory, pp. 278–352.

Perron, P., Zhu, X., 2005. Structural breaks with deterministic and stochastic trends. Journal of Econometrics 129, 65–119.

Phillips, P.C.B., 1987. Time series regression with a unit root. Econometrica 55, 277–301.

Picard, D., 1985. Testing and estimating change-points in time series. Advances in Applied Probability 17, 841–867.

Samworth, R.J., Wand, M.P., 2010. Asymptotics and optimal bandwidth selection for highest density region estimation. Annals of Statistics 38, 1767–1792.

Wang, D., Lin, K., Willett, R., 2019a. Statistically and computationally efficient change-point localization in regression settings. Journal of Machine Learning Research, forthcoming.

Wang, D., Yu, Y., Rinaldo, A., Willett, R., 2019b. Localizing changes in high-dimensional vector autoregressive processes. arXiv preprint arXiv:1909.06359.

Yao, Y., 1987. Approximating the distribution of the ML estimate of the change-point in a sequence of independent random variables. Annals of Statistics 15, 1321–1328.

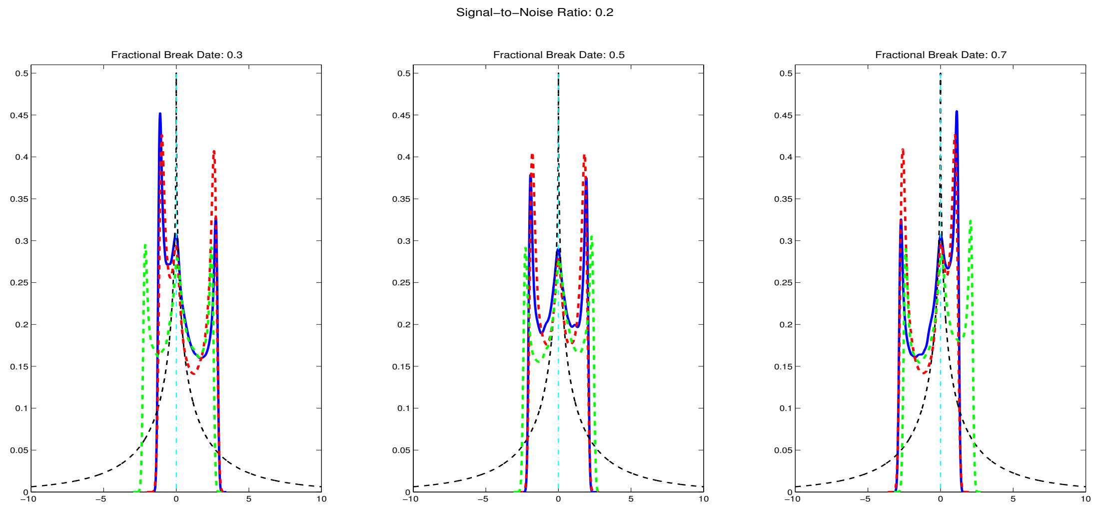

Figure 1: The probability density of $\rho(\widehat{T}_{b,\pi} - T_b^0)$ for model (5.1) with break magnitude $\delta_Z^0 = 0.2$ and true break fraction $\lambda_0 = 0.3$, 0.5 and 0.7 (the left, middle and right panel, respectively). The signal-to-noise ratio is $\delta_Z^0/\sigma_e = \delta_Z^0$ since $\sigma_e^2 = 1$. The blue solid (green broken) line is the density of the infeasible (reps. feasible) asymptotic distribution derived under a continuous record, the black broken line is the density of the asymptotic distribution from Bai (1997) and the red broken line is the density of the finite-sample distribution.

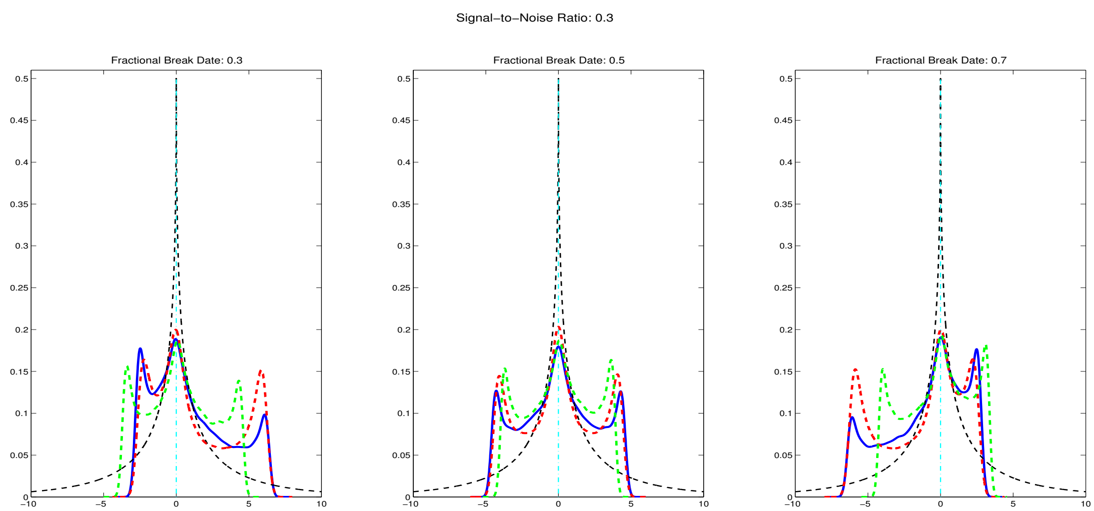

Figure 2: The probability density of $\rho(\widehat{T}_{b,\pi} - T_b^0)$ for model (5.1) with break magnitude $\delta_Z^0 = 0.3$ and true break fraction $\lambda_0 = 0.3$, 0.5 and 0.7 (the left, middle and right panel, respectively). The signal-to-noise ratio is $\delta_Z^0/\sigma_e = \delta_Z^0$ since $\sigma_e^2 = 1$. The blue solid (green broken) line is the density of the infeasible (reps. feasible) asymptotic distribution derived under a continuous record, the black broken line is the density of the asymptotic distribution from Bai (1997) and the red broken line is the density of the finite-sample distribution.

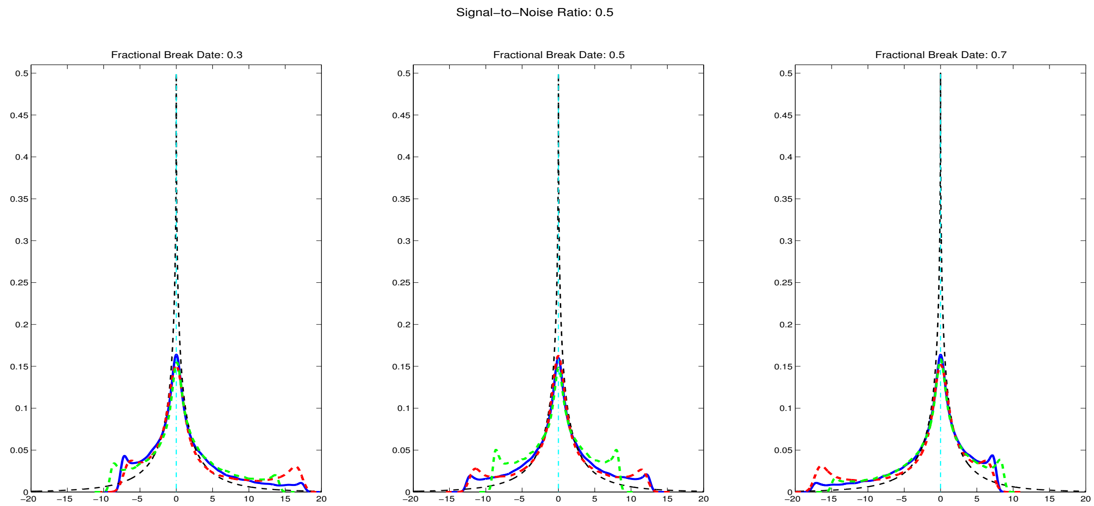

Figure 3: The probability density of $\rho(\widehat{T}_{b,\pi} - T_b^0)$ for model (5.1) with break magnitude $\delta_Z^0 = 0.5$ and true break fraction $\lambda_0 = 0.3$, 0.5 and 0.7 (the left, middle and right panel, respectively). The signal-to-noise ratio is $\delta_Z^0/\sigma_e = \delta_Z^0$ since $\sigma_e^2 = 1$. The blue solid (green broken) line is the density of the infeasible (reps. feasible) asymptotic distribution derived under a continuous record, the black broken line is the density of the asymptotic distribution from Bai (1997) and the red broken line is the density of the finite-sample distribution.

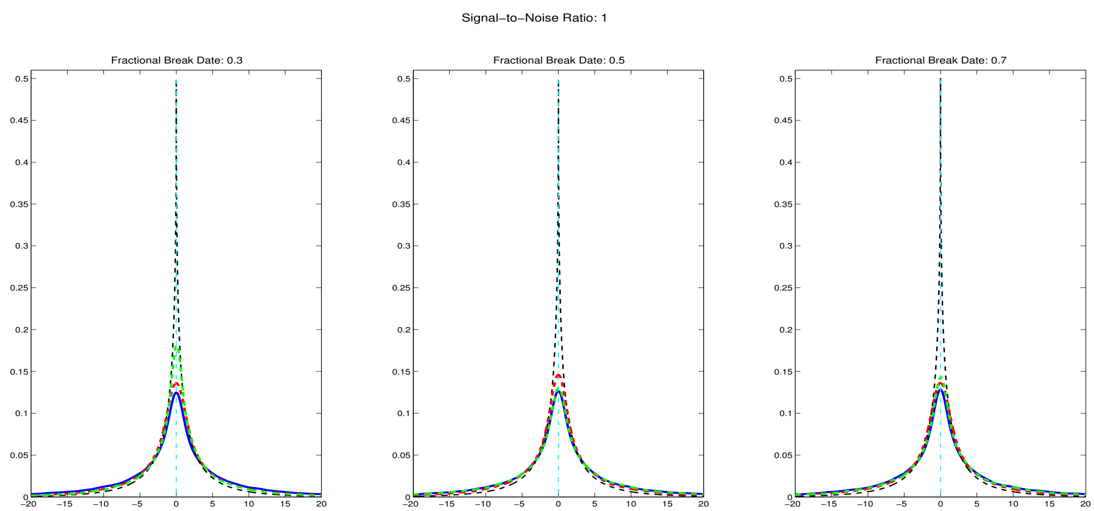

Figure 4: The probability density of $\rho(\widehat{T}_{b,\pi} - T_b^0)$ for model (5.1) with break magnitude $\delta_Z^0 = 1$ and true break fraction $\lambda_0 = 0.3$, 0.5 and 0.7 (the left, middle and right panel, respectively). The signal-to-noise ratio is $\delta_Z^0/\sigma_e = \delta_Z^0$ since $\sigma_e^2 = 1$. The blue solid (green broken) line is the density of the infeasible (reps. feasible) asymptotic distribution derived under a continuous record, the black broken line is the density of the asymptotic distribution from Bai (1997) and the red broken line is the density of the finite-sample distribution.

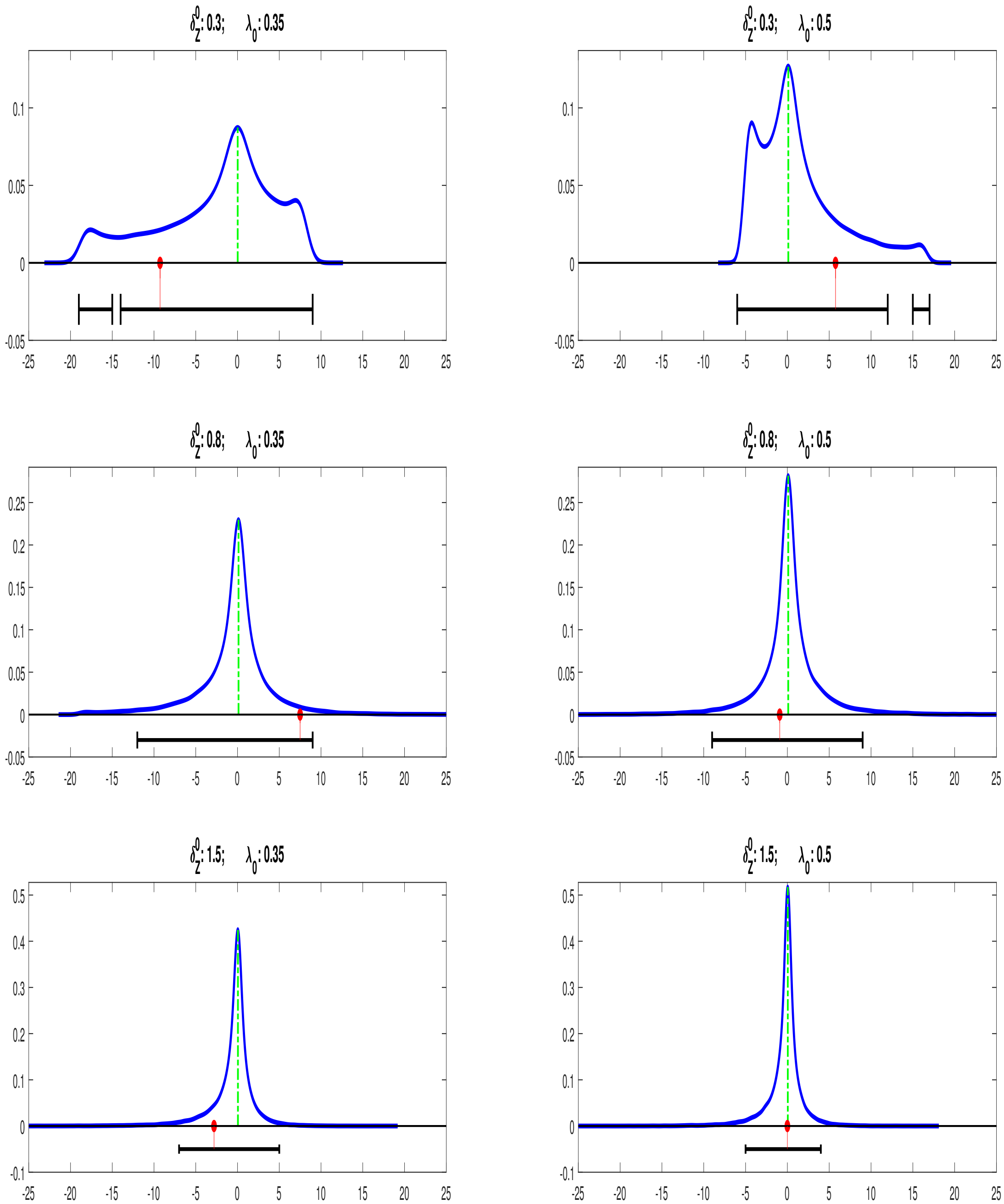

Figure 5: Highest Density Regions (HDRs) of the feasible probability density of $\rho(\widehat{T}_{b,\pi} - T_b^0)$ as described in Section 6. The significance level is $\alpha = 0.05$, the true break point is $\lambda_0 = 0.3$ and 0.5 (the left and right panels, respectively) and the break magnitude is $\delta_Z^0 = 0.3$, 0.8 and 1.5 (the top, middle and bottom panels, respectively). The horizontal axis is the support of $\rho\left(\widehat{T}_{b,\pi} - T_b^0\right)$. The red dot is the true value of the break point. The union of the black lines below the horizontal axis is the 95% HDR confidence region.

Figure 6: U.S. Industrial Production growth and estimates of the break dates. The sample period is 1990:01–2020:01.

Table 1: Coverage rate and length of the confidence set for the example of Section 6

	$\delta_Z^0 = 0.3$		$\delta_Z^0 = 0.8$		$\delta_Z^0 = 1.5$	
	Cov.	Lgth.	Cov.	Lgth.	Cov.	Lgth.
$\lambda_0 = 0.35$						
HDR	1	94	1	27	1	10
Bai (1997)	0	55	0	13	1	8
$\widehat{U}_T$.neq	1	95	1	37	1	24
$\lambda_0 = 0.5$						
HDR	1	82	1	14	1	4
Bai (1997)	1	67	1	18	1	5
$\widehat{U}_T$.neq	1	95	1	35	1	14

Coverage rate and length of the confidence sets corresponding to the example from Section 6. See also Figure 5. The significance level is $\alpha = 0.05$. Cov. and Lgth. refer to the coverage rate and average size of the confidence sets (i.e. average number of dates in the confidence sets), respectively. Cov=1 if the confidence set includes T_b^0 and Cov=0 otherwise. The sample size is $T = 100$.

Table 2: Small-sample coverage rate and length of the confidence set for model M1

		$\delta_Z^0 = 0.3$		$\delta_Z^0 = 0.6$		$\delta_Z^0 = 1$		$\delta_Z^0 = 2$	
		Cov.	Lgth.	Cov.	Lgth.	Cov.	Lgth.	Cov.	Lgth.
$\lambda_0 = 0.5$	HDR	0.938	131.35	0.941	69.05	0.943	24.02	0.962	6.89
	Bai (1997)	0.842	114.24	0.855	51.58	0.911	19.75	0.964	5.70
	$\widehat{U}_T$.eq	0.946	146.23	0.943	76.13	0.948	33.45	0.930	14.59
	ILR	0.954	147.25	0.956	78.17	0.965	23.87	0.973	5.25
$\lambda_0 = 0.35$	HDR	0.939	129.02	0.934	63.70	0.939	24.23	0.951	5.78
	Bai (1997)	0.855	111.45	0.855	49.52	0.914	19.39	0.956	5.62
	$\widehat{U}_T$.eq	0.933	148.74	0.933	75.94	0.933	33.08	0.933	14.43
	ILR	0.946	149.81	0.960	77.54	0.964	25.63	0.982	5.42
$\lambda_0 = 0.2$	HDR	0.941	127.29	0.940	62.13	0.942	22.06	0.946	5.73
	Bai (1997)	0.863	110.12	0.911	53.14	0.931	20.20	0.967	5.67
	$\widehat{U}_T$.eq	0.950	158.98	0.951	97.12	0.950	35.26	0.950	13.99
	ILR	0.956	162.32	0.956	96.45	0.965	33.31	0.976	5.96

The model is $y_t = \beta^0 + \delta_Z^0 \mathbf{1}_{\{t > \lfloor T\lambda_0 \rfloor\}} + e_t$, $e_t \sim i.i.d.\,\mathscr{N}(0, 1)$, $T = 200$. 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% size with the asymptotic critical value. The number of simulations is 5,000.

Table 3: Small-sample coverage rate and length of the confidence sets for model M2

		$\delta_Z^0 = 1$		$\delta_Z^0 = 1.5$		$\delta_Z^0 = 2$		$\delta_Z^0 = 3$	
		Cov.	Lgth.	Cov.	Lgth.	Cov.	Lgth.	Cov.	Lgth.
$\lambda_0 = 0.5$	HDR	0.916	30.68	0.944	14.77	0.969	8.34	0.995	4.55
	Bai (1997)	0.793	12.87	0.877	7.11	0.929	4.78	0.973	2.957
	$\widehat{U}_T$.eq	0.951	91.64	0.955	93.94	0.959	93.71	0.961	90.34
	ILR	0.951	46.31	0.967	34.19	0.977	26.48	0.991	16.49
$\lambda_0 = 0.35$	HDR	0.925	33.02	0.933	16.67	0.971	9.40	0.994	4.33
	Bai (1997)	0.804	13.00	0.876	7.11	0.923	4.94	0.974	2.93
	$\widehat{U}_T$.eq	0.952	91.22	0.945	92.61	0.957	92.48	0.964	93.08
	ILR	0.949	47.54	0.967	34.18	0.982	25.84	0.984	16.76
$\lambda_0 = 0.2$	HDR	0.937	34.66	0.953	19.24	0.954	11.42	0.994	5.36
	Bai (1997)	0.832	13.64	0.885	7.19	0.931	4.92	0.971	2.91
	$\widehat{U}_T$.eq	0.944	89.64	0.951	89.58	0.956	88.22	0.961	85.95
	ILR	0.946	49.13	0.970	33.54	0.980	24.48	0.989	12.51

The model is $y_t = \delta_Z^0 \left(1 - \nu^0\right) \mathbf{1}_{\{t > \lfloor T\lambda_0 \rfloor\}} + \nu^0 y_{t-1} + e_t$, $e_t \sim i.i.d.\,\mathscr{N}(0, 0.04)$, $\nu^0 = 0.8$, $T = 100$. The notes of Table 2 apply.

Table 4: Break dates in monthly U.S. industrial production: 1990:01–2020:01

Regime	Break date $\widehat{T}_b$	$\widehat{\beta}$	SD($\widehat{\beta}$)
[1990:01, 1992:01]	1992:1	-0.08	2.24
[1992:02, 2000:04]	2000:04	4.97	0.46
[2000:05, 2008:05]	2008:05	0.82	0.60
[2008:06, 2009:03]	2009:03	-18.63	4.74
[2009:04, 2015:12]	2015:12	4.94	1.03
[2016:01, 2020:01]	2020:01	-0.33	0.76

	Confidence set	Lgth.
T_1^0		
Bai (1997)	[1990:10, 1992:12]	25
HDR	[1990:01, 1992:10]	32
$\widehat{U}_T$.eq		55
T_2^0		
Bai (1997)	[1998:09, 2001:06]	33
HDR	[1999:11, 2001:2]	14
$\widehat{U}_T$.eq		48
T_3^0		
Bai (1997)	[2008:01, 2008:10]	9
HDR	[2008:05, 2008:12]	7
$\widehat{U}_T$.eq		7
T_4^0		
Bai (1997)	[2009:05, 2009:11]	6
HDR	[2008:12, 2009:06]	6
$\widehat{U}_T$.eq		10
T_5^0		
Bai (1997)	[2013:03, 2016:03]	37
HDR	[2013:12, 2016:03]	28
$\widehat{U}_T$.eq		47

Regime indicates the start and end date of the regime. Break date $\widehat{T}_b$ indicates the least-squares estimate of the break date where $b = 1, 2, 3, 4, 5$. $\widehat{\beta}$ is the estimate of β in a given regime. SD($\widehat{\beta}$) is the standard deviation of $\widehat{\beta}$.

A Supplemental Materials

The supplement for online publication [cf. Casini and Perron (2021c)] includes the followings: (i) it describes how to simulate the continuous record limiting distribution; (ii) it describes the limiting process in Theorem 4.1; (iii) it extends the benchmark model in Section 2 to include predictable processes; (iv) it includes all proofs of the results in the paper; (v) it presents additional small-sample evaluations of the HDR confidence sets.

Supplemental Material to

Continuous Record Asymptotics for Change-Point Models

Alessandro Casini Pierre Perron
University of Rome Tor Vergata Boston University

23rd October 2022
First Vesrion: 28th October 2015

Abstract

This supplemental material is structured as follows. Section S.A describes how to simulate the continuous record limiting distribution. Section S.B describes the limiting process in Theorem 4.1 of the main text. Section S.C extends the benchmark model in Section 2 of the main text to include predictable processes. Section S.D includes all proofs of the results in the paper. Section S.E presents additional small-sample evaluations of the HDR confidence sets.

S.A Simulation of the Limiting Distribution in Theorem 4.1

We discuss how to simulate the limiting distribution in Theorem 4.1 which is slightly different from simulating the limiting distribution in Theorem 4.2. However, the idea is similar in that we replace unknown quantities by consistent estimates. First, we replace N_b^0 by $\widehat{N}_b = \widehat{T}_b/T$. The ratio $||\delta^0||^2/\overline{\sigma}^2$ is consistently estimated by $||\widehat{\delta}||^2/(T^{-1}\sum_{k=1}^T \widehat{e}_{kh}^2)$ because under the "fast time scale" $h^{1/2}\sum_{k=1}^T \widehat{e}_{kh}^2 \xrightarrow{p} \overline{\sigma}^2$ (cf. Assumption 4.1). Now consider the term $\{-(\delta^0)' \langle Z_\Delta,\, Z_\Delta\rangle (v)\, \delta^0 + 2(\delta^0)'\mathscr{W}(v)\}$. For $v \leq 0$, this can be consistently estimated by

$$-T^{1/2}\left[\left(\widehat{\delta}\right)'\left(\sum\nolimits_{k=\widehat{T}_b+\lfloor v/h\rfloor}^{\widehat{T}_b} z_{kh}z_{kh}'\right)\widehat{\delta} - 2\widehat{\delta}'\widehat{\mathscr{W}_h}(v)\right], \tag{S.1}$$

where $\widehat{\mathscr{W}_h}$ is a simple-size dependent sequence of Gaussian processes whose marginal distribution is characterized by $h^{1/2}T\sum_{k=\widehat{T}_b+\lfloor v/h\rfloor}^{\widehat{T}_b} \widehat{e}_{kh}^2 z_{kh}z_{kh}'$ which is a consistent estimate of $\int_v^0 \Omega_{Ze,s}ds$. Thus, in the limit $\widehat{\mathscr{W}_h}(v)$ has the same marginal distribution as $\mathscr{W}(v)$, and it follows that the limiting distribution from Theorem 4.1 can be simulated. The proposed method is valid under a continuous-record asymptotic (i.e., under Assumption 4.1 and the adoption of the "fast time scale"). It can also be shown to be valid under a fixed-shifts framework.

S.B Description of the Limiting Process in Theorem 4.1

We describe the probability setup underlying the limit process of Theorem 4.1. Note that $Z_\Delta' e/h^{1/2} = h^{-1/2}\sum_{k=T_b+1}^{T_b^0} z_{kh}e_{kh}$ if $T_b \leq T_b^0$. Consider an additional measurable space $(\Omega^*,\, \mathscr{F}^*)$ and a transition probability $H(\omega,\, d\omega^*)$ from $(\Omega,\, \mathscr{F})$ into $(\Omega^*,\, \mathscr{F}^*)$. Next, we can define the products $\widetilde{\Omega} = \Omega \times \Omega^*$, $\widetilde{\mathscr{F}} = \mathscr{F} \otimes \mathscr{F}^*$, $\widetilde{P}(d\omega,\, d\omega^*) = P(d\omega)\, H(\omega,\, d\omega^*)$. This defines an extension $(\widetilde{\Omega},\, \widetilde{\mathscr{F}},\, \widetilde{P})$ of the original space $(\Omega,\, \mathscr{F},\, \{\mathscr{F}_t\}_{t\geq 0},\, P)$. We also consider another filtration $\{\widetilde{\mathscr{F}_t}\}_{t\geq 0}$ which takes the following product form $\widetilde{\mathscr{F}_t} = \cap_{s>t}\mathscr{F}_s \otimes \mathscr{F}_s^*$ where $\{\mathscr{F}_t^*\}_{t\geq 0}$ is a filtration on $(\Omega^*,\, \mathscr{F}^*)$. For the transition probability H, we consider the simple form $H(\omega,\, d\omega^*) = P^*(d\omega^*)$ for some probability measure P^* on $(\Omega^*,\, \mathscr{F}^*)$. This constitutes a "very good" product filtered extension. Next, assume that $(\Omega^*,\, \mathscr{F}^*,\, (\mathscr{F}_t^*)_{t\geq 0},\, P^*)$ supports p-dimensional $\{\mathscr{F}_t^*\}$-standard independent Wiener processes $W^{i*}(v)$ $(i = 1,\, 2)$. Finally, we postulate the process $\Omega_{Ze,t}$ with entries $\Sigma_Z^{(i,j)}\sigma_e^2$ to admit a progressively measurable $p \times p$ matrix-valued process (i.e., a symmetric "square-root" process) σ_{Ze}, satisfying $\Omega_{Ze} = \sigma_{Ze}\sigma_{Ze}'$, with the property that $||\sigma_{Ze}||^2 \leq K||\Omega_{Ze}||$ for some $K < \infty$. Define the process $\mathscr{W}(v) = \mathscr{W}_1(v)$ if $v \leq 0$, and $\mathscr{W}(v) = \mathscr{W}_2(v)$ if $v > 0$, where $\mathscr{W}_1(v) = \int_{N_b^0+v}^{N_b^0}\sigma_{Ze,s}dW_s^{1*}$ and $\mathscr{W}_2(v) = \int_{N_b^0}^{N_b^0+v}\sigma_{Ze,s}dW_s^{2*}$ with components $\mathscr{W}^{(j)}(v) = \sum_{r=1}^p\int_{N_b^0+v}^{N_b^0}\sigma_{Ze,s}^{(jr)}dW_s^{1*(r)}$ if $v \leq 0$ and $\mathscr{W}^{(j)}(v) = \sum_{r=1}^p\int_{N_b^0}^{N_b^0+v}\sigma_{Ze,s}^{(jr)}dW_s^{2*(r)}$ if $v > 0$. The process $\mathscr{W}(v)$ is well defined on the product extension $(\widetilde{\Omega},\, \widetilde{\mathscr{F}},\, \left\{\widetilde{\mathscr{F}_t}\right\}_{t\geq 0},\, \widetilde{P})$, and furthermore, conditionally on $\mathscr{F}$, is a two-sided centered continuous Gaussian process with independent increments and (conditional) covariance

$$\widetilde{\mathbb{E}}\left(\mathscr{W}^{(u)}(v)\,\mathscr{W}^{(j)}(v)\right) = \Omega_{\mathscr{W}}^{(u,j)}(v) = \begin{cases}\Omega_{\mathscr{W},1}^{(u,j)}(v), & \text{if } v \leq 0\\ \Omega_{\mathscr{W},2}^{(u,j)}(v), & \text{if } v > 0\end{cases}, \tag{S.1}$$

where $\Omega^{(u,j)}_{\mathscr{W},1}(v) = \int_{N_b^0+v}^{N_b^0} \Omega^{(u,j)}_{Ze,s} ds$ and $\Omega^{(u,j)}_{\mathscr{W},2}(v) = \int_{N_b^0}^{N_b^0+v} \Omega^{(u,j)}_{Ze,s} ds$. Therefore, $\mathscr{W}(v)$ is conditionally on $\mathscr{F}$, a continuous martingale with "deterministic" quadratic covariation process $\Omega_{\mathscr{W}}$. The continuity of $\Omega_{\mathscr{W}}$ signifies that $\mathscr{W}(v)$ is not only conditionally Gaussian but also a.s. continuous. Precise treatment of this result can be found in Section II.7 of Jacod and Shiryaev (2003).

S.C The Extended Model with Predictable Processes

S.C.1 The Extended Model

The assumptions on D_t and Z_t specify that they are continuous semimartingale of the form (2.3). This precludes predictable processes, which are often of interest in applications; e.g., a constant and/or a lagged dependent variable. Technically, these require a separate treatment since the coefficients associated with predictable processes are not identified under a fixed-span asymptotic setting. Let

$$\tau_{1,k} = \mu_{1,h} h + \alpha_{1,h} Y_{(k-1)h}, \qquad (k \leq \lfloor T\lambda_0 \rfloor),$$

and

$$\tau_{2,k} = \mu_{2,h} h + \alpha_{2,h} Y_{(k-1)h}, \qquad (k > \lfloor T\lambda_0 \rfloor + 1).$$

We consider the following extended model:

$$\Delta_h Y_k \triangleq \begin{cases} \tau_{1,k} + (\Delta_h D_k)' \nu^0 + (\Delta_h Z_k)' \delta^0_{Z,1} + \Delta_h e^*_k, & (k = 1, \ldots, T^0_b) \\ \tau_{2,k} + (\Delta_h D_k)' \nu^0 + (\Delta_h Z_k)' \delta^0_{Z,2} + \Delta_h e^*_k, & (k = T^0_b + 1, \ldots, T) \end{cases} \tag{S.1}$$

for some given initial value Y_0. We specify the parameters associated with the constant and the lagged dependent variable as being of higher order in h, or lower in T, as $h \downarrow 0$ so that some fixed true parameter values can be identified, i.e., $\mu_{1,h} \triangleq \mu^0_1 h^{-1/2}$, $\mu_{2,h} \triangleq \mu^0_2 h^{-1/2}$, $\mu_{\delta,h} \triangleq \mu_{2,h} - \mu_{1,h}$, $\alpha_{1,h} \triangleq \alpha^0_1 h^{-1/2}$, $\alpha_{2,h} \triangleq \alpha^0_2 h^{-1/2}$ and $\alpha_{\delta,h} \triangleq \alpha_{2,h} - \alpha_{1,h}$. Our framework is then similar to the small-diffusion setting studied previously [cf. Ibragimov and Has'minskiĭ (1980), Galtchouk and Konev (2001), Laredo (1990) and Sørensen and Uchida (2003)]. With $\mu_{\cdot,h}$ and $\alpha_{\cdot,h}$ independent of h and fixed, respectively, at the true values $\mu^0_\cdot$ and $\alpha^0_\cdot$, the continuous-time model is then equivalent to

$$\begin{aligned} Y_t = Y_0 &+ \int_0^t \left(\mu^0_1 + \mu^0_\delta \mathbf{1}_{\{s > N^0_b\}} \right) ds + \int_0^t \left(\alpha^0_1 + \alpha^0_\delta \mathbf{1}_{\{s > N^0_b\}} \right) Y_s ds \\ &+ D'_t \nu^0 + \int_0^t \left(\delta^0_{Z,1} + \delta^0_Z \mathbf{1}_{\{s > N^0_b\}} \right)' dZ_s + e^*_t, \end{aligned} \tag{S.2}$$

for $t \in [0, N]$, where $Y_t = \sum_{k=1}^{\lfloor t/h \rfloor} \Delta_h Y_k$, $D_t = \sum_{k=1}^{\lfloor t/h \rfloor} \Delta_h D_k$, $Z_t = \sum_{k=1}^{\lfloor t/h \rfloor} \Delta_h Z_k$ and $e^*_t = \sum_{k=1}^{\lfloor t/h \rfloor} \Delta_h e^*_k$. The results to be discussed below go through in this extended framework. However, some additional technical details are needed. Hence, we treat both cases with and without predictable components separately. Note that the model and results can be trivially extended to allow for more general forms of predictable processes (e.g., more lagged values), at the expense of additional technical details of no substance.

S.C.2 Asymptotic Results for the Model with Predictable Processes

In this section, we present asymptotic results allowing for predictable processes that include a constant and a lagged dependent variable among the regressors. Recall model (S.1). Let $\beta^0 = (\mu_1^0,\, \alpha_1^0,\, (\nu^0)',\, (\delta_{Z,1}^0)')'$, $\delta_Z^0 = (\mu_\delta^0,\, \alpha_\delta^0,\, (\delta_{Z,2}^0 - \delta_{Z,1}^0)')'$, $((\beta^0)',\, ((\delta_Z^0)'))' \in \Theta_0$, and $x_{kh} = ((\mu_{1,h}/\mu_1^0)h,\, (\alpha_{1,h}/\alpha_1^0)Y_{(k-1)h}h,\, \Delta_h D_k',\, \Delta_h Z_k')$. In matrix format, the model is $Y = X\beta^0 + Z_0\delta_Z^0 + e$, where now X is $T\times(p+q+2)$ and $Z_0 = X\overline{R}$, $\overline{R} \triangleq [(I_2,\, 0_{2\times p})',\, (0'_{(p+q)\times 2},\, R)]'$, with R as defined in Section 2. Natural estimates of β^0 and δ_Z^0 minimize the following criterion function,

$$\begin{aligned} h^{-1}\sum_{k=1}^{T}&\left(\Delta_h Y_k - \beta' \int_{(k-1)h}^{kh} X_s ds - \delta' \int_{(k-1)h}^{kh} Z_s ds\right)^2 \\ &= h^{-1}\sum_{k=1}^{T}\Bigg(\Delta_h Y_k - \mu_1^h h - \alpha_1^h \int_{(k-1)h}^{kh} Y_s ds - \pi' \Delta_h D_k \\ &\qquad - \delta_{Z,1}'\Delta_h Z_k \mathbf{1}\left\{k \le T_b\right\} - \delta_{Z,2}'\Delta_h Z_k \mathbf{1}\left\{k > T_b\right\}\Bigg)^2. \end{aligned} \tag{S.3}$$

Hence, we define our LS estimator as the minimizer of the following approximation to (S.3):

$$\begin{aligned} h^{-1}\sum_{k=1}^{T}\Bigg(&\Delta_h Y_k - \mu_1^h h - \alpha_1^h Y_{(k-1)h} h - \nu' \Delta_h D_k \\ &- \delta_{Z,1}'\Delta_h Z_k \mathbf{1}\left\{k \le T_b\right\} - \delta_{Z,2}'\Delta_h Z_k \mathbf{1}\left\{k > T_b\right\}\Bigg)^2. \end{aligned}$$

Such approximations are common [cf. Christopeit (1986), Lai and Wei (1983) and Mel'nikov and Novikov (1988) and the more recent work of Galtchouk and Konev (2001)]. Define $\Delta_h \widetilde{Y}_k \triangleq h^{1/2}\Delta_h Y_k$ and $\Delta_h \widetilde{V}_k = h^{1/2}\Delta_h V_k(\nu^0,\, \delta_{Z,1}^0,\, \delta_{Z,2}^0)$, where

$$\Delta_h V_k\left(\nu^0,\, \delta_{Z,1}^0,\, \delta_{Z,2}^0\right) \triangleq \begin{cases} (\nu^0)'\,\Delta_h D_k + \left(\delta_{Z,1}^0\right)'\Delta_h Z_k + \Delta_h e_k^*, & \text{if } k \le T_b^0 \\ (\nu^0)'\,\Delta_h D_k + \left(\delta_{Z,2}^0\right)'\Delta_h Z_k + \Delta_h e_k^*, & \text{if } k > T_b^0 \end{cases}.$$

The small-dispersion format of our model is then

$$\begin{aligned} \Delta_h \widetilde{Y}_k &= \left(\mu_1^0 h + \alpha_1^0 \widetilde{Y}_{(k-1)h} h\right)\mathbf{1}\left\{k \le T_b^0\right\} \\ &\quad + \left(\mu_2^0 h + \alpha_2^0 \widetilde{Y}_{(k-1)h} h\right)\mathbf{1}\left\{k > T_b^0\right\} + \Delta_h \widetilde{V}_k\left(\nu^0,\, \delta_{Z,1}^0,\, \delta_{Z,2}^0\right). \end{aligned} \tag{S.4}$$

This re-parametrization emphasizes that asymptotically our model describes small disturbances to the approximate dynamical system

$$d\widetilde{Y}_t^0/dt = \left(\mu_1^0 + \alpha_1^0 \widetilde{Y}_t^0\right)\mathbf{1}\left\{t \le N_b^0\right\} + \left(\mu_2^0 + \alpha_2^0 \widetilde{Y}_t^0\right)\mathbf{1}\left\{t > N_b^0\right\}. \tag{S.5}$$

The process $\left\{\widetilde{Y}_t^0\right\}_{t\geq 0}$ is the solution to the underlying ordinary differential equation. The LS estimate of the break point is then defined as $\widehat{T}_b \triangleq \arg\max_{T_b} Q_T(T_b)$, where

$$Q_T(T_b) \triangleq Q_T\left(\widehat{\beta}(T_b),\, \widehat{\delta}(T_b),\, T_b\right) = \widehat{\delta}'\left(Z_2' M Z_2\right)\widehat{\delta},$$

and the LS estimates of the regression parameters are

$$\widehat{\theta} \triangleq \arg\min_{\theta\in\Theta_0} h\left(S_T\left(\beta,\, \delta,\, \widehat{T}_b\right) - S_T\left(\beta^0,\, \delta_Z^0,\, T_b^0\right)\right),$$

where S_T is the sum of square residuals. With the exception of our small-dispersion assumption and consequent more lengthy derivations, our analysis remains the same as in the model without predictable processes. Hence, the asymptotic distribution of the break point estimator is derived under the same setting as in Section 4. We show that the limiting distribution is qualitatively equivalent to that in Theorem 4.1.

Assumption S.C.1. *Assumption 2.4 and 3.2 hold. Assumption 2.1 and 3.1 now apply to the last p (resp. q) elements of the process $\{Z_t\}_{t\geq 0}$ (resp. $\{D_t\}_{t\geq 0}$).*

Proposition S.C.1. *Consider model (S.1). Under Assumption 2.2 and S.C.1: (i) $\widehat{\lambda}_b \xrightarrow{P} \lambda_0$; (ii) for every $\varepsilon > 0$ there exists a $K > 0$ such that for all large T, $P(T|\widehat{\lambda}_b - \lambda_0| > K||\delta^0||^{-2}\overline{\sigma}^2) < \varepsilon$.*

Assumption S.C.2. *Let $\delta_h = h^{1/4}\delta^0$ and for $i = 1, 2$ $\mu_i^h = h^{1/4}\mu_i^0$ and $\alpha_i^h = h^{1/4}\alpha_i^0$, and assume that for all $t \in (N_b^0 - \epsilon,\, N_b^0 + \epsilon)$, with $\epsilon \downarrow 0$ and $T^{1-\kappa}\epsilon \to B < \infty$, $0 < \kappa < 1/2$, $\mathbb{E}[(\Delta_h e_t^*)^2 \,|\, \mathscr{F}_{t-h}] = \sigma_{h,t}^2 \Delta t$ P-a.s, where $\sigma_{h,t} \triangleq \sigma_h \sigma_{e,t}$ with $\sigma_h \triangleq h^{-1/4}\overline{\sigma}$.*

Furthermore, define the normalized residual $\Delta_h \widetilde{e}_t$ as in (4.1). We shall derive a stable convergence in distribution for $\overline{Q}_T(\cdot,\cdot)$ as defined in Section 4. The description of the limiting process is similar to the one presented in the previous section. However, here we shall condition on the σ-field $\mathscr{G}$ generated by all latent processes appearing in the model. In view of its properties, the σ-field $\mathscr{F}$ admits a regular version of the $\mathscr{G}$-conditional probability, denoted $H(\omega,\, d\omega^*)$. The limit process is then realized on the extension $(\widetilde{\Omega},\, \widetilde{\mathscr{F}},\, \{\widetilde{\mathscr{F}}_t\}_{t\geq 0},\, \widetilde{P})$ of the original filtered probability space as explained in Section S.B. We again introduce a two-sided Gaussian process $\mathscr{W}_{Ze}(\cdot)$ with a different dimension in order to accommodate for the presence of the predictable regressors in the first two columns of both X and Z. That is, $\mathscr{W}_{Ze}(\cdot)$ is a p-dimensional process which is $\mathscr{G}$-conditionally Gaussian and has P-a.s. continuous sample paths. We then have the following theorem.

Theorem S.C.1. *Consider model (S.4). Under Assumption 4.2, S.C.1-S.C.2: (i) $\widehat{\lambda}_b \xrightarrow{P} \lambda_0$; (ii) for every $\varepsilon > 0$ there exists a $K > 0$ such that for all large T, $P(T^{1-\kappa}|\widehat{\lambda}_b - \lambda_0| > K||\delta^0||^{-2}\overline{\sigma}^2) < \varepsilon$; (iii)*

$$N\left(\widehat{\lambda}_{b,\pi} - \lambda_0\right) \overset{\mathcal{L}-\mathrm{s}}{\Rightarrow} \underset{v\in\left[\frac{N\pi - N_b^0}{||\delta^0||^{-2}\overline{\sigma}^2},\, \frac{N(1-\pi)-N_b^0}{||\delta^0||^{-2}\overline{\sigma}^2}\right]}{\operatorname{argmax}} \left\{-\left(\delta^0\right)'\Lambda(v)\,\delta^0 + 2\left(\delta^0\right)'\mathscr{W}(v)\right\}, \tag{S.6}$$

where $\Lambda(v)$ is a process given by

$$\Lambda(v) \triangleq \begin{cases}\Lambda_1(v), & \text{if } v \leq 0\\ \Lambda_2(v), & \text{if } v > 0\end{cases}, \qquad \text{with}$$

$$\Lambda_1(v) \triangleq \begin{bmatrix} \int_{N_b^0+v}^{N_b^0} ds & \int_{N_b^0+v}^{N_b^0} \widetilde{Y}_s ds & 0_{1\times p} \\ \int_{N_b^0+v}^{N_b^0} \widetilde{Y}_s ds & \int_{N_b^0+v}^{N_b^0} \widetilde{Y}_s^2 ds & 0_{1\times p} \\ 0_{p\times 1} & 0_{p\times 1} & \langle Z, Z\rangle_1(v) \end{bmatrix},$$

and $\Lambda_2(v)$ is defined analogously, where $\langle Z, Z\rangle_1(v)$ is the $p\times p$ predictable quadratic covariation process of the pair $(Z_\Delta^{(u)}, Z_\Delta^{(j)})$, $3\leq u,j\leq p$ and $v\leq 0$. The process $\mathscr{W}(v)$ is, conditionally on $\mathscr{G}$, a two-sided centered Gaussian martingale with independent increments.

When $v\leq 0$, the limit process $\mathscr{W}(v)$ is defined as follows,

$$\mathscr{W}^{(j)}(v) = \begin{cases} \int_{N_b^0+v}^{N_b^0} dW_{e,s}, & j=1, \\ \int_{N_b^0+v}^{N_b^0} \widetilde{Y}_s dW_{e,s}, & j=2, \\ \mathscr{W}_{Ze}^{(j-2)}(v), & j=3,\ldots,p+2, \end{cases}$$

where $\mathscr{W}_{Ze}^{(i)}(v) \triangleq \sum_{r=1}^{p}\int_{N_b^0+v}^{N_b^0} \sigma_{Ze,s}^{(i,r)} dW_s^{1*(r)}$ $(i=1,\ldots,p)$ and analogously when $v>0$. That is, $\mathscr{W}_{Ze}(v)$ corresponds to the process $\mathscr{W}(v)$ used for the benchmark model (and so are W_s^{1*}, W_s^{2*} and $\Omega_{Ze,s}$ below). Its conditional covariance is given by

$$\widetilde{\mathbb{E}}\left(\mathscr{W}^{(u)}(v)\mathscr{W}^{(j)}(v)\right) = \Omega_{\mathscr{W}}^{(u,j)}(v) = \begin{cases} \Omega_{\mathscr{W},1}^{(u,j)}(v), & \text{if } v\leq 0 \\ \Omega_{\mathscr{W},2}^{(u,j)}(v), & \text{if } v>0 \end{cases}, \tag{S.7}$$

where $\Omega_{\mathscr{W},1}^{(u,j)}(v) = \int_{N_b^0+v}^{N_b^0} \sigma_{e,s}^2 ds$, if $u,j=1$; $\Omega_{\mathscr{W},1}^{(u,j)}(v) = \int_{N_b^0+v}^{N_b^0} \widetilde{Y}_s^2\sigma_{e,s}^2 ds$, if $u,j=2$; $\Omega_{\mathscr{W},1}^{(u,j)}(v) = \int_{N_b^0+v}^{N_b^0} \widetilde{Y}_s^2\sigma_{e,s}^2 ds$, if $1\leq u,j\leq 2$, $u\neq j$; $\Omega_{\mathscr{W},1}^{(u,j)}(v)=0$, if $u=1,2$, $j=3,\ldots,p$; $\Omega_{\mathscr{W},1}^{(u,j)}(v) = \int_{N_b^0+v}^{N_b^0} \Omega_{Ze,s}^{(u-2,j-2)} ds$ if $3\leq u,j\leq p+2$; and similarly for $\Omega_{\mathscr{W},2}^{(u,j)}(v)$. The asymptotic distribution is qualitatively the same as in Theorem 4.1. When the volatility processes are deterministic, we have convergence in law under the Skorhokod topology to the same limit process $\mathscr{W}(\cdot)$ with a Gaussian unconditional law. The case with stationary regimes is described as follows.

Assumption S.C.3. *$\Sigma^* = \{\mu_{\cdot,t},\,\Sigma_{\cdot,t},\,\sigma_{e,t}\}_{t\geq 0}$ is deterministic and the regimes are stationary.*

Let W_i^*, $i=1,2$, be two independent standard Wiener processes defined on $[0,\infty)$, starting at the origin when $s=0$. Let

$$\mathscr{V}(s) = \begin{cases} -\frac{|s|}{2} + W_1^*(s), & \text{if } s<0 \\ -\frac{(\delta^0)'\Lambda_2\delta^0}{(\delta^0)'\Lambda_1\delta^0}\frac{|s|}{2} + \left(\frac{(\delta^0)'\Omega_{\mathscr{W},2}\delta^0}{(\delta^0)'\Omega_{\mathscr{W},1}\delta^0}\right)^{1/2} W_2^*(s), & \text{if } s\geq 0. \end{cases}$$

Corollary S.C.1. *Under Assumption 4.2, S.C.1-S.C.3,*

$$\frac{\left((\delta^0)'\Lambda_1\delta^0\right)^2}{(\delta^0)'\Omega_{\mathscr{W},1}\delta^0} N\left(\widehat{\lambda}_{b,\pi}-\lambda_0\right) \Rightarrow \underset{s\in\mathcal{A}_1}{\operatorname{argmax}}\,\mathscr{V}(s), \tag{S.8}$$

where

$$\mathcal{A}_1 = \left[\frac{N\pi - N_b^0}{\|\delta^0\|^{-2}\overline{\sigma}^2}\frac{\left((\delta^0)'\Lambda_1\delta^0\right)^2}{(\delta^0)'\Omega_{\mathscr{W},1}\delta^0},\ \frac{N(1-\pi)-N_b^0}{\|\delta^0\|^{-2}\overline{\sigma}^2}\frac{\left((\delta^0)'\Lambda_1\delta^0\right)^2}{(\delta^0)'\Omega_{\mathscr{W},1}\delta^0}\right].$$

In the next two corollaries, we assume stationary errors across regimes. Corollary S.C.3 considers the basic case of a change in the mean of a sequence of *i.i.d.* random variables. Let

$$\mathscr{V}_{\mathrm{sta}}(s)=\begin{cases}-\frac{|s|}{2}+W_1^*(s), & \text{if } s<0\\ -\frac{(\delta^0)'\Lambda_2\delta^0}{(\delta^0)'\Lambda_1\delta^0}\frac{|s|}{2}+\left(\frac{(\delta^0)'\Lambda_2\delta^0}{(\delta^0)'\Lambda_1\delta^0}\right)^{1/2}W_2^*(s), & \text{if } s\geq 0\end{cases},$$

$$\mathscr{V}_{\mu,\mathrm{sta}}(s)=\begin{cases}-\frac{|s|}{2}+W_1^*(s), & \text{if } s<0\\ -\frac{|s|}{2}+W_2^*(s), & \text{if } s\geq 0\end{cases}.$$

Corollary S.C.2. *Under Assumption 4.2, S.C.1-S.C.3 and assuming that the second moments of the residual process are stationary across regimes,* $\sigma_{e,s}=\overline{\sigma}$ *for all* $0\leq s\leq N$*,*

$$\frac{(\delta^0)'\Lambda_1\delta^0}{\overline{\sigma}^2}N\left(\widehat{\lambda}_{b,\pi}-\lambda_0\right)\Rightarrow\underset{s\in\mathcal{A}_2}{\mathrm{argmax}}\,\mathscr{V}_{\mathrm{sta}}(s),$$

where

$$\mathcal{A}_2=\left[\frac{N\pi-N_b^0}{\|\delta^0\|^{-2}\overline{\sigma}^2}\frac{(\delta^0)'\Lambda_1\delta^0}{\overline{\sigma}^2},\ \frac{N(1-\pi)-N_b^0}{\|\delta^0\|^{-2}\overline{\sigma}^2}\frac{(\delta^0)'\Lambda_1\delta^0}{\overline{\sigma}^2}\right].$$

Corollary S.C.3. *Under Assumption 4.2, S.C.1-S.C.3, with* $\nu^0=0$*,* $\delta^0_{Z,i}=0$*, and* $\alpha^0_i=0$ *for* $i=1,\,2$*:*

$$\left(\delta^0/\overline{\sigma}\right)^2N\left(\widehat{\lambda}_{b,\pi}-\lambda_0\right)\Rightarrow\underset{s\in\left[\left(N\pi-N_b^0\right)(\delta^0/\overline{\sigma})^2,\ \left(N(1-\pi)-N_b^0\right)(\delta^0/\overline{\sigma})^2\right]}{\mathrm{argmax}}\mathscr{V}_{\mu,\mathrm{sta}}(s).$$

Remark S.C.1. The last corollary reports the result for the simple case of a shift in the mean of an *i.i.d.* process. This case was recently considered by Jiang et al. (2018) under a continuous-time setting in their Theorem 4.2-(b) which is similar to our Corollary S.C.3. Our limit theory differs in many respects, besides being obviously more general. Jiang et al. (2018) only develop an infeasible distribution theory for the break date estimator whereas we also derive a feasible version. This is because we introduce an assumption about the drift in order to "keep" it in the asymptotics. The limiting distribution is also derived under a different asymptotic experiment (cf. Assumption S.C.2 above and the change of time scale as discussed in Section 4). A direct consequence is that the estimate of the break fraction is shown to be consistent as $h\downarrow 0$ whereas Jiang et al. (2018) do not have such a result.

The results are similar to those in the benchmark model. However, the estimation of the regression parameters is more complicated because of the identification issues about the parameters associated with predictable processes. Nonetheless, our model specification allows us to construct feasible estimators. Given the small-dispersion specification in (S.4), we propose a two-step estimator. In fact, (S.5) essentially implies that asymptotically the evolution of the dependent variable is governed by a deterministic drift function given by $\mu^0_1+\alpha^0_1\widetilde{Y}^0_t$ (resp., $\mu^0_2+\alpha^0_2\widetilde{Y}^0_t$) if $t\leq N^0_b$ (resp., $t>N^0_b$). Thus, in a first step we construct least-squares estimates of μ^0_i and α^0_i $(i=1,\,2)$. Next, we subtract the estimate of the deterministic drift from the dependent variable so as to generate a residual component that will be used (after rescaling) as a new dependent variable in the second step where we construct the least-squares estimates of the parameters associated with the stochastic semimartigale regressors.

Proposition S.C.2. *Under Assumption 4.2, S.C.1-S.C.2, as* $h\downarrow 0$*,* $\widehat{\theta}\overset{P}{\to}\theta^0$*.*

The consistency of the estimate $\widehat{\theta}$ is all that is needed to carry out our inference procedures about the break point T^0_b presented in Section 6. The relevance of the result is that even though the drifts cannot in general be consistently estimated, we can, under our setting, estimate the parameters entering the limiting distribution; i.e., μ_i^0 and α_i^0.

S.D Mathematical Proofs

S.D.1 Additional Notations

For a matrix A, the orthogonal projection matrices P_A, M_A are defined as $P_A = A\left(A'A\right)^{-1}A'$ and $M_A = I - P_A$, respectively. For a matrix A, we use the vector-induced norm, i.e., $\|A\| = \sup_{x\neq 0}\|Ax\|/\|x\|$. Also, for a projection matrix P, $\|PA\| \leq \|A\|$. We denote the d-dimensional identity matrix by I_d. When the context is clear we omit the subscript notation in the projection matrices. We denote the $(i,\,j)$-th element of the outer product matrix $A'A$ as $\left(A'A\right)_{i,j}$ and the $i\times j$ upper-left (resp., lower-right) sub-block of $A'A$ as $\left[A'A\right]_{\{i\times j,\cdot\}}$ (resp., $\left[A'A\right]_{\{\cdot,i\times j\}}$). For a random variable ξ and a number $r\geq 1$, we write $\|\xi\|_r = \left(\mathbb{E}\|\xi\|^r\right)^{1/r}$. B and C are generic constants that may vary from line to line; we may sometime write C_r to emphasize the dependence of C on a number r. For two scalars a and b the symbol $a\wedge b$ means the infimum of $\{a,\,b\}$. The symbol "$\overset{\text{u.c.p.}}{\Rightarrow}$" signifies uniform locally in time convergence under the Skorokhod topology and recall that it implies convergence in probability. The symbol "$\overset{d}{\equiv}$" signifies equivalence in distribution. We also use the same notations as detailed in Section 2.

S.D.2 Preliminary Lemmas

Lemma S.D.1 is Lemma A.1 in Bai (1997). Let X_Δ be defined as in the display equation after (S.11).

Lemma S.D.1. *The following inequalities hold P-a.s.:*

$$\begin{aligned}&\left(Z_0'MZ_0\right)-\left(Z_0'MZ_2\right)\left(Z_2'MZ_2\right)^{-1}\left(Z_2'MZ_0\right)\\&\qquad\geq R'\left(X_\Delta'X_\Delta\right)\left(X_2'X_2\right)^{-1}\left(X_0'X_0\right)R,\quad T_b<T_b^0\end{aligned}\tag{S.1}$$

$$\begin{aligned}&\left(Z_0'MZ_0\right)-\left(Z_0'MZ_2\right)\left(Z_2'MZ_2\right)^{-1}\left(Z_2'MZ_0\right)\\&\qquad\geq R'\left(X_\Delta'X_\Delta\right)\left(X'X-X_2'X_2\right)^{-1}\left(X'X-X_0'X_0\right)R,\quad T_b\geq T_b^0.\end{aligned}\tag{S.2}$$

The following lemma presents the uniform approximation to the instantaneous covariation between continuous semimartingales. This will be useful in the proof of the convergence rate of our estimator. Below, the time window in which we study certain estimates is shrinking at a rate no faster than $h^{1-\epsilon}$ for some $0<\epsilon<1/2$.

Lemma S.D.2. *Let X_t (resp., $\widetilde{X}_t$) be a q (resp., p)-dimensional Itô continuous semimartingale defined on $[0,\,N]$. Let Σ_t denote the time t instantaneous covariation between X_t and $\widetilde{X}_t$. Choose a fixed number $\epsilon>0$ and ϖ satisfying $1/2-\epsilon\geq\varpi\geq\epsilon>0$. Further, let $B_T\triangleq\lfloor N/h-T^{\varpi}\rfloor$. Define the moving average of Σ_t as $\overline{\Sigma}_{kh}\triangleq\left(T^{\varpi}h\right)^{-1}\int_{kh}^{kh+T^{\varpi}h}\Sigma_s ds$, and let $\widehat{\Sigma}_{kh}\triangleq\left(T^{\varpi}h\right)^{-1}\sum_{i=1}^{\lfloor T^{\varpi}\rfloor}\Delta_h X_{k+i}\Delta_h\widetilde{X}_{k+i}'$. Then, $\sup_{1\leq k\leq B_T}||\widehat{\Sigma}_{kh}-\overline{\Sigma}_{kh}||=o_p(1)$. Furthermore, for each k and some $K>0$ with $N-K>kh>K$, $\sup_{T^{\epsilon}\leq T^{\varpi}\leq T^{1-\epsilon}}||\widehat{\Sigma}_{kh}-\overline{\Sigma}_{kh}||=o_p(1)$.*

Proof. By a polarization argument, we can assume that X_t and $\widetilde{X}_t$ are univariate without loss of generality, and by standard localization arguments, we can assume that the drift and diffusion coefficients of X_t and $\widetilde{X}_t$ are bounded. Then, by Itô Lemma,

$$\widehat{\Sigma}_{kh}-\overline{\Sigma}_{kh}\triangleq\frac{1}{T^{\varpi}h}\sum_{i=1}^{\lfloor T^{\varpi}\rfloor}\int_{(k+i-1)h}^{(k+i)h}\left(X_s-X_{(k+i-1)h}\right)d\widetilde{X}_s$$

$$+\frac{1}{T^{\varpi}h}\sum_{i=1}^{\lfloor T^{\varpi}\rfloor}\int_{(k+i-1)h}^{(k+i)h}\left(\widetilde{X}_s-\widetilde{X}_{(k+i-1)h}\right)dX_s.$$

For any $l\geq 1$, $||\widehat{\Sigma}_{kh}-\overline{\Sigma}_{kh}||_l\leq K_l T^{-\varpi/2}$, which follows from standard estimates for continuous Itô semimartignales. By a maximal inequality,

$$\left\|\sup_{1\leq k\leq B_T}\left|\widehat{\Sigma}_{kh}-\overline{\Sigma}_{kh}\right|\right\|_l\leq K_l T^{1/l}T^{-\varpi/2},$$

which goes to zero choosing $l>2/\varpi$. This proves the first claim. For the second, note that for $l\geq 1$,

$$\begin{aligned}\left\|\sup_{T^{\epsilon}\leq T^{\varpi}\leq T^{1-\epsilon}}\left|\widehat{\Sigma}_{kh}-\overline{\Sigma}_{kh}\right|\right\|_l&=\left\|\sup_{1\leq T^{\varpi-\epsilon}\leq T^{1-2\epsilon}}\left|\widehat{\Sigma}_{kh}-\overline{\Sigma}_{kh}\right|\right\|_l\\&\leq K_l T^{(1-2\epsilon)/l}T^{-\epsilon/2},\end{aligned}$$

and choose $l>(2-4\epsilon)/\epsilon$ to verify the claim. □

S.D.3 Preliminary Results

As it is customary in related contexts, we use a standard localization argument as explained in Section 1.d in Jacod and Shiryaev (2003) and thus we can replace Assumption 2.1-2.3 with the following stronger assumption.

Assumption S.D.1. *Let Assumption 2.1-2.3 hold. The process $\{Y_t,\, D_t,\, Z_t\}_{t\geq 0}$ takes value in some compact set, $\{\sigma_{\cdot,t}\}_{t\geq 0}$ is bounded càdlàg and the process $\{\mu_{\cdot,t}\}$ is bounded càdlàg or càglàd.*

The localization technique basically translates all the local conditions into global ones. We next introduce concepts and results which will be useful in some of the proofs below.

S.D.3.1 Approximate Variation, LLNs and CLTs

We review some basic definitions about approximate covariation and more general high-frequency statistics. Given a continuous-time semimartingales $X=\left(X^i\right)_{1\leq i\leq d}\in\mathbb{R}^d$ with zero initial value over the time horizon $[0,\,N]$, with P-a.s. continuous paths, the covariation of X over $[0,\,t]$ is denoted $[X,\,X]_t$. The $(i,\,j)$-element of the *quadratic covariation process* $[X,\,X]_t$ is defined as[4]

$$\left[X^i,\,X^j\right]_t=\operatorname*{plim}_{T\to\infty}\sum_{k=1}^{T}\left(X^i_{kh}-X^i_{(k-1)h}\right)\left(X^j_{kh}-X^j_{(k-1)h}\right),$$

where plim denotes the probability limit of the sum. $[X,\,X]_t$ takes values in the cone of all positive semidefinite symmetric $d\times d$ matrices and is continuous in t, adapted and of locally finite variation. Associated with this, we can define the (i,j)-element of the approximate covariation matrix as

$$\sum_{k\geq 1}\left({}_hX^i_{kh}-{}_hX^i_{(k-1)h}\right)\left({}_hX^j_{kh}-{}_hX^j_{(k-1)h}\right),$$

which consistently estimates the increments of the quadratic covariation $\left[X^i,\,X^j\right]$. It is an ex-post estimator of the covariability between the components of X over the time interval $[0,\,t]$. More precisely, as

[4] The reader may refer to Jacod and Protter (2012) or Jacod and Shiryaev (2003) for a complete introduction to the material of this section.

$h \downarrow 0$:

$$\sum_{k \geq 1}^{\lfloor t/h \rfloor} \left(X^i_{kh} - X^i_{(k-1)h} \right) \left(X^j_{kh} - X^j_{(k-1)h} \right) \xrightarrow{P} \int_0^t \Sigma^{(i,j)}_{XX,s} ds,$$

where $\Sigma^{(i,j)}_{XX,s}$ is referred to as the *spot (not integrated)* volatility.

After this brief review, we turn to the statement of the asymptotic results for some statistics to be encountered in the proofs below. We simply refer to Jacod and Protter (2012). More specifically, Lemma S.D.3-S.D.4 follow from their Theorem 3.3.1-(b), while Lemma S.D.5 follows from their Theorem 5.4.2.

Lemma S.D.3. *Under Assumption S.D.1, we have as $h \downarrow 0$, $T \to \infty$ with N fixed and for any $1 \leq i,\, j \leq p$,*

(i) $\left| (Z_2' e)_{i,1} \right| \xrightarrow{P} 0$ *where* $(Z_2' e)_{i,1} = \sum_{k=T_b+1}^{T} z^{(i)}_{kh} e_{kh}$;

(ii) $\left| (Z_0' e)_{i,1} \right| \xrightarrow{P} 0$ *where* $(Z_0' e)_{i,1} = \sum_{k=T_b^0+1}^{T} z^{(i)}_{kh} e_{kh}$;

(iii) $\left| (Z_2' Z_2)_{i,j} - \int_{(T_b+1)h}^{N} \Sigma^{(i,j)}_{ZZ,s} ds \right| \xrightarrow{P} 0$ *where* $(Z_2' Z_2)_{i,j} = \sum_{k=T_b+1}^{T} z^{(i)}_{kh} z^{(j)}_{kh}$;

(iv) $\left| (Z_0' Z_0)_{i,j} - \int_{(T_b^0+1)h}^{N} \Sigma^{(i,j)}_{ZZ,s} ds \xrightarrow{P} \right| 0$ *where* $(Z_0' Z_0)_{i,j} = \sum_{k=T_b^0+1}^{T} z^{(i)}_{kh} z^{(j)}_{kh}$.

For the following estimates involving X, we have, for any $1 \leq r \leq p$ and $1 \leq l \leq q + p$,

(v) $\left| (X e)_{l,1} \right| \xrightarrow{P} 0$ *where* $(X e)_{l,1} = \sum_{k=1}^{T} x^{(l)}_{kh} e_{kh}$;

(vi) $\left| (Z_2' X)_{r,l} - \int_{(T_b+1)h}^{N} \Sigma^{(r,l)}_{ZX,s} ds \right| \xrightarrow{P} 0$ *where* $(Z_2' X)_{r,l} = \sum_{k=T_b+1}^{T} z^{(r)}_{kh} x^{(l)}_{kh}$;

(vii) $\left| (Z_0' X)_{r,l} - \int_{(T_b^0+1)h}^{N} \Sigma^{(r,l)}_{ZX,s} ds \right| \xrightarrow{P} 0$ *where* $(Z_0' X)_{r,l} = \sum_{k=T_b^0+1}^{T} z^{(r)}_{kh} x^{(l)}_{kh}$.

Further, for $1 \leq u,\, d \leq q + p$,

(viii) $\left| (X' X)_{u,d} - \int_0^N \Sigma^{(u,d)}_{XX,s} ds \right| \xrightarrow{P} 0$ *where* $(X' X)_{u,d} = \sum_{k=1}^{T} x^{(u)}_{kh} x^{(d)}_{kh}$.

Lemma S.D.4. *Under Assumption S.D.1, we have as $h \downarrow 0$, $T \to \infty$ with N fixed, $\left| N_b^0 - N_b \right| > \gamma > 0$ and for any $1 \leq i,\, j \leq p$,*

(i) with $(Z_\Delta' Z_\Delta)_{i,j} = \sum_{k=T_b^0+1}^{T_b} z^{(i)}_{kh} z^{(j)}_{kh}$, *we have*

$$\begin{cases} \left| (Z_\Delta' Z_\Delta)_{i,j} - \int_{(T_b+1)h}^{T_b^0 h} \Sigma^{(i,j)}_{ZZ,s} ds \right| \xrightarrow{P} 0, & \text{if } T_b < T_b^0 \\ \left| (Z_\Delta' Z_\Delta)_{i,j} - \int_{T_b^0 h}^{(T_b+1)h} \Sigma^{(i,j)}_{ZZ,s} ds \right| \xrightarrow{P} 0, & \text{if } T_b > T_b^0 \end{cases};$$

and for $1 \leq r \leq p + q$,

(ii) with $(Z_\Delta' X_\Delta)_{i,r} = \sum_{k=T_b^0+1}^{T_b} z^{(i)}_{kh} x^{(r)}_{kh}$, *we have*

$$\begin{cases} \left| (Z_\Delta' X_\Delta)_{i,r} - \int_{(T_b+1)h}^{T_b^0 h} \Sigma^{(i,r)}_{ZX,s} ds \right| \xrightarrow{P} 0, & \text{if } T_b < T_b^0 \\ \left| (Z_\Delta' X_\Delta)_{i,r} - \int_{T_b^0 h}^{(T_b+1)h} \Sigma^{(i,r)}_{ZX,s} ds \right| \xrightarrow{P} 0, & \text{if } T_b > T_b^0 \end{cases}.$$

Next, we turn to the central limit theorems, they all feature a limiting process defined on an extension of the original probability space $(\Omega,\, \mathscr{F},\, P)$. In order to avoid non-useful repetitions, we present a general framework valid for all statistics considered in the paper. The first step is to carry out an extension of the original probability space $(\Omega,\, \mathscr{F},\, P)$. We accomplish this in the usual way. We first fix the original probability space $(\Omega,\, \mathscr{F},\, \{\mathscr{F}_t\}_{t \geq 0},\, P)$. Consider an additional measurable space $(\Omega^*,\, \mathscr{F}^*)$ and a transition probability $Q(\omega,\, d\omega^*)$ from $(\Omega,\, \mathscr{F})$ into $(\Omega^*,\, \mathscr{F}^*)$. Next, we can define the products $\widetilde{\Omega} = \Omega \times \Omega^*$, $\widetilde{\mathscr{F}} = \mathscr{F} \otimes \mathscr{F}^*$ and $\widetilde{P}(d\omega,\, d\omega^*) = P(d\omega)\, Q(\omega,\, d\omega^*)$. This defines the extension $(\widetilde{\Omega},\, \widetilde{\mathscr{F}},\, \widetilde{P})$ of the original space $(\Omega,\, \mathscr{F},\, \{\mathscr{F}_t\}_{t \geq 0},\, P)$. Any variable or process defined on either Ω or Ω^* is extended in the usual to $\widetilde{\Omega}$ as follows: for example, let Y_t be defined on Ω. Then we say that Y_t is extended in the usual way to $\widetilde{\Omega}$ by writing $Y_t(\omega, \omega^*) = Y_t(\omega)$. Further, we identify $\mathscr{F}_t$ with $\mathscr{F}_t \otimes \{\emptyset, \Omega^*\}$, so that we have a filtered space $(\widetilde{\Omega}, \widetilde{\mathscr{F}}, \{\mathscr{F}_t\}_{t\geq 0}, \widetilde{P})$. Finally, as for the filtration, we can consider another filtration $\{\widetilde{\mathscr{F}_t}\}_{t\geq 0}$ taking the product form $\widetilde{\mathscr{F}_t} = \cap_{s>t}\mathscr{F}_s \otimes \mathscr{F}^*_s$, where $\{\mathscr{F}^*_t\}_{t\geq 0}$ is a filtration on $(\Omega^*, \mathscr{F}^*)$. As for the transition probability Q we can consider the simple form $Q(\omega, d\omega^*) = P^*(d\omega^*)$ for some probability measure on $(\Omega^*, \mathscr{F}^*)$. This defines the way a product filtered extension $(\widetilde{\Omega}, \widetilde{\mathscr{F}}, \{\widetilde{\mathscr{F}_t}\}_{t\geq 0}, \widetilde{P})$ of the original filtered space $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{t\geq 0}, P)$ is constructed in this paper. Assume that the auxiliary probability space $(\Omega^*, \mathscr{F}^*, \{\mathscr{F}^*_t\}_{t\geq 0}, P^*)$ supports a p^2-dimensional standard Wiener process $W^\dagger_s$ which is adapted to $\{\widetilde{\mathscr{F}_t}\}$. We need some additional ingredients in order to describe the limiting process. We choose a progressively measurable "square-root" process σ^*_Z of the $\mathcal{M}^+_{p^2\times p^2}$-valued process $\widehat{\Sigma}_{Z,s}$, whose elements are given by $\widehat{\Sigma}^{(ij,kl)}_{Z,s} = \Sigma^{(ik)}_{Z,s}\Sigma^{(jl)}_{Z,s}$. Due to the symmetry of $\Sigma_{Z,s}$, the matrix with entries $(\sigma^{*,(ij,kl)}_{Z,s} + \sigma^{*,(ji,kl)}_{Z,s})/\sqrt{2}$ is a square-root of the matrix with entries $\widehat{\Sigma}^{(ij,kl)}_{Z,s} + \widehat{\Sigma}^{(il,jk)}_{Z,s}$. Then the process $\mathscr{U}_t$ with components $\mathscr{U}^{(r,j)}_t = 2^{-1/2}\sum_{k,l=1}^{p}\int_0^t(\sigma^{(rj,kl)}_{Z,s} + \sigma^{(jr,kl)}_{Z,s})dW^{\dagger(kl)}_s$ is, conditionally on $\mathscr{F}$, a continuous Gaussian process with independent increments and (conditional) covariance $\widetilde{\mathbb{E}}(\mathscr{U}^{(r,j)}(v)\,\mathscr{U}^{(k,l)}(v)\,|\,\mathscr{F}) = \int_{T_b^0h+v}^{T_b^0h}(\Sigma^{(rk)}_{Z,s}\Sigma^{(jl)}_{Z,s} + \Sigma^{(rl)}_{Z,s}\Sigma^{(jk)}_{Z,s})ds$, where $v\leq 0$. The CLT of interest is as follows.

Lemma S.D.5. *Let Z be a continuous Itô semimartingale satisfying Assumption S.D.1. Then,* $(Nh)^{-1/2}$ $(Z_2'Z_2 - ([Z, Z]_{Th} - [Z, Z]_{(T_b+1)h})) \overset{\mathcal{L}-\mathrm{s}}{\Rightarrow} \mathscr{U}$.

S.D.4 Proofs of the Results in Sections 3 and 4

S.D.4.1 Additional Notation

In some of the proofs we face a setting in which N_b is allowed to vary within a shrinking neighborhood of N_b^0. Some estimates only depend on observations in this window. For example, assume $T_b < T_b^0$ and consider $\sum_{k=T_b+1}^{T_b^0} x_{kh}x'_{kh}$. When N_b is allowed to vary within a shrinking neighborhood of N_b^0, this sum approximates a local window of asymptotically shrinking size. Introduce a sequence of integers $\{l_T\}$ that satisfies $l_T\to\infty$ and $l_Th\to 0$. Below we shall establish a $T^{1-\kappa}$-rate of convergence of $\widehat{\lambda}_b$ toward λ_0, considering the case where $N_b - N_b^0 = T^{-(1-\kappa)}$ for some $\kappa\in(0,\,1/2)$. Hence, define

$$\widehat{\Sigma}_X\left(T_b,\,T_b^0\right) \triangleq \sum_{k=T_b+1}^{T_b^0} x_{kh}x'_{kh} = \sum_{k=T_b^0+1-l_T}^{T_b^0} x_{kh}x'_{kh}, \tag{S.3}$$

where now $l_T = \lfloor T^\kappa\rfloor \to\infty$ and $l_Th = h^{1-\kappa}\to 0$. Note that $1/h^{1-\kappa}$ is the rate of convergence and the interpretation for $\widehat{\Sigma}_X(T_b,\,T_b^0)$ is that it involves asymptotically an infinite number of observations falling in the shrinking (at rate $h^{1-\kappa}$) block $((T_b-1)\,h,\,T_b^0h]$. Other statistics involving the regressors and errors are defined similarly:

$$\widehat{\Sigma}_{Xe}\left(T_b,\,T_b^0\right) \triangleq \sum_{k=T_b+1}^{T_b^0} x_{kh}e_{kh} = \sum_{k=T_b^0+1-l_T}^{T_b^0} x_{kh}e_{kh}, \tag{S.4}$$

and

$$\widehat{\Sigma}_{Ze}\left(T_b,\,T_b^0\right) \triangleq \sum_{k=T_b^0+1-l_T}^{T_b^0} z_{kh}e_{kh}. \tag{S.5}$$

Further, we let $\overline{\Sigma}_{Xe}\left(T_b,\,T_b^0\right)\triangleq h^{-(1-\kappa)}\int_{N_b}^{N_b^0}\Sigma_{Xe,s}ds$ and analogously when Z replaces X. We also define

$$\widehat{\Sigma}_{h,X}\left(T_b,\,T_b^0\right)\triangleq h^{-(1-\kappa)}\sum_{k=T_b^0+1-l_T}^{T_b^0}x_{kh}x_{kh}'. \tag{S.6}$$

The proofs of Section 4 are first given for the case where $\mu_{\cdot,t}$ from equation (2.3) are identically zero. In the last step, this is relaxed. Furthermore, throughout the proofs we proceed conditionally on the processes $\mu_{\cdot,t}$ and Σ_t^0 (defined in Assumption 2.3) so that they are treated as if they were deterministic. This is a natural strategy since the processes $\mu_{\cdot,t}$ are of higher order in h and they do not play any role for the asymptotic results [cf. Barndorff-Nielsen and Shephard (2004)].

S.D.4.2 Proof of Proposition 3.1

Proof. The concentrated sample objective function evaluated at $\widehat{T}_b$ is $Q_T(\widehat{T}_b)=\widehat{\delta}_{T_b}'\left(Z_2'MZ_2\right)\widehat{\delta}_{T_b}$. We have

$$\widehat{\delta}_{T_b}=\left(Z_2'MZ_2\right)^{-1}\left(Z_2'MY\right)=\left(Z_2'MZ_2\right)^{-1}\left(Z_2'MZ_0\right)\delta_Z^0+\left(Z_2'MZ_2\right)^{-1}Z_2Me,$$

and $\widehat{\delta}_{T_b^0}=\left(Z_0'MZ_0\right)^{-1}\left(Z_0'MY\right)=\delta_Z^0+\left(Z_0'MZ_0\right)^{-1}\left(Z_0'Me\right)$. Therefore,

$$\begin{aligned}Q_T\left(T_b\right)-Q_T\left(T_b^0\right)&=\widehat{\delta}_{T_b}'\left(Z_2'MZ_2\right)\widehat{\delta}_{T_b}-\widehat{\delta}_{T_b^0}'\left(Z_0'MZ_0\right)\widehat{\delta}_{T_b^0} &\text{(S.7)}\\ &=\left(\delta_Z^0\right)'\left\{\left(Z_0'MZ_2\right)\left(Z_2'MZ_2\right)^{-1}\left(Z_2'MZ_0\right)-Z_0'MZ_0\right\}\delta_Z^0 &\text{(S.8)}\\ &\quad+g_e\left(T_b\right), &\text{(S.9)}\end{aligned}$$

where

$$\begin{aligned}g_e\left(T_b\right)&=2\left(\delta_Z^0\right)'\left(Z_0'MZ_2\right)\left(Z_2'MZ_2\right)^{-1}Z_2Me-2\left(\delta_Z^0\right)'\left(Z_0'Me\right) &\text{(S.10)}\\ &\quad+e'MZ_2\left(Z_2'MZ_2\right)^{-1}Z_2Me-e'MZ_0\left(Z_0'MZ_0\right)^{-1}Z_0'Me. &\text{(S.11)}\end{aligned}$$

Denote

$$\begin{aligned}&X_\Delta\triangleq X_2-X_0=\left(0,\,\ldots,\,0,\,x_{(T_b+1)h},\ldots,\,x_{T_b^0h},\,0,\ldots,\right)', &&\text{for } T_b<T_b^0\\ &X_\Delta\triangleq-\left(X_2-X_0\right)=\left(0,\,\ldots,\,0,\,x_{(T_b^0+1)h},\ldots,\,x_{T_bh},\,0,\ldots,\right)', &&\text{for } T_b>T_b^0\\ &X_\Delta\triangleq 0, &&\text{for } T_b=T_b^0.\end{aligned}$$

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

$$r\left(T_b\right)\triangleq\frac{\left(\delta_Z^0\right)'\left\{\left(Z_0'MZ_0\right)-\left(Z_0'MZ_2\right)\left(Z_2'MZ_2\right)^{-1}\left(Z_2'MZ_0\right)\right\}\delta_Z^0}{\left|T_b-T_b^0\right|}. \tag{S.12}$$

We arbitrarily define $r\left(T_b\right)=\left(\delta_Z^0\right)'\delta_Z^0$ when $T_b=T_b^0$. We write (S.7) as

$$Q_T\left(T_b\right)-Q_T\left(T_0\right)=-\left|T_b-T_b^0\right|r\left(T_b\right)+g_e\left(T_b\right),\qquad\text{for all } T_b. \tag{S.13}$$

By definition, $\widehat{T}_b$ is an extremum estimator and thus satisfies $g_e(\widehat{T}_b) \geq \left|\widehat{T}_b - T_b^0\right| r(\widehat{T}_b)$. Therefore,

$$\begin{aligned}
P\left(\left|\widehat{\lambda}_b - \lambda_0\right| > K\right) &= P\left(\left|\widehat{T}_b - T_b^0\right| > TK\right) \\
&\leq P\left(\sup_{\left|T_b - T_b^0\right| > TK} \left|g_e\left(T_b\right)\right| \geq \inf_{\left|T_b - T_b^0\right| > TK} \left|T_b - T_b^0\right| r\left(T_b\right)\right) \\
&\leq P\left(\sup_{p \leq T_b \leq T-p} \left|g_e\left(T_b\right)\right| \geq TK \inf_{\left|T_b - T_b^0\right| > TK} r\left(T_b\right)\right) \qquad \text{(S.14)} \\
&= P\left(r_T^{-1} \sup_{p \leq T_b \leq T-p} \left|g_e\left(T_b\right)\right| \geq K\right),
\end{aligned}$$

where we recall that $p \leq T_b \leq T-p$ is needed for identification, and $r_T \triangleq T \inf_{\left|T_b - T_b^0\right| > TK} r\left(T_b\right)$. Lemma S.D.6 below shows that r_T is positive and bounded away from zero. Thus, it is sufficient to verify that the stochastic component is negligible as $h \downarrow 0$, i.e.,

$$\sup_{p \leq T_b \leq T-p} \left|g_e\left(T_b\right)\right| = o_p\left(1\right). \qquad \text{(S.15)}$$

The first term of $g_e\left(T_b\right)$ is

$$2\left(\delta_Z^0\right)' \left(Z_0' M Z_2\right) \left(Z_2' M Z_2\right)^{-1/2} \left(Z_2' M Z_2\right)^{-1/2} Z_2 M e. \qquad \text{(S.16)}$$

Lemma S.D.5 implies that for any $1 \leq j \leq p$, $\left(Z_2 e\right)_{j,1} / \sqrt{h} = O_p\left(1\right)$ and for any $1 \leq i \leq q+p$, $\left(Xe\right)_{i,1} / \sqrt{h} = O_p\left(1\right)$. These hold because they both involve a positive fraction of the data. Furthermore, from Lemma S.D.3, we also have that $Z_2' M Z_2$ and $Z_0' M Z_2$ are $O_p\left(1\right)$. Therefore, the supremum of $\left(Z_0' M Z_2\right) \left(Z_2' M Z_2\right)^{-1/2}$ over all T_b is

$$\sup_{T_b} \left(Z_0' M Z_2\right) \left(Z_2' M Z_2\right)^{-1} \left(Z_2' M Z_0\right) \leq Z_0' M Z_0 = O_p\left(1\right),$$

by Lemma S.D.3. By Assumption (2.1)-(iii), $\left(Z_2' M Z_2\right)^{-1/2} Z_2' M e$ is $O_p\left(1\right) O_p(\sqrt{h})$ uniformly, which implies that (S.16) is $O_p(\sqrt{h})$ uniformly over $p \leq T_b \leq T-p$. As for the second term of (S.10), $Z_0' M e = O_p(\sqrt{h})$. The first term in (S.11) is uniformly $o_p\left(1\right)$ and the same holds for the last term. Therefore, combining these results, $\sup_{T_b} \left|g_e\left(T_b\right)\right| = O_p(\sqrt{h})$ uniformly when $|\widehat{\lambda}_b - \lambda_0| > K$. Therefore for some $B > 0$, these arguments combined with Lemma S.D.6 below result in $P(r_B^{-1} \sup_{p \leq T_b \leq T-p} \left|g_e\left(T_b\right)\right| \geq K) \leq \varepsilon$, from which it follows that the right-hand side of (S.14) is weakly smaller than ε. This concludes the proof since $\varepsilon > 0$ was arbitrary. □

Lemma S.D.6. *For $B > 0$, let $r_B = \inf_{\left|T_b - T_b^0\right| > TB} Tr\left(T_b\right)$. There exists a $\kappa > 0$ such that for every $\varepsilon > 0$, there exists a $B < \infty$ such that $P\left(r_B \geq \kappa\right) \leq 1 - \varepsilon$, i.e., r_B is positive and bounded away from zero with high probability.*

Proof. Assume $T_b \leq T_b^0$ and observe that $r_T \geq r_B$ for an appropriately chosen B. From the first inequality result in Lemma S.D.1,

$$r\left(T_b\right) \geq \left(\delta_Z^0\right)' R' \left(X_\Delta' X_\Delta / \left(T_b^0 - T_b\right)\right) \left(X_2' X_2\right)^{-1} \left(X_0' X_0\right) R \delta_Z^0.$$

When multiplied by T, we have

$$\begin{aligned}Tr\left(T_{b}\right) &\geq T\left(\delta_{Z}^{0}\right)' R' \frac{X_{\Delta}' X_{\Delta}}{T_{b}^{0}-T_{b}}\left(X_{2}' X_{2}\right)^{-1}\left(X_{0}' X_{0}\right) R \delta_{Z}^{0} \\ &= \left(\delta_{Z}^{0}\right)' R' \frac{X_{\Delta}' X_{\Delta}}{N_{b}^{0}-N_{b}}\left(X_{2}' X_{2}\right)^{-1}\left(X_{0}' X_{0}\right) R \delta_{Z}^{0}.\end{aligned}$$

Note that $0<K<B<h\left(T_{b}^{0}-T_{b}\right)<N$. Then,

$$Tr\left(T_{b}\right) \geq\left(\delta_{Z}^{0}\right)' R'\left(X_{\Delta}' X_{\Delta} / N\right)\left(X_{2}' X_{2}\right)^{-1}\left(X_{0}' X_{0}\right) R \delta_{Z}^{0},$$

and by standard estimates for Itô semimartingales, $X_{\Delta}' X_{\Delta}=O_{p}(1)$ (i.e., use the Burkhölder-Davis-Gundy inequality, recalling that $|\widehat{N}_{b}-N_{b}^{0}|>B N$). Hence, we conclude that $Tr\left(T_{b}\right) \geq\left(\delta_{Z}^{0}\right)' R' O_{p}(1 / N) O_{p}(1) R \delta_{Z}^{0}$ $\geq \kappa>0$, where κ is some positive constant. The last inequality follows whenever $X_{\Delta}' X_{\Delta}$ is positive definite since $R' X_{\Delta}' X_{\Delta}\left(X_{2}' X_{2}\right)^{-1}\left(X_{0}' X_{0}\right) R$ can be rewritten as $R'[(X_{0}' X_{0})^{-1}+(X_{\Delta}' X_{\Delta})^{-1}] R$. According to Lemma S.D.3, $X_{2}' X_{2}$ is $O_{p}(1)$. The same argument applies to $X_{0}' X_{0}$, which together with the the fact that R has full common rank in turn implies that we can choose a $B>0$ such that $r_{B}=\inf_{|T_{b}-T_{b}^{0}|>T B} Tr\left(T_{b}\right)$ satisfies $P\left(r_{B} \geq \kappa\right) \leq 1-\varepsilon$. The case with $T_{b}>T_{b}^{0}$ is similar and omitted. □

S.D.4.3 Proof of Proposition 3.2

Proof. Given the consistency result, one can restrict attention to the local behavior of the objective function for those values of T_{b} in $\mathbf{B}_{T} \triangleq\left\{T_{b}: T \eta \leq T_{b} \leq T(1-\eta)\right\}$, where $\eta>0$ satisfies $\eta \leq \lambda_{0} \leq 1-\eta$. By Proposition 3.1, the estimator $\widehat{T}_{b}$ will visit the set $\mathbf{B}_{T}$ with large probability as $T \rightarrow \infty$. That is, for any $\varepsilon>0$, $P\left(\widehat{T}_{b} \notin \mathbf{B}_{T}\right)<\varepsilon$ for sufficiently large T. We show that for large T, $\widehat{T}_{b}$ eventually falls in the set $\mathbf{B}_{K, T} \triangleq\left\{T_{b}:\left|N_{b}-N_{b}^{0}\right| \leq K T^{-1}\right\}$, for some $K>0$. For any $K>0$, define the intersection of $\mathbf{B}_{T}$ and the complement of $\mathbf{B}_{K, T}$ by

$$\mathbf{D}_{K, T} \triangleq\left\{T_{b}: N \eta \leq N_{b} \leq N(1-\eta),\left|N_{b}-N_{b}^{0}\right|>K T^{-1}\right\}.$$

Notice that

$$\begin{aligned}&\left\{\left|\widehat{\lambda}_{b}-\lambda_{0}\right|>K T^{-1}\right\}= \\ &\quad\left\{\left|\widehat{\lambda}_{b}-\lambda_{0}\right|>K T^{-1} \cap \widehat{\lambda}_{b} \in(\eta, 1-\eta)\right\} \\ &\quad \cup\left\{\left|\widehat{\lambda}_{b}-\lambda_{0}\right|>K T^{-1} \cap \widehat{\lambda}_{b} \notin(\eta, 1-\eta)\right\} \\ &\quad \subseteq\left\{\left|\widehat{\lambda}_{b}-\lambda_{0}\right|>K\left(T^{-1}\right) \cap \widehat{\lambda}_{b} \in(\eta, 1-\eta)\right\} \cup\left\{\widehat{\lambda}_{b} \notin(\eta, 1-\eta)\right\},\end{aligned}$$

and so

$$\begin{aligned}P\left(\left|\widehat{\lambda}_{b}-\lambda_{0}\right|>K T^{-1}\right) \leq\; & P\left(\widehat{\lambda}_{b} \notin(\eta, 1-\eta)\right) \\ & +P\left(\left|\widehat{T}_{b}-T_{b}^{0}\right|>K \cap \widehat{\lambda}_{b} \in(\eta, 1-\eta)\right),\end{aligned}$$

and for large T,

$$\begin{aligned}P\left(\left|\widehat{\lambda}_{b}-\lambda_{0}\right|>K T^{-1}\right) &\leq \varepsilon+P\left(\left|\widehat{\lambda}_{b}-\lambda_{0}\right|>K T^{-1} \cap \widehat{\lambda}_{b} \in(\eta, 1-\eta)\right) \\ &\leq \varepsilon+P\left(\sup_{T_{b} \in \mathbf{D}_{K, T}} Q_{T}\left(T_{b}\right) \geq Q_{T}\left(T_{b}^{0}\right)\right).\end{aligned}$$

Therefore it is enough to show that the second term above is negligible as $h \downarrow 0$. Suppose $T_b < T_b^0$. Since $\widehat{T}_b = \arg\max Q_T(T_b)$, it is enough to show that $P(\sup_{T_b \in \mathbf{D}_{K,T}} Q_T(T_b) \geq Q_T(T_b^0)) < \varepsilon$. Note that this implies $|T_b - T_b^0| > KN^{-1}$. Therefore, we have to deal with a setting where the time span in $\mathbf{D}_{K,T}$ between N_b and N_b^0 is actually shrinking. The difficulty arises from the quantities depending on the difference $|N_b - N_b^0|$. We can rewrite $Q_T(T_b) \geq Q_T(T_b^0)$ as $g_e(T_b) / |T_b - T_b^0| \geq r(T_b)$, with $g_e(T_b)$ and $r(T_b)$ as defined above. Thus, we need to show,

$$P\left(\sup_{T_b \in \mathbf{D}_{K,T}} h^{-1} \frac{g_e(T_b)}{|T_b - T_b^0|} \geq \inf_{T_b \in \mathbf{D}_{K,T}} h^{-1} r(T_b)\right) < \varepsilon.$$

By Lemma S.D.1,

$$\inf_{T_b \in \mathbf{D}_{K,T}} r(T_b) \geq \inf_{T_b \in \mathbf{D}_{K,T}} \left(\delta_Z^0\right)' R' \frac{X_\Delta' X_\Delta}{|T_b - T_b^0|} \left(X_2' X_2\right)^{-1} \left(X_0' X_0\right) R \delta_Z^0.$$

The asymptotic results used so far rely on statistics involving integrated covariation between continuous semimartingales. However, since $|T_b - T_b^0| > K/N$, the context is different and the same results do not apply because the time horizon is decreasing as the sample size increases for quantities depending on $|N_b - N_b^0|$. Thus, we shall apply asymptotic results for the local approximation of the covariation between processes. Moreover, when $|T_b - T_b^0| > K/N$, there are at least K terms in this sum with asymptotically vanishing moments. That is, for any $1 \leq i,\, j \leq q+p$, we have $\mathbb{E}[x_{kh}^{(i)} x_{kh}^{(j)} | \mathscr{F}_{(k-1)h}] = \Sigma_{X,(k-1)h}^{(i,j)} h$, and note that $x_{kh}/\sqrt{h}$ is i.n.d. with finite variance and thus by Assumption 3.1 we can always choose a K large enough such that $\left(h\,|T_b - T_b^0|\right)^{-1} X_\Delta' X_\Delta = \left(h\,|T_b - T_b^0|\right)^{-1} \sum_{k=T_b+1}^{T_b^0} x_{kh} x_{kh}' = A > 0$ for all $T_b \in \mathbf{D}_{K,T}$. This shows that $\inf_{T_b \in \mathbf{D}_{K,T}} h^{-1} r(T_b)$ is bounded away from zero. Note that for the other terms in $r(T_b)$ we can use the same arguments since they do not depend on $|N_b - N_b^0|$. Hence,

$$P\left(\sup_{T_b \in \mathbf{D}_{K,T}} h^{-1} \left(T_b^0 - T_b\right)^{-1} g_e(T_b) \geq B/N\right) < \varepsilon, \tag{S.17}$$

for some $B > 0$. Consider the terms of $g_e(T_b)$ in (S.10). When $T_b \in \mathbf{D}_{K,T}$, Z_2 involves at least a positive fraction $N\eta$ of the data. From Lemma S.D.3, as $h \downarrow 0$, it follows that

$$\begin{aligned}
h^{-1} & \left(T_b^0 - T_b\right)^{-1} e' M Z_2 \left(Z_2' M Z_2\right)^{-1} Z_2 M e \\
& = \left(T_b^0 - T_b\right)^{-1} h^{-1} O_p\left(h^{1/2}\right) O_p(1)\, O_p\left(h^{1/2}\right) = \frac{O_p(1)}{T_b^0 - T_b},
\end{aligned}$$

uniformly in T_b. Choose K large enough so that the probability that the right-hand size is larger than B/N is less than $\varepsilon/4$. A similar argument holds for the second term in (S.11). Next consider the first term of $g_e(T_b)$ in (S.10). Using $Z_2 = Z_0 \pm Z_\Delta$ we can deduce that

$$\begin{aligned}
& \left(\delta_Z^0\right)' \left(Z_0' M Z_2\right) \left(Z_2' M Z_2\right)^{-1} Z_2 M e \\
& \quad = \left(\delta_Z^0\right)' \left(\left(Z_2' \pm Z_\Delta'\right) M Z_2\right) \left(Z_2' M Z_2\right)^{-1} Z_2 M e \\
& \quad = \left(\delta_Z^0\right)' Z_0' M e \pm \left(\delta_Z^0\right)' Z_\Delta' M e \\
& \qquad \pm \left(\delta_Z^0\right)' \left(Z_\Delta' M Z_2\right) \left(Z_2' M Z_2\right)^{-1} Z_2 M e,
\end{aligned} \tag{S.18}$$

from which it follows that

$$\begin{aligned}&\left|2\left(\delta_Z^0\right)'\left(Z_0'MZ_2\right)\left(Z_2'MZ_2\right)^{-1}Z_2Me-2\left(\delta_Z^0\right)'\left(Z_0'Me\right)\right|\\&=\left|\left(\delta_Z^0\right)'Z_\Delta'Me\right|+\left|\left(\delta_Z^0\right)'\left(Z_\Delta'MZ_2\right)\left(Z_2'MZ_2\right)^{-1}\left(Z_2Me\right)\right|.\end{aligned}\tag{S.19}$$

First, we can apply Lemma S.D.3 [(vi) and (viii)], and Lemma S.D.4 [(i)-(ii)], together with Assumption 2.1-(iii), to terms that do not involve $|N_b-N_b^0|$, i.e.,

$$\begin{aligned}h^{-1}\left(\delta_Z^0\right)'\left(Z_\Delta'MZ_2\right)&=h^{-1}\left(\delta_Z^0\right)'\left(Z_\Delta'Z_2\right)-h^{-1}\left(\delta_Z^0\right)'\left(Z_\Delta'X_\Delta\left(X'X\right)^{-1}X'Z_2\right)\\&=\frac{\left(\delta_Z^0\right)'\left(Z_\Delta'Z_\Delta\right)}{h}-\left(\delta_Z^0\right)'\left(\frac{Z_\Delta'X_\Delta}{h}\left(X'X\right)^{-1}X'Z_2\right).\end{aligned}$$

Consider $Z_\Delta'Z_\Delta$. By the same reasoning as above, whenever $T_b\in\mathbf{D}_{K,T}$, $\left(Z_\Delta'Z_\Delta\right)/h\left(T_b^0-T_b\right)=O_p(1)$ for K large enough. The term $Z_\Delta'X_\Delta/h\left(T_b^0-T_b\right)$ is also $O_p(1)$ uniformly. Thus, it follows from Lemma S.D.5 that the second term of (S.19) is $O_p(h^{1/2})$. Next, note that $Z_\Delta'Me=Z_\Delta'e-Z_\Delta'X\left(X'X\right)^{-1}X'e$. We can write

$$\begin{aligned}\frac{Z_\Delta'Me}{\left(T_b^0-T_b\right)h}=&\frac{1}{\left(T_b^0-T_b\right)h}\sum_{k=T_b+1}^{T_b^0}z_{kh}e_{kh}\\&-\frac{1}{\left(T_b^0-T_b\right)h}\left(\sum_{k=T_b+1}^{T_b^0}z_{kh}x_{kh}'\right)\left(X'X\right)^{-1}\left(X'e\right).\end{aligned}$$

Note that the sequence $\{h^{-1/2}z_{kh}h^{-1/2}x_{kh}\}$ is i.n.d. with finite mean identically in k. There are at least K terms in this sum, so $(\sum_{k=T_b+1}^{T_b^0}z_{kh}x_{kh}')/\left(T_b^0-T_b\right)h$ is $O_p(1)$ for a large enough K in view of Assumption 3.1. Then,

$$\frac{1}{\left(T_b^0-T_b\right)h}\left(\sum_{k=T_b+1}^{T_b^0}z_{kh}x_{kh}'\right)\left(X'X\right)^{-1}\left(X'e\right)=O_p(1)\,O_p(1)\,O_p\left(h^{1/2}\right),\tag{S.20}$$

when K is large. Thus,

$$\frac{1}{\left(T_b^0-T_b\right)h}g_e\left(T_b\right)=\frac{1}{\left(T_b^0-T_b\right)h}\left(\delta^0\right)'2Z_\Delta'e+\frac{O_p(1)}{T_b^0-T_b}+O_p\left(h^{1/2}\right).\tag{S.21}$$

We can now prove (S.17) using (S.21). To this end, we sneed a $K>0$, such that

$$\begin{aligned}&P\left(\sup_{T_b\in\mathbf{D}_{K,T}}\left\|\left(\delta_Z^0\right)'\frac{2}{h}\frac{1}{T_b^0-T_b}\sum_{k=T_b+1}^{T_b^0}z_{kh}e_{kh}\right\|>\frac{B}{4N}\right)\\&\leq P\left(\sup_{T_b\leq T_b^0-KN^{-1}}\left\|\frac{1}{h}\frac{1}{T_b^0-T_b}\sum_{k=T_b+1}^{T_b^0}z_{kh}e_{kh}\right\|>\frac{B}{8N\left\|\delta_Z^0\right\|}\right)<\varepsilon.\end{aligned}\tag{S.22}$$

Note that $|T_b-T_b^0|$ is bounded away from zero in $\mathbf{D}_{K,T}$. Observe that $(z_{kh}/\sqrt{h})(e_{kh}/\sqrt{h})$ are independent in k and have zero mean and finite second moments. Hence, by the Hájek-Réiny inequality [see Lemma

A.6 in Bai and Perron (1998)],

$$P\left(\sup_{T_b\leq T_b^0-KN^{-1}}\left\|\frac{1}{T_b^0-T_b}\sum_{k=T_b+1}^{T_b^0}\frac{z_{kh}}{\sqrt{h}}\frac{e_{kh}}{\sqrt{h}}\right\|>\frac{B}{8\left\|\delta_Z^0\right\|N}\right)$$
$$\leq A\frac{64\left\|\delta_Z^0\right\|^2N^2}{B^2}\frac{1}{KN^{-1}},$$

where $A>0$. We can choose K large enough such that the right-hand side is less than $\varepsilon/4$. Combining the above arguments, we deduce the claim in (S.17) which then concludes the proof of Proposition 3.2. □

S.D.4.4 Proof of Proposition 3.3

We focus on the case with $T_b\leq T_0$. The arguments for the other case are similar and omitted. From Proposition 3.1 the distance $|\widehat{\lambda}_b-\lambda_0|$ can be made arbitrary small. Proposition 3.2 gives the associated rate of convergence: $T(\widehat{\lambda}_b-\lambda_0)=O_p(1)$. Given the consistency result for $\widehat{\lambda}_b$, we can apply a restricted search. In particular, by Proposition 3.2, for large $T>\overline{T}$, we know that $\{T_b\notin\mathbf{D}_{K,T}\}$, or equivalently $|T_b-T_b^0|\leq K$, with high probability for some K. Essentially, what we shall show is that from the results of Proposition 3.1-3.2 the error in replacing T_b^0 with $\widehat{T}_b$ is stochastically small and thus it does not affect the estimation of the parameters β^0, $\delta_{Z,1}^0$ and $\delta_{Z,2}^0$. Toward this end, we first find a lower bound on the convergence rate for $\widehat{\lambda}_b$ that guarantees its estimation to be asymptotically independent from that of the regression parameters. This result will also be used in later proofs. We shall see that the rate of convergence established in Proposition 3.2 is strictly faster than the lower bound. Below, we use $\widehat{T}_b$ in order to construct Z_2 and define $\widehat{Z}_0\triangleq Z_2$.

Lemma S.D.7. *Fix $\gamma\in(0,\,1/2)$ and some constant $A>0$. For all large $T>\overline{T}$, if $|\widehat{N}_b-N_b^0|\leq AO_p\left(h^{1-\gamma}\right)$, then $X'(Z_0-\widehat{Z}_0)=O_p(h^{1-\gamma})$ and $Z_0'(Z_0-\widehat{Z}_0)=O_p(h^{1-\gamma})$.*

Proof. Note that the setting of Proposition 3.2 satisfies the conditions of this lemma because $\widehat{N}_b-N_b^0=O_p(h)\leq AO_p\left(h^{1-\gamma}\right)$ as $h\downarrow0$. By assumption, there exists some constant $C>0$ such that $P(h^{\gamma}|\widehat{T}_b-T_b^0|>C)<\varepsilon$. We have to show that although we only know $|\widehat{T}_b-T_b^0|\leq Ch^{-\gamma}$, the error when replacing T_b^0 by $\widehat{T}_b$ in the construction of Z_2 goes to zero fast enough. This is achieved because $|\widehat{N}_b-N_b^0|\to0$ at least at rate $h^{1-\gamma}$, which is faster than the standard convergence rate for regression parameters (i.e., $\sqrt{T}$-rate). Without loss of generality we take $C=1$. We have

$$h^{-1/2}X'\left(Z_0-\widehat{Z}_0\right)=h^{1/2-\gamma}\frac{1}{h^{1-\gamma}}\sum_{T_b^0-\lfloor T^{\gamma}\rfloor}^{T_b^0}x_{kh}z_{kh}.$$

Notice that, as $h\downarrow0$, the number of terms in the sum on the right-hand side, for all $T>\overline{T}$, increases to infinity at rate $1/h^{\gamma}$. Since $\widehat{N}_b$ approaches N_b^0 at rate $T^{-(1-\gamma)}$, the quantity $X'(Z_0-\widehat{Z}_0)/h^{1-\gamma}$ is a consistent estimate of the so-called instantaneous or spot covariation between X and Z at time N_b^0. Theorem 9.3.2 part (i) in Jacod and Protter (2012) can be applied since the "window" is decreasing at rate $h^{1-\gamma}$ and the same factor $h^{1-\gamma}$ is in the denominator. Thus, we have as $h\downarrow0$,

$$X_{\Delta}'Z_{\Delta}/h^{1-\gamma}\xrightarrow{P}\Sigma_{XX,N_b^0},\tag{S.23}$$

which implies that $h^{-1/2}X'\left(Z_0-\widehat{Z}_0\right)=O_p(h^{1/2-\gamma})$. This shows that the order of the error in replacing

Z_0 by $Z_2 = \widehat{Z}_0$ goes to zero at a fast enough rate. That is, by definition we can write

$$Y = X\beta^0 + \widehat{Z}_0\delta_Z^0 + \left(Z_0 - \widehat{Z}_0\right)\delta_Z^0 + e,$$

from which it follows that

$$X'\widehat{Z}_0 = X'Z_0 + o_p(1), \; X'\left(Z_0 - \widehat{Z}_0\right)\delta_Z^0 = o_p(1),$$

and $Z_0'(Z_0 - \widehat{Z}_0)\delta_Z^0 = o_p(1)$. To see this, consider for example

$$X'\left(\widehat{Z}_0 - Z_0\right) = \sum_{T_b^0 - \lfloor T^\gamma \rfloor}^{T_b^0} x_{kh}z_{kh} = \frac{h^{1-\gamma}}{h^{1-\gamma}} \sum_{T_b^0 - \lfloor T^\gamma \rfloor}^{T_b^0} x_{kh}z_{kh} = h^{1-\gamma}O_p(1),$$

which clearly implies that $X'\widehat{Z}_0 = X'Z_0 + o_p(1)$. The other case can be proved similarly. This concludes the proof of the Lemma. □

Using Lemma S.D.7, the proof of the proposition becomes simple.

Proof of Proposition 3.3. By standard arguments,

$$\sqrt{T}\begin{bmatrix}\widehat{\beta} - \beta^0 \\ \widehat{\delta} - \delta_Z^0\end{bmatrix} = \begin{bmatrix}X'X & X'\widehat{Z}_0 \\ \widehat{Z}_0'X & \widehat{Z}_0'\widehat{Z}_0\end{bmatrix}^{-1} \sqrt{T}\begin{bmatrix}X'e + X'\left(Z_0 - \widehat{Z}_0\right)\delta_Z^0 \\ \widehat{Z}_0'e + \widehat{Z}_0'\left(Z_0 - \widehat{Z}_0\right)\delta_Z^0\end{bmatrix},$$

from which it follows that

$$\begin{bmatrix}X'X & X'\widehat{Z}_0 \\ \widehat{Z}_0'X & \widehat{Z}_0'\widehat{Z}_0\end{bmatrix}^{-1} \frac{1}{h^{1/2}}X'\left(Z_0 - \widehat{Z}_0\right)\delta_Z^0 = O_p(1)\,o_p(1) = o_p(1),$$

and a similar reasoning applies to $\widehat{Z}_0'(Z_0 - \widehat{Z}_0)\delta_Z^0$. All other terms involving $\widehat{Z}_0$ can be treated in analogous fashion. In particular, the $O_p(1)$ result above follows from Lemma S.D.3-S.D.4. The rest of the arguments (including mixed normality) follows from Barndorff-Nielsen and Shephard (2004) and Li et al. (2017) and are omitted. □

S.D.4.5 Proof of Proposition 4.1

Proof of part (i) of Proposition 4.1. Below C is a generic positive constant which may change from line to line. Let $\widetilde{e}$ denote the vector of normalized residuals $\widetilde{e}_t$ defined by (4.1). Recall that $\widehat{T}_b = \arg\max_{T_b} Q_T(T_b)$, $Q_T(\widehat{T}_b) = \widehat{\delta}_{T_b}'(Z_2'MZ_2)\widehat{\delta}_{T_b}$, and the decomposition

$$\begin{aligned} Q_T(T_b) - Q_T\left(T_b^0\right) &= \widehat{\delta}_{T_b}'(Z_2'MZ_2)\widehat{\delta}_{T_b} - \widehat{\delta}_{T_b^0}'(Z_0'MZ_0)\widehat{\delta}_{T_b^0} && \text{(S.24)} \\ &= \delta_h'\left\{(Z_0'MZ_2)(Z_2'MZ_2)^{-1}(Z_2'MZ_0) - Z_0'MZ_0\right\}\delta_h && \text{(S.25)} \\ &\quad + g_e(T_b), && \text{(S.26)} \end{aligned}$$

where

$$\begin{aligned} g_e(T_b) &= 2\delta_h'(Z_0'MZ_2)(Z_2'MZ_2)^{-1}Z_2Me - 2\delta_h'(Z_0'Me) && \text{(S.27)} \\ &\quad + e'MZ_2(Z_2'MZ_2)^{-1}Z_2Me - e'MZ_0(Z_0'MZ_0)^{-1}Z_0'Me. && \text{(S.28)} \end{aligned}$$

Since $g_e(\widehat{T}_b) \geq |\widehat{T}_b - T_b^0| r(\widehat{T}_b)$, we have

$$
\begin{aligned}
& P\left(\left|\widehat{\lambda}_b - \lambda_0\right| > K\right) \\
& = P\left(\left|\widehat{T}_b - T_b^0\right| > TK\right) \\
& \leq P\left(\sup_{\left|T_b - T_b^0\right| > TK} h^{-1/2} \left|g_e\left(T_b\right)\right| \geq \inf_{\left|T_b - T_b^0\right| > TK} h^{-1/2} \left|T_b - T_b^0\right| r\left(T_b\right)\right) \\
& \leq P\left(\sup_{p \leq T_b \leq T-p} h^{-1/2} \left|g_e\left(T_b\right)\right| \geq TK \inf_{\left|T_b - T_b^0\right| > TK} h^{-1/2} r\left(T_b\right)\right) \\
& = P\left(r_T^{-1} \sup_{p \leq T_b \leq T-p} h^{-1/2} \left|g_e\left(T_b\right)\right| \geq K\right),
\end{aligned} \tag{S.29}
$$

where $r_T = T \inf_{\left|T_b - T_b^0\right| > TK} h^{-1/2} r\left(T_b\right)$, which is positive and bounded away from zero by Lemma S.D.8. Thus, it is sufficient to verify that

$$
\sup_{p \leq T_b \leq T-p} h^{-1/2} \left|g_e\left(T_b\right)\right| = o_p\left(1\right). \tag{S.30}
$$

Consider the first term of $g_e\left(T_b\right)$:

$$
\begin{aligned}
& 2\delta_h' \left(Z_0' M Z_2\right) \left(Z_2' M Z_2\right)^{-1/2} \left(Z_2' M Z_2\right)^{-1/2} Z_2 M e \\
& \qquad \leq 2h^{1/4} \left(\delta^0\right)' \left(Z_0' M Z_2\right) \left(Z_2' M Z_2\right)^{-1/2} \left(Z_2' M Z_2\right)^{-1/2} Z_2 M e.
\end{aligned} \tag{S.31}
$$

For any $1 \leq j \leq p$, $\left(Z_2 \widetilde{e}\right)_{j,1} / \sqrt{h} = O_p\left(1\right)$ by Theorem S.D.5, and similarly, for any $1 \leq i \leq q+p$, $\left(X \widetilde{e}\right)_i / \sqrt{h} = O_p\left(1\right)$. Furthermore, from Lemma S.D.3 we also have that $Z_2' M Z_2$ and $Z_0' M Z_2$ are $O_p\left(1\right)$. Therefore, the supremum of $\left(Z_0' M Z_2\right) \left(Z_2' M Z_2\right)^{-1/2}$ over all T_b is such that

$$
\sup_{T_b} \left(Z_0' M Z_2\right) \left(Z_2' M Z_2\right)^{-1} \left(Z_2' M Z_0\right) \leq Z_0' M Z_0 = O_p\left(1\right),
$$

by Lemma S.D.3. By Assumption 2.1-(iii), $\left(Z_2' M Z_2\right)^{-1/2} Z_2 M \widetilde{e}$ is $O_p\left(1\right) O_p(\sqrt{h})$ uniformly, which implies that (S.31) is $O_p(\sqrt{h})$ uniformly over $p \leq T_b \leq T-p$. In view of Assumption 4.1 [recall (4.1)], we need to study the behavior of $\left(X' e\right)_{j,1}$ for $1 \leq j \leq p+q$. Note first that $|\widehat{\lambda}_b - \lambda_0| > K$ or $N > |\widehat{N}_b - N_b^0| > KN$. Then, by Itô formula, proceeding as in the proof of Lemma S.D.2, we have a standard result for the local volatility of a continuous Itô semimartingale; namely that for some $A > 0$ (recall the condition $T^{1-\kappa}\epsilon \rightarrow B > 0$),

$$
\left\|\mathbb{E}\left(\frac{1}{\epsilon} \sum_{T_b^0 - \lfloor T^\kappa \rfloor}^{T_b^0} x_{kh} \widetilde{e}_{kh} - \frac{1}{\epsilon} \int_{N_b^0 - \epsilon}^{N_b^0} \Sigma_{Xe,s} ds | \mathscr{F}_{\left(T_b^0 - 1\right)h}\right)\right\| \leq A h^{1/2}.
$$

From Assumption 4.2, since $\Sigma_{Xe,t} = 0$ for all $t \geq 0$, we have

$$
\begin{aligned}
X' e & = \sum_{k=1}^{T_b^0 - \lfloor T^\kappa \rfloor} x_{kh} \widetilde{e}_{kh} + h^{-1/4} \sum_{k = T_b^0 - \lfloor T^\kappa \rfloor + 1}^{T_b^0 + \lfloor T^\kappa \rfloor} x_{kh} \widetilde{e}_{kh} + \sum_{k = T_b^0 + \lfloor T^\kappa \rfloor + 1}^{T} x_{kh} \widetilde{e}_{kh} \\
& = O_p\left(h^{1/2}\right) + h^{-1/4} O_p\left(h^{1-\kappa+1/2}\right) + O_p\left(h^{1/2}\right) = O_p\left(h^{1/2}\right).
\end{aligned} \tag{S.32}
$$

The same bound applies to $Z_2'e$ and $Z_0'e$. Thus, (S.31) is such that

$$\begin{aligned}2h^{-1/2}h^{1/4}\left(\delta^0\right)'\left(Z_0'MZ_2\right)\left(Z_2'MZ_2\right)^{-1/2}\left(Z_2'MZ_2\right)^{-1/2}Z_2Me \\ = 2h^{-1/2}h^{1/4}\left\|\delta^0\right\|O_p\left(1\right)O_p\left(h^{1/2}\right)=O_p\left(1\right)O_p\left(h^{1/4}\right).\end{aligned}$$

As for the second term of (S.27),

$$h^{-1/2}\delta_h'\left(Z_0'Me\right)=2h^{-1/4}\left(\delta^0\right)'\left(Z_0'Me\right)=Ch^{-1/4}O_p\left(h^{1/2}\right)=CO_p\left(h^{1/4}\right),$$

using (S.32). Again using (S.32), the first term in (S.28) is such that, uniformly in T_b,

$$\begin{aligned}&h^{-1/2}e'MZ_2\left(Z_2'MZ_2\right)^{-1}Z_2Me \\ &= h^{-1/2}BO_p\left(h^{1/2}\right)O_p\left(1\right)O_p\left(h^{1/2}\right)=O_p\left(h^{1/2}\right).\end{aligned} \tag{S.33}$$

Similarly, the last term in (S.28) is $O_p(\sqrt{h})$. Therefore, combining these results, we have $h^{-1/2}\sup_{T_b}|g_e\left(T_b\right)|$ $= BO_p(h^{1/4})$, from which it follows that the right-hand side of (S.29) is weakly smaller than ε.

Lemma S.D.8. *For $B>0$, let $r_{B,h}=\inf_{|T_b-T_b^0|>TB}Th^{-1/2}r\left(T_b\right)$. There exists an $A>0$ such that for every $\varepsilon>0$, there exists a $B<\infty$ such that $P\left(r_{B,h}\geq A\right)\leq 1-\varepsilon$.*

Proof. Assume $N_b\leq N_b^0$, and observe that $r_T\geq r_{B,h}$ for an appropriately chosen B. From the first inequality result in Lemma S.D.1,

$$\begin{aligned}Th^{-1/2}r\left(T_b\right)&\geq Th^{-1/2}h^{1/2}\left(\delta^0\right)'R'\frac{X_\Delta'X_\Delta}{T_b^0-T_b}\left(X_2'X_2\right)^{-1}\left(X_0'X_0\right)R\delta^0 \\ &= \left(\delta^0\right)'R'\left(X_\Delta'X_\Delta/\left(N_b^0-N_b\right)\right)\left(X_2'X_2\right)^{-1}\left(X_0'X_0\right)R\delta^0.\end{aligned}$$

Note that $B<h\left(T_b^0-T_b\right)<N$. Then

$$Th^{-1/2}r\left(T_b\right)\geq\left(\delta^0\right)'R'\left(X_\Delta'X_\Delta/N\right)\left(X_2'X_2\right)^{-1}\left(X_0'X_0\right)R\delta^0>A,$$

by the same argument as in Lemma S.D.6. Following the same reasoning as in the proof of Lemma S.D.6, we can choose a $B>0$ such that $r_{B,h}=\inf_{|T_b-T_b^0|>TB}Th^{-1/2}r\left(T_b\right)$ satisfies $P(r_{B,h}\geq A)\leq 1-\varepsilon$. □

Proof of part (ii) of Proposition 4.1. Suppose $T_b<T_b^0$. Let

$$\mathbf{D}_{K,T}=\left\{T_b:\ N\eta\leq N_b\leq N\left(1-\eta\right),\ \left|N_b-N_b^0\right|>K\left(T^{1-\kappa}\right)^{-1}\right\}.$$

It is enough to show that $P(\sup_{T_b\in\mathbf{D}_{K,T}}Q_T\left(T_b\right)\geq Q_T(T_b^0))<\varepsilon$. The difficulty is again to control the estimates that depend on $|N_b-N_b^0|$. We shall show that

$$P\left(\sup_{T_b\in\mathbf{D}_{K,T}}h^{-3/2}\frac{g_e\left(T_b,\,\delta_h\right)}{\left|T_b-T_b^0\right|}\geq\inf_{T_b\in\mathbf{D}_{K,T}}h^{-3/2}r\left(T_b\right)\right)<\varepsilon.$$

By Lemma S.D.1,

$$\inf_{T_b\in\mathbf{D}_{K,T}}r\left(T_b\right)\geq\inf_{T_b\in\mathbf{D}_{K,T}}\delta_h'R'\frac{X_\Delta'X_\Delta}{T_b^0-T_b}\left(X_2'X_2\right)^{-1}\left(X_0'X_0\right)R\delta_h,$$

and, since $|T_b - T_b^0| > KT^{\kappa}$, it is important to consider $X'_\Delta X_\Delta = \sum_{k=T_b+1}^{T_b^0} x_{kh} x'_{kh}$. We shall apply asymptotic results for the local approximation of the covariation between processes. Consider

$$\frac{X'_\Delta X_\Delta}{h\left(T_b^0 - T_b\right)} = \frac{1}{h\left(T_b^0 - T_b\right)} \sum_{k=T_b+1}^{T_b^0} x_{kh} x'_{kh}.$$

By Theorem 9.3.2-(i) in Jacod and Protter (2012), as $h \downarrow 0$

$$\frac{1}{h\left(T_b^0 - T_b\right)} \sum_{k=T_b+1}^{T_b^0} x_{kh} x'_{kh} \xrightarrow{P} \Sigma_{XX, N_b^0}, \tag{S.34}$$

since $|N_b - N_b^0|$ shrinks at a rate no faster than $Kh^{1-\kappa}$ and $1/Kh^{1-\kappa} \to \infty$. By Lemma S.D.2 this approximation is uniform, establishing that

$$\begin{aligned} h^{-3/2} & \inf_{T_b \in \mathbf{D}_{K,T}} (\delta_h)' R' \frac{X'_\Delta X_\Delta}{T_b^0 - T_b} \left(X'_2 X_2\right)^{-1} \left(X'_0 X_0\right) R \delta_h \\ & = \inf_{T_b \in \mathbf{D}_{K,T}} \left(\delta^0\right)' R' \frac{X'_\Delta X_\Delta}{h\left(T_b^0 - T_b\right)} \left(X'_2 X_2\right)^{-1} \left(X'_0 X_0\right) R \delta^0, \end{aligned}$$

is bounded away from zero. Thus, it is sufficient to show that

$$P\left(\sup_{T_b \in \mathbf{D}_{K,T}} h^{-3/2} \frac{g_e\left(T_b,\, \delta_h\right)}{|T_b - T_b^0|} \geq B\right) < \varepsilon, \tag{S.35}$$

for some $B > 0$. Consider the terms of $g_e\left(T_b\right)$ in (S.28). Using $Z_2 = Z_0 \pm Z_\Delta$, we deduce for the first term that

$$\begin{aligned} & \delta'_h \left(Z'_0 M Z_2\right) \left(Z'_2 M Z_2\right)^{-1} Z_2 M e \\ & \quad = \delta'_h \left(\left(Z'_2 \pm Z_\Delta\right) M Z_2\right) \left(Z'_2 M Z_2\right)^{-1} Z_2 M e \\ & \quad = \delta'_h Z'_0 M e \pm \delta'_h Z'_\Delta M e \pm \delta'_h \left(Z'_\Delta M Z_2\right) \left(Z'_2 M Z_2\right)^{-1} Z_2 M e. \end{aligned} \tag{S.36}$$

First, we can apply Lemma S.D.3 [(vi)-(viii)], together with Assumption 2.1-(iii), to terms not involving $|N_b - N_b^0|$. The third term is such that

$$K^{-1} h^{-(1-\kappa)} \left(Z'_\Delta M Z_2\right) = \frac{Z'_\Delta Z_2}{Kh^{1-\kappa}} - \frac{Z'_\Delta X_\Delta}{Kh^{1-\kappa}} \left(X' X\right)^{-1} X' Z_2. \tag{S.37}$$

Consider $Z'_\Delta Z_\Delta$ (the argument for $Z'_\Delta X_\Delta$ is analogous). By Lemma S.D.2, $Z'_\Delta Z_\Delta / Kh^{1-\kappa}$ uniformly approximates the moving average of $\Sigma_{ZZ,t}$ over $(N_b^0 - KT^{\kappa}h,\, N_b^0]$. Hence, as $h \downarrow 0$,

$$Z'_\Delta Z_\Delta / Kh^{1-\kappa} = B O_p\left(1\right), \tag{S.38}$$

for some $B > 0$, uniformly in T_b. The second term in (S.37) is thus also $O_p\left(1\right)$ uniformly, using Lemma S.D.3. Then, using (S.32) and (S.37) into the third term of (S.36), we have

$$\begin{aligned} & \frac{1}{K} h^{-(1-\kappa)-1/2} \left(\delta_h\right)' \left(Z'_\Delta M Z_2\right) \left(Z'_2 M Z_2\right)^{-1} Z_2 M e \\ & \quad \leq \frac{1}{K} h^{-1/4} \left(\delta^0\right)' \left(\frac{Z'_\Delta M Z_2}{h^{1-\kappa}}\right) \left(Z'_2 M Z_2\right)^{-1} Z_2 M e \end{aligned} \tag{S.39}$$

$$\leq h^{-1/4} \frac{Z_{\Delta}' M Z_2}{K h^{1-\kappa}} O_p(1) O_p\left(h^{1/2}\right) \leq O_p\left(h^{1/4}\right),$$

where $(Z_2' M Z_2)^{-1} = O_p(1)$. So the right-and side of (S.39) is less than $\varepsilon/4$ in probability. Therefore, for the second term of (S.36),

$$\begin{aligned}
&K^{-1} h^{-(1-\kappa)-1/2} \delta_h' Z_{\Delta}' M e \\
&\quad = \frac{h^{-1/2}}{K h^{1-\kappa}} \delta_h' \sum_{k=T_b+1}^{T_b^0} z_{kh} e_{kh} - \frac{h^{-1/2}}{h^{1-\kappa}} \delta_h' \left(\sum_{k=T_b+1}^{T_b^0} z_{kh} x_{kh}' \right) (X'X)^{-1} (X'e) \\
&\quad \leq \frac{h^{-1/2}}{K h^{1-\kappa}} \delta_h' \sum_{k=T_b+1}^{T_b^0} z_{kh} e_{kh} - B \frac{1}{K} \frac{h^{-1/4}}{h^{1-\kappa}} \left(\delta^0\right) \left(\sum_{k=T_b+1}^{T_b^0} z_{kh} x_{kh}' \right) (X'X)^{-1} (X'e) \\
&\quad \leq \frac{h^{-1/2}}{K h^{1-\kappa}} \delta_h' \sum_{k=T_b+1}^{T_b^0} z_{kh} e_{kh} - h^{-1/4} O_p(1) O_p\left(h^{1/2}\right).
\end{aligned} \tag{S.40}$$

Thus, using (S.36), (S.27) is such that

$$\begin{aligned}
&2\delta_h' Z_0' M e \pm 2\delta_h' Z_{\Delta}' M e \pm 2\delta_h' (Z_{\Delta}' M Z_2) (Z_2' M Z_2)^{-1} Z_2 M e - 2\delta_h' (Z_0' M e) \\
&\quad = 2\delta_h' Z_{\Delta}' M e \pm 2\delta_h' (Z_{\Delta}' M Z_2) (Z_2' M Z_2)^{-1} Z_2 M e \\
&\quad \leq \frac{h^{-1/2}}{K h^{1-\kappa}} \left(\delta^0\right)' \sum_{k=T_b+1}^{T_b^0} z_{kh} \widetilde{e}_{kh} - h^{-1/4} O_p(1) O_p\left(h^{1/2}\right) + O_p\left(h^{-1/4}\right),
\end{aligned}$$

in view of (S.39) and (S.40). Next, consider (S.28). We can use the decomposition $Z_2 = Z_0 \pm Z_{\Delta}$ and show that all terms involving the matrix Z_{Δ} are negligible. To see this, consider the first term when multiplied by $K^{-1} h^{-(3/2-\kappa)}$, then

$$\begin{aligned}
&K^{-1} h^{-(3/2-\kappa)} e' M Z_2 (Z_2' M Z_2)^{-1} Z_2 M e \\
&\quad = K^{-1} h^{-(3/2-\kappa)} e' M Z_0 (Z_2' M Z_2)^{-1} Z_2 M e \\
&\qquad \pm K^{-1} h^{-(3/2-\kappa)} e' M Z_{\Delta} (Z_2' M Z_2)^{-1} Z_2 M e.
\end{aligned} \tag{S.41}$$

By the same argument as in (S.32), $Z_2' M e = O_p(h^{1/2})$. Using the Burkhölder-Davis-Gundy inequality, the estimates for the local volatility of continuous Itô semimartingales yield

$$\begin{aligned}
\widetilde{e}' M Z_{\Delta} &= \widetilde{e}' Z_{\Delta} - \widetilde{e}' X (X'X)^{-1} X' Z_{\Delta} \\
&= O_p\left(K h^{1/2+1-\kappa}\right) - O_p\left(h^{1/2}\right) O_p(1) O_p\left(K h^{1-\kappa}\right).
\end{aligned}$$

Thus, the second term in (S.41) is such that

$$\begin{aligned}
&K^{-1} h^{-(3/2-\kappa)} \widetilde{e}' M Z_{\Delta} (Z_2' M Z_2)^{-1} Z_2 M e \\
&\quad = B\left(K^{-1} h^{-(3/2-\kappa)}\right) O_p\left(K h^{1-\kappa+1/2}\right) O_p(1) O_p\left(h^{1/2}\right) \\
&\quad = B O_p\left(h^{1/2}\right).
\end{aligned} \tag{S.42}$$

Next, let us consider (S.28). The key here is to recognize that, on $\mathbf{D}_{K,T}$, T_b and T_b^0 lies on the same window with right-hand point N_b^0. Thus the difference between the two terms in (S.28) is asymptotically negligible. First, note that using (S.32),

$$\widetilde{e}'MZ_0\left(Z_0'MZ_0\right)^{-1}Z_0M\widetilde{e}=O_p\left(h^{1/2}\right)O_p\left(1\right)O_p\left(h^{1/2}\right)=O_p\left(h\right).$$

Applying $Z_0=Z_2\pm Z_\Delta$ repeatedly in (S.41), and noting that the cross-product terms involving Z_Δ are $o_p\left(1\right)$ by the same reasoning as in (S.42), we obtain that the difference between the first and second term of (S.28) is negligible. The more intricate step is the one arising from

$$\begin{aligned}&e'MZ_0\left(Z_0'MZ_2\pm Z_\Delta'MZ_2\right)^{-1}Z_0'Me-e'MZ_0\left(Z_0'MZ_0\right)^{-1}Z_0'Me\\&=e'MZ_0\left[\left(Z_0'MZ_2\pm Z_\Delta'MZ_2\right)^{-1}-\left(Z_0'MZ_0\right)^{-1}\right]Z_0'Me.\end{aligned}$$

On $\mathbf{D}_{K,T}$, $\left|N_b-N_b^0\right|=O_p(Kh^{1-\kappa})$, and so each term involving Z_Δ is of higher order. By using the continuity of probability limits, the matrix in square brackets goes to zero at rate $h^{1-\kappa}$. Then, this expression when multiplied by $h^{-(3/2-\kappa)}K^{-1}$, and after using the same rearrangements as above, can be shown to satisfy [recall also (S.32)]

$$\begin{aligned}&h^{-(3/2-\kappa)}K^{-1}e'MZ_0\left[\left(Z_0'MZ_2\pm Z_\Delta'MZ_2\right)^{-1}-\left(Z_0'MZ_0\right)^{-1}\right]Z_0'Me\\&=h^{-(3/2-\kappa)}K^{-1}O_p\left(h\right)\left[\left(Z_0'MZ_2\pm Z_\Delta'MZ_2\right)^{-1}-\left(Z_0'MZ_0\right)^{-1}\right]\\&=h^{-(3/2-\kappa)}K^{-1}O_p\left(h\right)\\&\quad\times\left[\left(Z_0'MZ_0\pm Z_0'MZ_\Delta'\pm Z_\Delta'MZ_2\right)^{-1}-\left(Z_0'MZ_0\right)^{-1}\right]\\&=h^{-(3/2-\kappa)}K^{-1}O_p\left(h\right)o_p\left(h^{1-\kappa}\right)=O_p\left(h^{1/2}\right)o_p\left(1\right).\end{aligned}$$

Therefore, (S.28) is stochastically small uniformly in $T_b\in\mathbf{D}_{K,T}$ when T is large. Altogether, we have

$$\begin{aligned}h^{-1/2}\frac{g_e\left(T_b\right)}{\left|T_b-T_b^0\right|}\leq 2\frac{h^{-1/2}}{Kh^{1-\kappa}}\delta_h'\sum_{k=T_b+1}^{T_b^0}z_{kh}e_{kh}\\-h^{-1/4}O_p\left(1\right)O_p\left(h^{1/2}\right)+O_p\left(h^{-1/4}\right).\end{aligned}$$

Thus, it remains to find a bound for the first term above. By Itô's formula, standard estimates for the local volatility of continuous Itô semimartingales yield for every T_b,

$$\mathbb{E}\left(\left\|\widehat{\Sigma}_{Ze}\left(T_2,\,T_b^0\right)-\overline{\Sigma}_{Ze}\left(T_2,\,T_b^0\right)\right\|\,|\,\mathscr{F}_{T_bh}\right)\leq Bh^{1/2},\tag{S.43}$$

for some $B>0$. Let $R_{1,h}=\sum_{k=T_b^0-(B+1)\lfloor T^\kappa\rfloor+1}^{T_b^0}z_{kh}\widetilde{e}_{kh}$, $R_{2,h}\left(T_b\right)=\sum_{k=T_b+1}^{T_2^0-(B+1)\lfloor T^\kappa\rfloor}z_{kh}e_{kh}$ and note that $\sum_{k=T_2+1}^{T_2^0}z_{kh}e_{kh}=R_{1,h}+R_{2,h}\left(T_b\right)$. Then, for any $C>0$,

$$\begin{aligned}&P\left(\sup_{T_b<T_b^0-KT^\kappa}2\frac{h^{-1/2}}{Kh^{1-\kappa}}\delta_h'\left\|\sum_{k=T_b+1}^{T_b^0}z_{kh}e_{kh}\right\|\geq C\right)\\&=P\left(\sup_{T_b<T_b^0-KT^\kappa}\frac{h^{-1/2}}{Kh^{1-\kappa}}\delta_h'\left\|R_{1,h}+R_{2,h}\left(T_b\right)\right\|\geq 2^{-1}C\right)\\&\leq P\left(\frac{1}{Kh^{1-\kappa}}\left\|R_{1,h}\right\|>4^{-1}C\left\|\delta^0\right\|^{-1}h^{1/2}\right)\end{aligned}\tag{S.44}$$

$$+P\left(\sup_{T_b<T_b^0-KT^\kappa}\frac{K^{-1}}{h^{1-\kappa}}\left\|R_{2,h}\left(T_b\right)\right\|>4^{-1}C\left\|\delta^0\right\|^{-1}h^{1/4}\right).$$

Consider first the second probability. By Markov's inequality,

$$\begin{aligned}
&P\left(\sup_{T_b<T_b^0-KT^\kappa}\frac{1}{Kh^{1-\kappa}}\left\|R_{2,h}\left(T_b\right)\right\|>4^{-1}C\left\|\delta^0\right\|^{-1}h^{1/4}\right)\\
&\leq P\left(\sup_{T_b<T_b^0-KT^\kappa}\left\|\frac{1}{Kh^{1-\kappa}}R_{2,h}\left(T_b\right)\right\|>4^{-1}C\left\|\delta^0\right\|^{-1}h^{1/4}\right)\\
&\leq \left(K/B\right)T^\kappa P\left(\left\|\frac{1}{Kh^{1-\kappa}}R_{2,h}\left(T_b\right)\right\|>4^{-1}C\left\|\delta^0\right\|^{-1}h^{1/4}\right)\\
&\leq \frac{\left(4\left(B+1\right)\left\|\delta^0\right\|\right)^r}{C^r}h^{-r/4}\frac{K}{B}T^\kappa\mathbb{E}\left(\left|\frac{1}{\left(B+1\right)Kh^{1-\kappa}}\left\|R_{2,h}\left(T_b\right)\right\|\right|^r\right)\\
&\leq C_r\left(B+1\right)B^{-1}\left\|\delta^0\right\|^r h^{-r/4}T^\kappa h^{r/2}\leq C_r\left\|\delta^0\right\|^r h^{r/2-\kappa-r/4}\to 0,
\end{aligned}$$

for a sufficiently large $r>0$. We now turn to $R_{1,h}$. We have,

$$\begin{aligned}
&P\left(\frac{1}{Kh^{1-\kappa}}\left\|R_{1,h}\right\|>2^{-1}C\left\|\delta^0\right\|^{-1}h^{1/2}\right)\\
&\leq P\Bigg(\frac{\left(B+1\right)}{K}\left\|\left(B+1\right)^{-1}h^{-(1-\kappa)}\sum_{k=T_b^0-(B+1)\lfloor T^\kappa\rfloor+1}^{T_b^0}z_{kh}\widetilde{e}_{kh}\right\|\\
&\quad>\frac{C}{4}\left\|\delta^0\right\|^{-1}h^{1/2}\Bigg)\\
&\leq P\left(\left(B+1\right)K^{-1}O_{\mathbb{P}}\left(1\right)>4^{-1}C\left\|\delta^0\right\|^{-1}\right)\to 0,
\end{aligned}$$

by choosing K large enough where we have used (S.43). Altogether, the right-hand side of (S.44) is less than ε, which concludes the proof. □

Proof of part (iii) of Proposition 4.1. Observe that Lemma S.D.7 applies under this setting. Then, we have,

$$\sqrt{T}\begin{bmatrix}\widehat{\beta}-\beta_0\\ \widehat{\delta}-\delta_h\end{bmatrix}=\begin{bmatrix}X'X & X'\widehat{Z}_0\\ \widehat{Z}_0'X & \widehat{Z}_0'\widehat{Z}_0\end{bmatrix}^{-1}\sqrt{T}\begin{bmatrix}X'e+X'\left(Z_0-\widehat{Z}_0\right)\delta_h\\ \widehat{Z}_0'e+\widehat{Z}_0'\left(Z_0-\widehat{Z}_0\right)\delta_h\end{bmatrix},$$

so that we need to show that

$$\begin{bmatrix}X'X & X'\widehat{Z}_0\\ \widehat{Z}_0'X & \widehat{Z}_0'\widehat{Z}_0\end{bmatrix}^{-1}\frac{1}{h^{1/2}}X'\left(Z_0-\widehat{Z}_0\right)\delta_h\xrightarrow{P}0,$$

and the limiting distribution of $X'e/h^{1/2}$ is Gaussian. The first claim can be proved in a manner analogous to that in the proof of Proposition 3.3. For the second claim, we have the following decomposition from (S.32),

$$X'e=\sum_{k=1}^{T_b^0-\lfloor T^\kappa\rfloor}x_{kh}\widetilde{e}_{kh}+h^{-1/4}\sum_{T_b^0-\lfloor T^\kappa\rfloor+1}^{T_b^0+\lfloor T^\kappa\rfloor}x_{kh}\widetilde{e}_{kh}+\sum_{k=T_b^0+\lfloor T^\kappa\rfloor+1}^{T}x_{kh}\widetilde{e}_{kh}$$

$$\triangleq R_{1,h} + R_{2,h} + R_{3,h}.$$

By Theorem S.D.5, $h^{-1/2} R_{1,h} \overset{\mathcal{L}\text{-s}}{\to} \mathscr{M}\mathscr{N}(0, V_1)$, where $V_1 \triangleq \lim_{T \to \infty} T \sum_{k=1}^{T_b^0 - \lfloor T^\kappa \rfloor} \mathbb{E}(x_{kh} x'_{kh} \widetilde{e}^2_{kh})$. Similarly, $h^{-1/2} R_{3,h} \overset{\mathcal{L}\text{-s}}{\to} \mathscr{M}\mathscr{N}(0, V_3)$, where $V_3 \triangleq \lim_{T \to \infty} T \sum_{k=T_b^0 + \lfloor T^\kappa \rfloor + 1}^{T} \mathbb{E}(x_{kh} x'_{kh} \widetilde{e}^2_{kh})$. If $\kappa \in (0, 1/4)$, $h^{-(1-\kappa)} \sum_{T_b^0 - \lfloor T^\kappa \rfloor + 1}^{T_b^0 + \lfloor T^\kappa \rfloor} x_{kh} \widetilde{e}_{kh} \overset{P}{\to} \Sigma_{Xe, N_b^0}$ by Theorem 9.3.2 in Jacod and Protter (2012) and so $h^{-1/2} R_{2,h} = h^{-3/4} \sum_{T_b^0 - \lfloor T^\kappa \rfloor}^{T_b^0 + \lfloor T^\kappa \rfloor} x_{kh} \widetilde{e}_{kh} \overset{P}{\to} 0$. If $\kappa = 1/4$, then $h^{-1/2} R_{2,h} \to \Sigma_{Xe, N_b^0}$ in probability again by Theorem 9.3.2 in Jacod and Protter (2012). Since by Assumption 4.2 $\Sigma_{Xe,t} = 0$ for all $t \geq 0$, whenever $\kappa \in (0, 1/4]$, $X'e/h^{1/2}$ is asymptotically normally distributed. The rest of the proof is simple and follows the same steps as in Proposition 3.3. □

S.D.4.6 Proof of Proposition 4.2

Proof. By Lemma 4.1,

$$Q_T(T_b) - Q_T\left(T_b^0\right) = -\delta'_h\left(Z'_\Delta Z_\Delta\right)\delta_h \pm 2\delta'_h\left(Z'_\Delta e\right) + o_p\left(h^{3/2-\kappa}\right).$$

Divide both sides by h to yield,

$$\begin{aligned} h^{-1}\left(Q_T(T_b) - Q_T\left(T_b^0\right)\right) = -h^{1/2}\left(\delta^0\right)'\left(\frac{Z'_\Delta}{\sqrt{h}}\frac{Z_\Delta}{\sqrt{h}}\right)\delta^0 \\ \pm 2\left(\delta^0\right)'\left(\frac{Z'_\Delta}{\sqrt{h}}\frac{\widetilde{e}}{\sqrt{h}}\right) + o_p\left(h^{1/2-\kappa}\right). \end{aligned}$$

Note that $z_{kh}/\sqrt{h} \sim i.n.d.\, \mathscr{N}(0, \Sigma_{kh})$ and $\widetilde{e}_{kh}/\sqrt{h} \sim i.n.d.\, \mathscr{N}(0, \sigma^2_{e,kh})$. Thus,

$$\begin{aligned} & h^{-1+\kappa/2}\left(Q_T(T_b) - Q_T\left(T_b^0\right)\right) \\ & = -\frac{h^{1/2}}{\sqrt{T^\kappa}}\left(\delta^0\right)'\left(\frac{Z'_\Delta}{\sqrt{h}}\frac{Z_\Delta}{\sqrt{h}}\right)\delta^0 \\ & \quad \pm \frac{2}{\sqrt{T^\kappa}}\left(\delta^0\right)'\left(\frac{Z'_\Delta}{\sqrt{h}}\frac{\widetilde{e}}{\sqrt{h}}\right) + o_p\left(h^{1/2-\kappa/2}\right) \\ & = O_p\left(h^{1/2}\right) \pm \frac{2}{\sqrt{T^\kappa}}\left(\delta^0\right)'\left(\frac{Z'_\Delta}{\sqrt{h}}\frac{\widetilde{e}}{\sqrt{h}}\right) + o_p\left(h^{1/2-\kappa}\right). \end{aligned}$$

Also $T_b = T_b^0 + \lfloor vT^\kappa \rfloor$, and

$$h^{-1+\kappa/2}\left(Q_T(T_b) - Q_T\left(T_b^0\right)\right) \Rightarrow 2\left(\delta^0\right)'\mathscr{W}(v).$$

The continuous mapping theorem then yields the desired result. □

S.D.4.7 Proof of Lemma 4.1

First, we begin with the following simple identity. Throughout the proof, B is a generic constant which may change from line to line.

Lemma S.D.9. *The following identity holds*

$$\begin{aligned} & (\delta_h)'\left\{Z'_0 M Z_0 - \left(Z'_0 M Z_2\right)\left(Z'_2 M Z_2\right)^{-1}\left(Z'_2 M Z_0\right)\right\}\delta_h \\ & \quad = (\delta_h)'\left\{Z'_\Delta M Z_\Delta - \left(Z'_\Delta M Z_2\right)\left(Z'_2 M Z_2\right)^{-1}\left(Z'_2 M Z_\Delta\right)\right\}\delta_h. \end{aligned}$$

Proof. The proof follows simply from the fact that $Z_0' M Z_2 = Z_2' M Z_2 \pm Z_\Delta' M Z_2$ and so

$$\begin{aligned}
&(\delta_h)' \left\{ Z_0' M Z_0 - \left(Z_2' M Z_2 \pm Z_\Delta' M Z_2 \right) \left(Z_2' M Z_2 \right)^{-1} \left(Z_2' M Z_0 \right) \right\} \delta_h \\
&\quad = (\delta_h)' \{ Z_\Delta' M Z_0 - \left(Z_\Delta' M Z_2 \right) \left(Z_2' M Z_2 \right)^{-1} \left(Z_2' M Z_2 \right) \\
&\qquad - \left(Z_\Delta' M Z_2 \right) \left(Z_2' M Z_2 \right)^{-1} \left(Z_2' M Z_\Delta \right) \} \delta_h \\
&\quad = (\delta_h)' \left\{ Z_\Delta' M Z_\Delta - \left(Z_\Delta' M Z_2 \right) \left(Z_2' M Z_2 \right)^{-1} \left(Z_2' M Z_\Delta \right) \right\} \delta_h . \square
\end{aligned}$$

Proof of Lemma 4.1. By the definition of $Q_T(T_b) - Q_T(T_0)$ and Lemma S.D.9,

$$\begin{aligned}
&Q_T(T_b) - Q_T(T_0) \\
&= -\delta_h' \left\{ Z_\Delta' M Z_\Delta - \left(Z_\Delta' M Z_2 \right) \left(Z_2' M Z_2 \right)^{-1} \left(Z_2' M Z_\Delta \right) \right\} \delta_h + g_e(T_b, \delta_h),
\end{aligned} \tag{S.45}$$

where

$$g_e(T_b, \delta_h) = 2\delta_h' \left(Z_0' M Z_2 \right) \left(Z_2' M Z_2 \right)^{-1} Z_2 M e - 2\delta_h' \left(Z_0' M e \right) \tag{S.46}$$

$$+ e' M Z_2 \left(Z_2' M Z_2 \right)^{-1} Z_2 M e - e' M Z_0 \left(Z_0' M Z_0 \right)^{-1} Z_0' M e. \tag{S.47}$$

Recall that $N_b(u) \in \mathcal{D}(C)$ implies $T_b(u) = T_b^0 + uT^\kappa$, $u \in [-C, C]$. We consider the case $u \leq 0$. By Theorem 9.3.2-(i) in Jacod and Protter (2012) combined with Lemma S.D.2, we have uniformly in u as $h \downarrow 0$

$$\frac{1}{h^{1-\kappa}} \sum_{k=T_b^0 + uT^\kappa}^{T_b^0} x_{kh} x_{kh}' \xrightarrow{P} \Sigma_{XX, N_b^0}. \tag{S.48}$$

Since $Z_\Delta' X = Z_\Delta' X_\Delta$, we will use this result also for $Z_\Delta' X / h^{1-\kappa}$. With the notation of Section S.D.4.1 [recall (S.6)], by the Burkhölder-Davis-Gundy inequality, we have that standard estimates for the local volatility, is such that

$$\left\| \mathbb{E} \left(\widehat{\Sigma}_{ZX} \left(T_b, T_b^0 \right) - \Sigma_{ZX, (T_b^0 - 1)h} | \mathscr{F}_{(T_b^0 - 1)h} \right) \right\| \leq B h^{1/2}. \tag{S.49}$$

Equation (S.48)-(S.49) can be used to yield, uniformly in T_b,

$$\psi_h^{-1} Z_\Delta' X \left(X' X \right)^{-1} X' Z_\Delta = O_p(1) X' Z_\Delta, \tag{S.50}$$

and

$$Z_\Delta' M Z_2 = Z_\Delta' Z_\Delta - Z_\Delta' X \left(X' X \right)^{-1} X' Z_2 = O_p(\psi_h) - O_p(\psi_h) O_p(1) O_p(1). \tag{S.51}$$

Now, expand the first term of (S.45),

$$\delta_h' Z_\Delta' M Z_\Delta \delta_h = \delta_h' Z_\Delta' Z_\Delta \delta_h - \delta_h' Z_\Delta' X \left(X' X \right)^{-1} X' Z_\Delta \delta_h. \tag{S.52}$$

By Lemma S.D.3, $\left(X' X \right)^{-1} = O_p(1)$ and recall that $\delta_h = h^{1/4} \delta^0$. Then,

$$\psi_h^{-1} \delta_h' Z_\Delta' M Z_\Delta \delta_h = \psi_h^{-1} \delta_h' Z_\Delta' Z_\Delta \delta_h - \psi_h^{-1} \delta_h' Z_\Delta' X \left(X' X \right)^{-1} X' Z_\Delta \delta_h. \tag{S.53}$$

By (S.50), the second term above is such that

$$\left\|\delta^0\right\|^2 h^{1/2}\frac{Z_\Delta' X}{\psi_h}\left(X'X\right)^{-1}X'Z_\Delta=\left\|\delta^0\right\|^2 h^{1/2}O_p\left(1\right)X'Z_\Delta, \tag{S.54}$$

uniformly in $T_b\left(u\right)$. Therefore,

$$\psi_h^{-1}\delta_h' Z_\Delta' M Z_\Delta \delta_h=\psi_h^{-1}\delta_h' Z_\Delta' Z_\Delta \delta_h-\left\|\delta^0\right\|^2 h^{1/2}O_p\left(1\right)O_p\left(\psi_h\right). \tag{S.55}$$

In the last equality the second term of $\delta' Z_\Delta' M Z_\Delta \delta$ is always of higher order. This suggests that the term involving regressors whose parameters are allowed to shift plays a primary role in the asymptotic analysis. The second term is a complicated function of cross products of all regressors around the time of the change. Because of the fast rate of convergence, these high order product estimates around the break date will be negligible. We use this result repeatedly in the derivations that follow. The second term of (S.45) when multiplied by ψ_h^{-1} is, uniformly in $T_b\left(u\right)$,

$$\psi_h^{-1}\delta_h\left(Z_\Delta' M Z_2\right)\left(Z_2' M Z_2\right)^{-1}\left(Z_2' M Z_\Delta\right)\delta_h'=\left\|\delta^0\right\|^2 h^{1/2}O_p\left(1\right)O_p\left(1\right)O_p\left(\psi_h\right),$$

where we have used the fact that $Z_\Delta' M Z_2/\psi_h=O_p\left(1\right)$ [cf. (S.51)]. Hence, the second term of (S.45), when multiplied by ψ_h^{-1}, is $O_p\left(h^{3/2-\kappa}\right)$ uniformly in T_b. Finally, let us consider $g_e\left(T_b,\,\delta_h\right)$. Recall that $\widetilde{e}_{kh}$ defined in (4.1) is i.n.d. with zero mean and conditional variance $\sigma_{e,k-1}^2 h$. Upon applying the continuity of probability limits repeatedly, one first obtains that the difference between the two terms in (S.47) goes to zero at a fast enough rate as in the last step of the proof of Proposition 4.1-(ii). That is, for T large enough, we can find a c_T sufficiently small such that,

$$\psi_h^{-1}\left[e' M Z_2\left(Z_2' M Z_2\right)^{-1}Z_2 M e-e' M Z_0\left(Z_0' M Z_0\right)^{-1}Z_0' M e\right]=o_p\left(c_T h\right).$$

Next, consider the first two terms of $g_e\left(T_b,\,\delta_h\right)$. Using $Z_0' M Z_2=Z_2' M Z_2\pm Z_\Delta' M Z_2$, it is easy to show that

$$\begin{aligned}&2h^{1/4}\left(\delta^0\right)'\left(Z_0' M Z_2\right)\left(Z_2' M Z_2\right)^{-1}Z_2 M e-2h^{1/4}\left(\delta^0\right)'\left(Z_0' M e\right)\\&\qquad=2h^{1/4}\left(\delta^0\right)' Z'_\Delta M e\pm 2h^{1/4}\left(\delta^0\right)' Z'_\Delta M Z_2\left(Z_2' M Z_2\right)^{-1}Z_2' M e.\end{aligned} \tag{S.56}$$

Note that, uniformly in $T_b\left(u\right)$,

$$\begin{aligned}&\psi_h^{-1}h^{1/4}\left(\delta^0\right)' Z'_\Delta M Z_2\\&\quad=h^{1/4}\left(\delta^0\right)' Z'_\Delta Z_\Delta+\left(\delta^0\right)' h^{1/4}\frac{Z_\Delta' X}{\psi_h}\left(X'X\right)^{-1}X'Z_2\\&\quad=h^{1/4}\left(\delta^0\right)'\frac{Z_\Delta' Z_\Delta}{\psi_h}+\left(\delta^0\right)' h^{1/4}O_p\left(1\right)\\&\quad=h^{1/4}\left\|\delta^0\right\|O_p\left(1\right)+\left\|\delta^0\right\|h^{1/4}O_p\left(1\right),\end{aligned}$$

where we have used (S.48) and the fact that $\left(X'X\right)^{-1}$ and $X'Z_2$ are each $O_p\left(1\right)$. Recall the decomposition in (S.32):

$$X'e=O_p\left(h^{1-\kappa+1/4}\right)+O_p\left(h^{1/2}\right). \tag{S.57}$$

Thus, the last term in (S.56) multiplied by ψ_h^{-1} is such that

$$\begin{aligned}&\psi_h^{-1} 2h^{1/4}\left(\delta^0\right)' Z'_{\Delta} M Z_2 \left(Z_2' M Z_2\right)^{-1} Z_2' M e\\ &\quad = h^{1/4}\left\|\delta^0\right\| O_p(1)\, O_p(1)\left[O_p\left(h^{1-\kappa+1/4}\right) + O_p\left(h^{1/2}\right)\right]\\ &\quad = \left\|\delta^0\right\| h^{1/4} O_p(1)\, O_p\left(h^{1/2}\right) = \left\|\delta^0\right\| O_p\left(h^{3/4}\right).\end{aligned}$$

The first term of (S.56) can be decomposed further as follows

$$2h^{1/4}\left(\delta^0\right)' Z'_{\Delta} M e = 2h^{1/4}\left(\delta^0\right)' Z'_{\Delta} e - 2h^{1/4}\left(\delta^0\right) Z'_{\Delta} X \left(X'X\right)^{-1} X'e.$$

Then, when multiplied by ψ_h^{-1}, the second term above is such that, uniformly in T_b,

$$\begin{aligned}&h^{1/4}\left(\delta^0\right)' \left(Z'_{\Delta} X/\psi_h\right)\left(X'X\right)^{-1} X'e\\ &\quad = h^{1/4}\left(\delta^0\right)' O_p(1)\, O_p(1)\left[O_p\left(h^{1-\kappa+1/4}\right) + O_p\left(h^{1/2}\right)\right] = O_p\left(h^{3/4}\right),\end{aligned}$$

where we have used (S.48) and (S.57). Combining the last results, we have uniformly in T_b,

$$\begin{aligned}\psi_h^{-1} g_e\left(T_b,\, \delta_h\right) &= 2h^{1/4}\left(\delta^0\right)' \left(Z'_{\Delta} e/\psi_h\right)\\ &\quad + O_p\left(h^{3/4}\right) + \left\|\delta^0\right\| O_p\left(h^{3/4}\right) + o_p\left(c_T h\right),\end{aligned}$$

when T is large and c_T is a sufficiently small number. Then,

$$\begin{aligned}&\psi_h^{-1}\left(Q_T\left(T_b\right) - Q_T\left(T_b^0\right)\right)\\ &\quad = -\delta_h\left(Z'_{\Delta} Z_{\Delta}/\psi_h\right)\delta_h \pm 2\delta_h'\left(Z'_{\Delta} e/\psi_h\right)\\ &\qquad + O_p\left(h^{3/2-\kappa}\right) + O_p\left(h^{3/4}\right) + \left\|\delta^0\right\| O_p\left(h^{3/4}\right) + o_p\left(c_T h\right).\end{aligned}$$

Therefore, for T large enough,

$$\psi_h^{-1}\left(Q_T\left(T_b\right) - Q_T\left(T_b^0\right)\right) = -\delta_h\left(Z'_{\Delta} Z_{\Delta}/\psi_h\right)\delta_h \pm 2\delta_h'\left(Z'_{\Delta} e/\psi_h\right) + o_p\left(h^{1/2}\right).$$

This concludes the proof of Lemma 4.1. □

S.D.4.8 Proof of Theorem 4.1

Proof. Let us focus on the case $T_b(v) \le T_b^0$ (i.e., $v \le 0$). The change of time scale is obtained by a change in variable. On the old time scale, by Proposition 4.1, $N_b(v)$ varies on the time interval $[N_b^0 - |v|\, h^{1-\kappa},\, N_b^0 + |v|\, h^{1-\kappa}]$ with $v \in [-C,\, C]$. Lemma 4.1 shows that the conditional first moment of $Q_T\left(T_b(v)\right) - Q_T\left(T_b^0\right)$ is determined by that of $-\delta_h'\left(Z'_{\Delta} Z_{\Delta}\right)\delta_h \pm 2\delta_h'\left(Z'_{\Delta} e\right).$ Next, we rescale time with $s \mapsto t \triangleq \psi_h^{-1} s$ on $\mathcal{D}(C)$. This is achieved by rescaling the criterion function $Q_T\left(T_b(u)\right) - Q_T\left(T_b^0\right)$ by the factor ψ_h^{-1}. First, note that the processes Z_t and e_t^* [recall (2.3) and (4.1)] are rescaled as follows on $\mathcal{D}(C)$. Let $Z_{\psi,s} \triangleq \psi_h^{-1/2} Z_s$, $W_{\psi,e,s} \triangleq \psi_h^{-1/2} W_{e,s}$ and note that

$$dZ_{\psi,s} = \psi_h^{-1/2}\sigma_{Z,s} dW_{Z,s}, \qquad dW_{\psi,e,s} = \psi_h^{-1/2}\sigma_{e,s} dW_{e,s}, \qquad \text{with } s \in \mathcal{D}(C). \tag{S.58}$$

For $s \in \left[N_b^0 - Ch^{1-\kappa},\, N_b^0 + Ch^{1-\kappa}\right]$, let $v = \psi_h^{-1}\left(N_b^0 - s\right)$ and, by using the properties of $W_{\cdot,s}$ and the fact that $\sigma_{Z,s}$, $\sigma_{e,s}$ are $\mathscr{F}_s$-measurable, we have

$$dZ_{\psi,t} = \sigma_{Z,t} dW_{Z,t}, \qquad dW_{\psi,e,t} = \sigma_{e,t} dW_{e,t}, \qquad \text{with } t \in \mathcal{T}\left(C\right). \tag{S.59}$$

This can be used into the following quantities for $N_b\left(v\right) \in \mathcal{D}\left(C\right)$. First,

$$\psi_h^{-1} Z_\Delta' Z_\Delta = \sum_{k=T_b(v)+1}^{T_b^0} z_{\psi,kh} z_{\psi,kh},$$

which by (S.58)-(S.59) is such that

$$\psi_h^{-1} Z_\Delta' Z_\Delta = \sum_{k=T_b^0+\lfloor v/h \rfloor}^{T_b^0} z_{kh} z_{kh}', \qquad\qquad v \in \mathcal{D}^*\left(C\right). \tag{S.60}$$

Using the same argument:

$$\psi_h^{-1} Z_\Delta' \widetilde{e} = \sum_{k=T_b^0+\lfloor v/h \rfloor}^{T_b^0} z_{kh} \widetilde{e}_{kh}, \qquad\qquad v \in \mathcal{D}^*\left(C\right). \tag{S.61}$$

Now $N_b\left(v\right)$ varies on $\mathcal{D}^*\left(C\right)$. Furthermore, for sufficiently large T, Lemma 4.1 gives

$$Q_T\left(T_b\right) - Q_T\left(T_b^0\right) = -\delta_h\left(Z_\Delta' Z_\Delta\right)\delta_h \pm 2\delta_h'\left(Z_\Delta' e\right) + o_p\left(h^{1/2}\right),$$

and thus, when multiplying by $h^{-1/2}$, we have

$$\overline{Q}_T\left(T_b\right) = -\left(\delta^0\right)' Z_\Delta' Z_\Delta \left(\delta^0\right) \pm 2\left(\delta^0\right)'\left(h^{-1/2} Z_\Delta' \widetilde{e}\right) + o_p\left(1\right),$$

since on $\mathcal{D}^*\left(C\right)$, $e_{kh} \sim \text{i.n.d.}\, \mathscr{N}(0,\, \sigma_{h,k-1}^2 h)$, $\sigma_{h,k} = O(h^{-1/4})\sigma_{e,k}$ and $\widetilde{e}_{kh}$ is the normalized error [i.e., $\widetilde{e}_{kh} \sim \text{i.n.d.}\, \mathscr{N}(0,\, \sigma_{e,k-1}^2 h)$] defined in (4.1). Hence, according to the re-parametrization introduced in the main text, we examine the behavior of

$$\overline{Q}_T\left(\theta^*\right) = -\left(\delta^0\right)'\left(\sum_{k=T_b+1}^{T_b^0} z_{kh} z_{kh}'\right)\delta^0 + 2\left(\delta^0\right)'\left(h^{-1/2}\sum_{k=T_b+1}^{T_b^0} z_{kh}\widetilde{e}_{kh}\right). \tag{S.62}$$

For the first term, a law of large numbers will be applied which yields convergence in probability toward some quadratic covariation process. For the second term, we observe that the finite-dimensional convergence follows essentially from results in Jacod and Protter (2012) (we indicate the precise theorems below) after some adaptation to our context. Hence, we shall then verify the asymptotic stochastic equicontinuity of the sequence of processes $\{\overline{Q}_T\left(\cdot\right),\, T \geq 1\}$. Let us associate to the continuous-time index t a corresponding $\mathcal{D}^*\left(C\right)$-specific index t_v. This means that each t_v identifies a distinct t in $\mathcal{D}^*\left(C\right)$ through v as defined above. More specifically, for each $\left(\cdot,\, v\right) \in \mathcal{D}^*\left(C\right)$, define the new functions

$$J_{Z,h}\left(v\right) \triangleq \sum_{k=T_b(v)+1}^{T_b^0} z_{kh} z_{kh}', \qquad\qquad J_{e,h}\left(v\right) \triangleq \sum_{k=T_b(v)+1}^{T_b^0} z_{kh}\widetilde{e},$$

for $(T_b(v)+1)h \leq t_v < (T_b(v)+2)h$. For $v \leq 0$, the lower limit of the summation is $T_b(v)+1 = T_b^0 + \lfloor v/h \rfloor$ and thus the number of observations in each sum increases at rate $1/h$. The functions $\{J_{Z,h}(v)\}$ and $\{J_{e,h}(v)\}$ have discontinuous, although *càdlàg*, paths and thus they belong to $\mathbb{D}(\mathcal{D}^*(C), \mathbb{R})$. Since $Z_t^{(j)}$ $(j = 1, \ldots, p)$ is a continuous Itô semimartingale, we have by Theorem 3.3.1 in Jacod and Protter (2012) that $J_{Z,h}(v) \overset{\text{u.c.p.}}{\Rightarrow} [Z, Z]_1(v)$, where $[Z, Z]_1(v) \triangleq [Z, Z]_{h\lfloor N_b^0/h \rfloor} - [Z, Z]_{h\lfloor t_v/h \rfloor}$, and recall by Assumption 2.3 that $[Z, Z]_1(v)$ is equivalent to $\langle Z, Z\rangle_1(v)$ where $\langle Z, Z\rangle_1(v) = \langle Z, Z\rangle_{h\lfloor t_v/h \rfloor}(v)$. Next, let $\mathscr{W}_h(v) = h^{-1/2}J_{e,h}(v)$ and $\mathscr{W}_1(v) = \int_{N_b^0+v}^{N_b^0} \sigma_{Ze,s} dW_s^{1*}$ where W_s^{1*} is defined in Section S.B. By Theorem 5.4.2 in Jacod and Protter (2012) we have $\mathscr{W}_h(v) \overset{\mathcal{L}-\text{s}}{\Rightarrow} \mathscr{W}_1(v)$ under the Skorokhod topology. Note the that both limit processes $[Z, Z]_1(v)$ and $\mathscr{W}_1(v)$ are continuous. This restores the compatibility of the Skorokhod topology with the natural linear structure of $\mathbb{D}(\mathcal{D}^*(C), \mathbb{R})$. For $v \leq 0$, the finite-dimensional stable convergence in law for $\overline{Q}_T(\cdot)$ then follows:

$$\overline{Q}_T(\theta^*) \overset{\mathcal{L}_f-\text{s}}{\to} -\left(\delta^0\right)' \langle Z, Z\rangle_1(v)\, \delta^0 + 2\left(\delta^0\right)' \mathscr{W}_1(v),$$

where $\overset{\mathcal{L}_f-\text{s}}{\to}$ signifies finite-dimensional stable convergence in law. Similarly, for $v > 0$,

$$\overline{Q}_T(\theta^*) \overset{\mathcal{L}_f-\text{s}}{\to} -\left(\delta^0\right)' \langle Z, Z\rangle_2(v)\, \delta^0 + 2\left(\delta^0\right)' \mathscr{W}_2(v).$$

Next, we verify the asymptotic stochastic equicontinuity of the sequence of processes $\{\overline{Q}_T(\cdot), T \geq 1\}$.[5] For $1 \leq i \leq p$, let $\zeta_{h,k}^{(i)} \triangleq z_{kh}^{(i)} \widetilde{e}_{kh}$, $\zeta_{h,k}^{*(i)} \triangleq \mathbb{E}[z_{kh}^{(i)} \widetilde{e}_{kh} | \mathscr{F}_{(k-1)h}]$, and $\zeta_{h,k}^{**(i)} \triangleq \zeta_{h,k}^{(i)} - \zeta_{h,k}^{*(i)}$. For $1 \leq i, j \leq p$, let $\zeta_{Z,h,k}^{(i,j)} \triangleq z_{kh}^{(i)} z_{kh}^{(j)} - \Sigma_{Z,(k-1)h}^{(i,j)} h$,

$$\zeta_{Z,h,k}^{*(i,j)} \triangleq \mathbb{E}\left[z_{kh}^{(i)} z_{kh}^{(j)} - \Sigma_{Z,(k-1)h}^{(i,j)} h | \mathscr{F}_{(k-1)h}\right],$$

and $\zeta_{Z,h,k}^{**(ij)} \triangleq \zeta_{Z,h,k}^{(ij)} - \zeta_{Z,h,k}^{*(ij)}$. Then, we have the following decomposition for $\overline{Q}_T^{\text{c}}(\theta^*) \triangleq \overline{Q}_T(\theta^*) + (\delta^0)' \langle Z, Z\rangle_1(v)\, \delta^0$ (if $v \leq 0$, and defined analogously for $v > 0$),

$$\overline{Q}_T^{\text{c}}(\theta^*) = \sum_{r=1}^{4} \overline{Q}_{r,T}(\theta^*), \tag{S.63}$$

where

$$\overline{Q}_{1,T}(\theta^*) \triangleq -\left(\delta^0\right)' \left(\sum_k \zeta^*{}_{Z,h,k}\right) \delta^0, \qquad \overline{Q}_{2,T}(\theta^*) \triangleq -\left(\delta^0\right)' \left(\sum_k \zeta^{**}{}_{Z,h,k}\right) \delta^0,$$
$$\overline{Q}_{3,T}(\theta^*) \triangleq \left(\delta^0\right)' \left(h^{-1/2} \sum_k \zeta^*{}_{h,k}\right), \quad \text{and} \quad \overline{Q}_{3,T}(\theta^*) \triangleq \left(\delta^0\right)' \left(h^{-1/2} \sum_k \zeta^*{}_{h,k}\right);$$

where $\sum_k$ stands for $\sum_{k=T_b^0+\lfloor v/h \rfloor}^{T_b^0}$. Then,

$$\sup_{(\theta, v) \in \mathcal{D}^*(C)} \left\|\overline{Q}_{3,T}(\theta^*)\right\| \leq K \left\|\delta^0\right\| h^{-1/2} \sum_k \left\|\zeta_{h,k}^*\right\| \overset{P}{\to} 0, \tag{S.64}$$

[5] Although in this proof it is not necessary to consider a neighborhood about δ^0 while proving stochastic equicontinuity, this step will be needed to justify our inference methods later. Thus, this proof is more general and may be useful in other contexts.

which follows from Jacod and Rosenbaum (2013) given that $\Sigma_{Ze,k}=0$ identically by Assumption 4.2. As for $\overline{Q}_{1,T}(\theta,\,v)$, we prove stochastic equicontinuity directly, using the definition in Andrews (1994). Choose any $\varepsilon>0$ and $\eta>0$. Consider any $(\theta,\,v)\,,\,(\bar{\theta},\,\bar{v})$ with $v<0<\bar{v}$ (the other cases can be proven similarly) and $\bar{\delta}=\delta+c_{p\times 1}$, where $c_{p\times 1}$ is a $p\times 1$ vector with each entry equals to $c\in\mathbb{R}$, with $0<c\leq\tau<\infty$, then

$$
\begin{aligned}
&|\overline{Q}_{1,T}(\theta^{*})-\overline{Q}_{1,T}\left(\bar{\theta}^{*}\right)|\\
&=\left|\bar{\delta}'\left(\sum_{k=T_b^0+1}^{T_b(\bar{v})}\zeta^{*}{}_{Z,h,k}\right)\bar{\delta}-\delta'\left(\sum_{k=T_b(v)+1}^{T_b^0}\zeta^{*}_{Z,h,k}\right)\delta\right|\\
&=\left|c'_{p\times 1}\left(\sum_{k=T_b^0+1}^{T_b^0+\lfloor\bar{v}/h\rfloor}\zeta^{*}_{Z,h,k}\right)c_{p\times 1}+\delta'\left(\sum_{k=T_b^0+1}^{T_b(\bar{v})}\zeta^{*}_{Z,h,k}-\sum_{k=T_b^0+\lfloor v/h\rfloor}^{T_b^0}\zeta^{*}_{Z,h,k}\right)\delta\right|\\
&\leq K(\sum_{k=T_b^0+1}^{T_b(\bar{v})}\left\|\zeta^{*}_{Z,h,k}\right\|\left\|c_{p\times 1}\right\|^{2}+\left\|\sum_{k=T_b^0+1}^{T_b^0+\lfloor\bar{v}/h\rfloor}\zeta^{*}_{Z,h,k}-\sum_{k=T_b^0+\lfloor v/h\rfloor}^{T_b^0}\zeta^{*}_{Z,h,k}\right\|\left\|\delta\right\|^{2}\\
&\leq K\left(\left(pc^{2}\right)\sum_{k=T_b^0+1}^{T_b^0+\lfloor\bar{v}/h\rfloor}\left\|\zeta^{*}_{Z,h,k}\right\|+\sum_{k=T_b^0+\lfloor v/h\rfloor}^{T_b(\bar{v})}\left\|\zeta^{*}_{Z,h,k}\right\|\left\|\delta\right\|^{2}\right).
\end{aligned}
$$

By Itô's formula $||\zeta^{*}_{Z,h,k}||=O(h^{3/2})$, and so

$$
\begin{aligned}
\left|\overline{Q}_{1,T}(\theta^{*})-\overline{Q}_{1,T}\left(\bar{\theta}^{*}\right)\right|&\leq K\left(c^{2}h^{-1}O_{p}\left(h^{3/2}\right)O(\tau)+\left\|\delta\right\|^{2}h^{-1}O_{p}\left(h^{3/2}\right)O(\tau)\right)\\
&\leq K\left(c^{2}O_{p}\left(h^{1/2}\right)O(\tau)+\left\|\delta\right\|^{2}O_{p}\left(h^{1/2}\right)O(\tau)\right),
\end{aligned}
$$

which goes to zero uniformly in $\theta^{*}\in\Theta$ as $\tau\rightarrow 0$. Next, consider $\overline{Q}_{2,T}(\theta^{*})$ and observe that for any $r\geq 1$, standard estimates for Itô semimartingales yields $\mathbb{E}(||\zeta^{**}_{Z,h,k}||^{r}|\,\mathscr{F}_{(k-1)h})\leq K_{r}h^{r}$. Then, by using a maximal inequality and choosing $r>2$,

$$
\left(\mathbb{E}\left[\sup_{(\theta,\,v)\in\mathcal{D}^{*}(C)}\left|\overline{Q}_{2,T}(\theta^{*})\right|\right]^{r}\right)^{1/r}\leq K_{r}\left\|\delta^{0}\right\|^{2}h^{-2/r}h\leq K_{r}h^{1-2/r}\rightarrow 0, \tag{S.65}
$$

and thus we can use Markov's inequality together with the latter result to verify that $\overline{Q}_{2,T}(\theta^{*})$ is stochastically equicontinuous. Turning to $\overline{Q}_{4,T}(\theta^{*})$,

$$
\begin{aligned}
&\left|\overline{Q}_{4,T}\left(\bar{\theta}^{*}\right)-\overline{Q}_{4,T}(\theta^{*})\right|\\
&=\left|\bar{\delta}'\left(h^{-1/2}\sum_{k=T_b^0+1}^{T_b^0+\lfloor\bar{v}/h\rfloor}\zeta^{*}_{e,h,k}\right)-\delta'\left(h^{-1/2}\sum_{k=T_b^0+\lfloor v/h\rfloor}^{T_b^0}\zeta^{*}_{e,h,k}\right)\right|\\
&=\left|c'_{p\times 1}\left(h^{-1/2}\sum_{k=T_b^0+1}^{T_b^0+\lfloor\bar{v}/h\rfloor}\zeta^{*}_{e,h,k}\right)\right.\\
&\qquad\left.+\,\delta'\left(h^{-1/2}\sum_{k=T_b^0+1}^{T_b^0+\lfloor\bar{v}/h\rfloor}\zeta^{*}_{e,h,k}-h^{-1/2}\sum_{k=T_b^0+\lfloor v/h\rfloor}^{T_b^0}\zeta^{*}_{e,h,k}\right)\right|
\end{aligned}
$$

$$
\begin{aligned}
&\leq K(h^{-1/2}\sum_{k=T_b^0+1}^{T_b^0+\lfloor \overline{v}/h \rfloor}\left\|\zeta_{e,h,k}^*\right\|\|c_{p\times 1}\| \\
&\quad+\left\|h^{-1/2}\sum_{k=T_b^0+1}^{T_b^0+\lfloor \overline{v}/h \rfloor}\zeta_{e,h,k}^* - h^{-1/2}\sum_{k=T_b^0+\lfloor v/h \rfloor}^{T_b^0}\zeta_{e,h,k}^*\right\|\|\delta\|) \\
&\leq K\left(pch^{-1/2}\sum_{k=T_b^0+1}^{T_b^0+\lfloor \overline{v}/h \rfloor}\left\|\zeta_{e,h,k}^*\right\| + h^{-1/2}\sum_{k=T_b^0+\lfloor v/h \rfloor}^{T_b^0+\lfloor \overline{v}/h \rfloor}\left\|\zeta_{e,h,k}^*\right\|\|\delta\|\right).
\end{aligned}
$$

By the Burkhölder-Davis-Gundy inequality, $||\zeta_{e,h,k}^*|| \leq Kh^{3/2}$ (recall $\Sigma_{Ze,t} = 0$ for all $t \geq 0$), so that

$$
\begin{aligned}
\left|\overline{Q}_{4,T}(\theta^*) - \overline{Q}_{4,T}\left(\bar{\theta}^*\right)\right| &\leq K(c^2h^{-1/2}h^{-1}h^{3/2}O(\tau) \\
&\quad+\|\delta\|^2 h^{-1/2}h^{-1}h^{3/2}O(\tau)) \\
&\leq K\left(c^2O(\tau) + \|\delta\|^2 O(\tau)\right).
\end{aligned}
$$

Then for every $\eta > 0$, with $\mathbf{B}(\tau, (\theta, v))$ a closed ball of radius $\tau > 0$ around θ^*, the quantity

$$
\limsup_{h\downarrow 0} P\left[\sup_{\theta^*\in\Theta:\,\bar{\theta}^*\in\mathbf{B}(\tau,\,\theta^*)}\left|\overline{Q}_{4,T}(\theta^*) - \overline{Q}_{4,T}\left(\bar{\theta}^*\right)\right| > \eta\right], \tag{S.66}
$$

can be made arbitrary less than $\varepsilon > 0$ as $h \downarrow 0$, by choosing τ small enough. Combining (S.64), (S.65) and (S.66), we conclude that the process $\{\overline{Q}_T(\theta, v), T \geq 1\}$ is asymptotically stochastic equicontinuous. Since the finite-dimensional convergence was demonstrated above, this suffices to guarantee the stable convergence in law of the process $\{\overline{Q}_T(\theta, v), T \geq 1\}$ toward a two-sided Gaussian limit process with drift $(\delta^0)'[Z, Z]_.(\cdot)\,\delta^0$, having P-a.s. continuous sample paths with $\mathscr{F}$-conditional covariance matrix given in (S.1). Because $N(\widehat{\lambda}_{b,\pi} - \lambda_0) = O_p(1)$ under the new "fast time scale", and $\mathcal{D}^*(C)$ is compact, then the main assertion of the theorem follows from the continuous mapping theorem for the argmax functional. In view of Section S.D.4.10, a result which shows the negligibility of the drift term, the proof of Theorem 4.1 is complete. □

S.D.4.9 Proof of Theorem 4.2

Proof. By Theorem 4.1 and using the property of the Gaussian law of the limiting process,

$$
\overline{Q}_T(\theta, v) \overset{\mathcal{L}-s}{\Rightarrow} \mathscr{H}(v) = \begin{cases} -(\delta^0)'\langle Z, Z\rangle_1(v)\,\delta^0 + 2\left((\delta^0)'\,\Omega_{\mathscr{W},1}(\delta^0)\right)^{1/2} W_1^*(v), & \text{if } v \leq 0 \\ -(\delta^0)'\langle Z, Z\rangle_2(v)\,\delta^0 + 2\left((\delta^0)'\,\Omega_{\mathscr{W},2}(\delta^0)\right)^{1/2} W_2^*(v), & \text{if } v > 0. \end{cases}
$$

By a change in variable $v = \vartheta^{-1}s$ with $\vartheta = ((\delta^0)'\langle Z, Z\rangle_1\,\delta^0)^2/(\delta^0)'\Omega_{\mathscr{W},1}(\delta^0)$, we can show that

$$
\begin{aligned}
&\underset{v\in\mathcal{A}}{\operatorname{argmax}}\,\mathscr{H}(v) \\
&\overset{d}{\equiv} \underset{s\in\mathcal{A}^*}{\operatorname{argmax}}\,\mathscr{V}(s),
\end{aligned}
$$

where

$$\mathscr{V}(s)=\begin{cases}-\frac{|s|}{2}+W_1^*(s), & \text{if } s<0\\ -\frac{(\delta^0)'\langle Z,Z\rangle_2\delta^0}{(\delta^0)'\langle Z,Z\rangle_1\delta^0}\frac{|s|}{2}+\left(\frac{(\delta^0)'\Omega_{\mathscr{W},2}(\delta^0)}{(\delta^0)'\Omega_{\mathscr{W},1}(\delta^0)}\right)^{1/2}W_2^*(s), & \text{if } s\geq 0,\end{cases}$$

and we have used the facts that $W(s)\overset{d}{\equiv}W(-s)$, $W(cs)\overset{d}{\equiv}|c|^{1/2}W(s)$, and for any $c>0$ and for any function $f(s)$, $\arg\max_s cf(s)=\arg\max_s f(s)$. Thus,

$$\begin{aligned}&\underset{v\in\mathcal{A}}{\operatorname{argmax}}\mathscr{H}(v)\\ &\overset{d}{\equiv}\underset{s\in\mathcal{A}^*}{\operatorname{argmax}}\\ &\quad\left(\frac{\left((\delta^0)'\langle Z,\,Z\rangle_1\delta^0\right)^2}{(\delta^0)'\Omega_{\mathscr{W},1}(\delta^0)}\right)^{-1}\mathscr{V}(s),\end{aligned}$$

and finally by the continuous mapping theorem for the argmax functional,

$$\begin{aligned}&\frac{\left((\delta^0)'\langle Z,\,Z\rangle_1\delta^0\right)^2}{(\delta^0)'\Omega_{\mathscr{W},1}(\delta^0)}N\left(\widehat{\lambda}_{b,\pi}-\lambda_0\right)\\ &\Rightarrow\underset{s\in\mathcal{A}^*}{\operatorname{argmax}}\mathscr{V}(s).\end{aligned}$$

This concludes the proof. □

S.D.4.10 Negligibility of the Drift Term

We are in the setting of Section 3-4. In Proposition 3.1-3.3 and 4.1, the drift processes $\mu_{\cdot,t}$ from (2.3) are clearly of higher order in h and so they are negligible. In Theorem 4.1, we first changed the time scale and then normalized the criterion function by the factor $h^{-1/2}$. The change of time scale now results in

$$dZ_{\psi,s}=\psi_h^{-1/2}\mu_{Z,s}ds+\psi_h^{-1/2}\sigma_{Z,s}dW_{Z,s},\quad dW_{\psi,e,s}=\psi_h^{-1/2}\sigma_{e,s}dW_{e,s},\tag{S.67}$$

with $s\in\mathcal{D}(C)$. Given $s\mapsto t=\psi_h^{-1}s$, we have $\psi_h^{-1/2}\mu_{Z,s}ds=\psi_h^{-1/2}\mu_{Z,s}\psi_h(ds/\psi_h)=\mu_{Z,s}\psi_h^{\vartheta}dt$ with $\vartheta=1/2$. Then, as in (S.59), $dZ_{\psi,t}=\psi_h^{\vartheta}\mu_{Z,t}dt+\sigma_{Z,t}dW_{Z,t}$ and $dW_{\psi,e,t}=\sigma_{e,t}dW_{e,t}$ with $t\in\mathcal{D}^*(C)$. Thus, the change of time scale effectively makes the drift $\mu_{Z,s}ds$ of even higher order. We show a stronger result in that we demonstrate its negligibility even in the case $\vartheta=0$; hence, we show that the limit law of (S.62) remains the same when $\mu_{\cdot,t}$ are nonzero. We set, for any $1\leq i\leq p$ and $1\leq j\leq q+p$,

$$\begin{aligned}&\mu_{Z,k}^{*(i)}\triangleq\int_{(k-1)h}^{kh}\mu_{Z,s}^{(i)}ds, &&\mu_{X,k}^{*(j)}\triangleq\int_{(k-1)h}^{kh}\mu_{X,s}^{(j)}ds,\\ &z_{0,kh}^{(i)}\triangleq\sum_{r=1}^{p}\int_{(k-1)h}^{kh}\sigma_{Z,s}^{(i,r)}dW_Z^{(r)}, \quad\text{and} &&x_{0,kh}^{(j)}\triangleq\sum_{r=1}^{q+p}\int_{(k-1)h}^{kh}\sigma_{X,s}^{(j,r)}dW_X^{(r)}.\end{aligned}$$

Note that

$$z_{kh}^{(i)}x_{kh}^{(j)}=\mu_{Z,k}^{*(i)}\mu_{X,k}^{*(j)}+\mu_{Z,k}^{*(i)}x_{0,kh}^{(j)}+z_{0,kh}^{(i)}\mu_{X,k}^{*(j)}+z_{0,kh}^{(i)}x_{0,kh}^{(j)}.$$

Recall that $\mu_{\cdot,k}^{*(\cdot)}$ is $O(h)$ uniformly in k, and note that $\mu_{Z,k}^{*(i)}x_{0,kh}^{(j)}+\mu_{Z,k}^{*(i)}z_{0,kh}^{(i)}$ follows a Gaussian law with zero mean and variance of order $O(h^3)$. Also note that on $\mathcal{D}^*(C)$, $T_b^0-T_b-1\asymp 1/h$, where $a_h\asymp b_h$ if for some $c \geq 1$, $b_h/c \leq a_h \leq cb_h$. Then,

$$\begin{aligned}\sum_{k=T_b+1}^{T_b^0} z_{kh}^{(i)} x_{kh}^{(j)} &= \sum_{k=T_b+1}^{T_b^0} \mu_{Z,k}^{*(i)} \mu_{X,k}^{*(j)} + \sum_{k=T_b+1}^{T_b^0} \mu_{Z,k}^{*(i)} x_{0,kh}^{(j)} \\ &\quad + \sum_{k=T_b+1}^{T_b^0} z_{0,kh}^{(i)} \mu_{X,k}^{*(j)} + \sum_{k=T_b+1}^{T_b^0} z_{0,kh}^{(i)} x_{0,kh}^{(j)} \\ &= o\left(h^{1/2}\right) + o_p\left(h^{1/2}\right) + \sum_{k=T_b+1}^{T_b^0} z_{0,kh}^{(i)} x_{0,kh}^{(j)}.\end{aligned}$$

Therefore, conditionally on $\Sigma^0 = \{\mu_{\cdot,t},\, \sigma_{\cdot,t}\}_{t\geq 0}$, the limit law of

$$\overline{Q}_T\left(\theta^*\right) = -\left(\delta^0\right)'\left(\sum_{k=T_b+1}^{T_b^0} z_{kh} z_{kh}'\right)\delta^0 + 2\left(\delta^0\right)'\left(h^{-1/2}\sum_{k=T_b+1}^{T_b^0} z_{kh}\widetilde{e}_{kh}\right),$$

is the same as the limit law of

$$-\left(\delta^0\right)'\left(\sum_{k=T_b+1}^{T_b^0} z_{0,kh} z_{0,kh}'\right)\delta^0 + 2\left(\delta^0\right)'\left(h^{-1/2}\sum_{k=T_b+1}^{T_b^0} z_{0,kh}\widetilde{e}_{kh}\right),$$

which completes the proof of Theorem 4.1. □

S.D.4.11 Proof of Proposition 5.1

Proof. Replace ξ_1, ξ_2, ρ and ϑ in (4.7) by their corresponding estimates ξ_1, ξ_2, ρ and ϑ, respectively. Multiply both sides of (4.7) by h^{κ} and apply a change in variable $v = s/h^{\kappa}$. Consider the case $s < 0$. On the "fast time scale", $W_{\cdot}^*$ is replaced by $\widehat{W}_{1,h}(s) = W_{1,h}^*(sh^{\kappa})$ $(s<0)$, where $W_{1,h}^*(s)$ is a sample-size dependent Wiener process. It follows that

$$-h^{-\kappa}\frac{|s|}{2} + h^{-\kappa}W_{1,h}^*\left(h^{\kappa}s\right) = -\frac{|v|}{2} + W_1^*\left(v\right).$$

A similar argument can be applied when $s \geq 0$. Let $\widehat{\mathscr{V}}(s)$ denote $\mathscr{V}(s)$ constructed with the proposed estimates in place of the population parameters. Then,

$$\begin{aligned}h^{-\kappa}\underset{s\in\left[\left(\pi-\widehat{\lambda}_b\right)\widehat{\vartheta},\,\left(1-\pi-\widehat{\lambda}_b\right)\widehat{\vartheta}\right]}{\operatorname{argmax}}\widehat{\mathscr{V}}(s) &= \underset{v\in\left[\left(\pi-\widehat{\lambda}_b\right)\widehat{\vartheta}/h^{2\kappa},\,\left(1-\pi-\widehat{\lambda}_b\right)\widehat{\vartheta}/h^{2\kappa}\right]}{\operatorname{argmax}}\widehat{\mathscr{V}}(v) \\ &\Rightarrow \underset{v\in\left[\left(\pi-\lambda_0\right)\vartheta,\,\left(1-\pi-\lambda_0\right)\vartheta\right]}{\operatorname{argmax}}\mathscr{V}(v),\end{aligned}$$

which is equal to the right-hand side of (4.7) since

$$\vartheta = \left\|\delta^0\right\|^2\overline{\sigma}^{-2}\left(\left(\delta^0\right)'\langle Z,\, Z\rangle_1\delta^0\right)^2/\left(\delta^0\right)'\Omega_{\mathscr{W},1}\left(\delta^0\right).$$

Therefore, equation (4.7) holds when we use the proposed plug-in estimates. □

S.D.5 Proofs of the Results in Section S.C.2

The steps are similar to those used for the case when the model does not include predictable processes. However, we need to rely occasionally on different asymptotic results since the latter processes have distinct statistical properties. Recall that the dependent variable $\Delta_h Y_k$ in model (S.2) is the increment of a discretized process which cannot be identified as an ordinary diffusion. However, its normalized version, $\widetilde{Y}_{(k-1)h} \triangleq h^{1/2} Y_{(k-1)h}$, is well-defined and we exploit this property in the proof. $\Delta_h Y_k$ has first conditional moment of order $O(h^{-1/2})$, it has unbounded variation and does not belong to the usual class of semimartingales.[6] The predictable process $\{Y_{(k-1)h}\}_{k=1}^{T}$ derived from it has different properties. Its "quadratic variation" exists, and thus it is finite in any fixed time interval. That is, the integrated second moments of the regressor $Y_{(k-1)h}$ are finite, i.e., we have

$$\sum_{k=1}^{T} \left(Y_{(k-1)h} h\right)^2 = \sum_{k=1}^{T} \left(h^{1/2} Y_{(k-1)h} h^{1/2}\right)^2 = h \sum_{k=1}^{T} \left(\widetilde{Y}_{(k-1)h}\right)^2 = O_p\left(1\right),$$

by a standard approximation for Riemann sums and recalling that $\widetilde{Y}_{(k-1)h}$ is scaled to be $O_p\left(1\right)$. Then it is easy to see that $\{\widetilde{Y}_{(k-1)h}\}_{k=1}^{T}$ has nice properties. It is left-continuous, adapted, and of finite variation in any finite time interval. When used as the integrand of a stochastic integral, the integral itself makes sense. Importantly, its quadratic variation is null and the process is orthogonal to any continuous local martingale. These properties will be used in the sequel. In analogy to the previous section, we use a localization procedure and thus we need the following assumption related to Assumption S.D.1.

Assumption S.D.2. *Assumption 4.2 and S.C.1 hold, the process $\{\widetilde{Y}_t,\, D_t,\, Z_t\}_{t\geq 0}$ takes value in some compact set and the processes $\{\mu_{\cdot,t},\, \sigma_{\cdot,t}\}_{t\geq 0}$ (except $\{\mu^h_{\cdot,t}\}_{t\geq 0}$) are bounded.*

Recall the notation $M = I - X\left(X'X\right)^{-1} X'$, where now

$$X = \begin{bmatrix} h^{1/2} & Y_0 h & \Delta_h D_1' & \Delta_h Z_1' \\ h^{1/2} & Y_1 h & \Delta_h D_2' & \Delta_h Z_2' \\ \vdots & \vdots & \vdots & \vdots \\ h^{1/2} & Y_{Th} h & \Delta_h D_T' & \Delta_h Z_T' \end{bmatrix}_{T\times(q+p+2)} . \tag{S.68}$$

Thus, $X'X$ is a $(q+p+2)\times(q+p+2)$ matrix given by $\begin{bmatrix} a_1 & a_2 & a_3 & a_4 \end{bmatrix}$, where

$$a_1 = \begin{bmatrix} \sum_{k=1}^{T} h \\ h^{1/2} \sum_{k=1}^{T} \left(Y_{(k-1)h} h\right) \\ \sum_{k=1}^{T} h^{1/2} \left(\Delta_h D_k\right) \\ \sum_{k=1}^{T} h^{1/2} \left(\Delta_h Z_k\right) \end{bmatrix}, \qquad a_2 = \begin{bmatrix} h^{1/2} \sum_{k=1}^{T} \left(Y_{(k-1)h} h\right) \\ \sum_{k=1}^{T} \left(Y_{(k-1)h}^2 \cdot h^2\right) \\ \sum_{k=1}^{T} \left(\Delta_h D_k\right) \left(Y_{(k-1)h} h\right) \\ \sum_{k=1}^{T} \left(\Delta_h Z_k\right) \left(Y_{(k-1)h} h\right) \end{bmatrix},$$

$$a_3 = \begin{bmatrix} \sum_{k=1}^{T} h^{1/2} \left(\Delta_h D_k'\right) \\ \sum_{k=1}^{T} \left(\Delta_h D_k'\right) \left(Y_{(k-1)h} h\right) \\ X_D' X_D \\ X_Z' X_D \end{bmatrix}, \qquad a_4 = \begin{bmatrix} \sum_{k=1}^{T} h^{1/2} \left(\Delta_h Z'_k\right) \\ \sum_{k=1}^{T} \left(\Delta_h Z_k'\right) \left(Y_{(k-1)h} h\right) \\ X_D' X_Z \\ X_Z' X_Z \end{bmatrix},$$

[6] For an introduction to the terminology used in this sub-section, we refer the reader to first chapters in Jacod and Shiryaev (2003).

where $X_D'X_D$ is a $q \times q$ matrix whose (j, r)-th component is the approximate covariation between the j-th and r-th element of D, with $X_D'X_Z$ defined similarly. In view of the properties of $Y_{(k-1)h}$ outlined above and Assumption S.D.2, $X'X$ is $O_p(1)$ as $h \downarrow 0$. The limit matrix is symmetric positive definite where the only zero elements are in the $2 \times (q+p)$ upper right sub-block, and by symmetry in the $(q+p) \times 2$ lower left sub-block. Furthermore, we have

$$X'e = \begin{bmatrix} \sum_{k=1}^{T} h^{1/2} e_{kh} \\ \sum_{k=1}^{T} \left(Y_{(k-1)h} h\right) e_{kh} \\ \sum_{k=1}^{T} \Delta_h D_k e_{kh} \\ \sum_{k=1}^{T} \Delta_h Z_k e_{kh} \end{bmatrix}. \tag{S.69}$$

The other statistics are omitted in order to save space. Again the proofs are first given for the case where the drift processes $\mu_{Z,t}$, $\mu_{D,t}$ of the semimartingale regressors Z and D are identically zero. In the last step we extend the results to nonzero $\mu_{Z,t}$, $\mu_{D,t}$. We also start by conditioning on the processes $\mu_{Z,t}$, $\mu_{D,t}$ and on all the volatility processes so that they are treated as if they were deterministic. We begin with a preliminary lemma.

Lemma S.D.10. *For $1 \leq i \leq 2$, $3 \leq j \leq p+2$ and $\gamma > 0$, $\sum_{k=\lfloor s/h \rfloor}^{\lfloor t/h \rfloor} z_{kh}^{(i)} z_{kh}^{(j)} \overset{\text{u.c.p.}}{\Rightarrow} 0$, for all $N > t > s + \gamma > s > 0$.*

Proof. Without loss of generality consider any $3 \leq j \leq p+2$ and $N > t > s > 0$. We have $\sum_{k=\lfloor s/h \rfloor}^{\lfloor t/h \rfloor} z_{kh}^{(1)} z_{kh}^{(j)} = \sum_{k=\lfloor s/h \rfloor}^{\lfloor t/h \rfloor} \sqrt{h}(\Delta_h M_{Z,k}^{(j)})$, with further $\mathbb{E}[z_{kh}^{(1)} z_{kh}^{(j)} | \mathscr{F}_{(k-1)h}] = 0$, $|z_{kh}^{(1)} z_{kh}^{(j)}| \leq K$ for some K by Assumption S.D.2. Thus $\{z_{kh}^{(i)} z_{kh}^{(j)}, \mathscr{F}_{kh}\}$ is a martingale difference array. Then, for any $\eta > 0$,

$$\begin{aligned} P & \left(\sum_{k=\lfloor s/h \rfloor}^{\lfloor t/h \rfloor} \left| z_{kh}^{(1)} z_{kh}^{(j)} \right|^2 > \eta \right) \\ & \leq \frac{K}{\eta} \mathbb{E} \left(\sum_{k=\lfloor s/h \rfloor}^{\lfloor t/h \rfloor} h^2 \left(\Delta_h M_{Z,k}^{(j)} \right)^2 \right) \leq \frac{K}{\eta} h O_p(t-s) \to 0, \end{aligned}$$

where the second inequality follows from the Burkhölder-Davis-Gundy inequality with parameter $r = 2$. This shows that the array $\{|z_{kh}^{(i)} z_{kh}^{(j)}|^2\}$ is asymptotically negligible. By Lemma 2.2.11 in the Appendix of Jacod and Protter (2012), we verify the claim for $i = 1$. For the case $i = 2$ note that $z_{kh}^{(2)} z_{kh}^{(j)} = (Y_{(k-1)h} h)(\Delta_h M_{Z,k}^{(j)})$, and recall that $\widetilde{Y}_{(k-1)h} = h^{1/2} Y_{(k-1)h} = O_p(1)$. Thus, the same proof remains valid for the case $i = 2$. The assertion of the lemma follows. □

S.D.5.1 Proof of Proposition S.C.1

Proof of part (i) of Proposition S.C.1. Following the same steps that led to (S.12), we can write

$$Q_T(T_b) - Q_T(T_0) = -\left|T_b - T_b^0\right| d(T_b) + g_e(T_b), \qquad \text{for all } T_b, \tag{S.70}$$

where

$$d(T_b) \triangleq \frac{(\delta_Z^0)' \left\{ (Z_0' M Z_0) - (Z_0' M Z_2)(Z_2' M Z_2)^{-1} (Z_2' M Z_0) \right\} \delta_Z^0}{|T_b - T_b^0|}, \tag{S.71}$$

and we arbitrarily define $d(T_b) = (\delta_Z^0)' \delta_Z^0$ when $T_b = T_b^0$. Let $d_T = T \inf_{|T_b - T_b^0| > TK} d(T_b)$; it is positive and bounded away from zero by Lemma S.D.11 below. Then

$$
\begin{aligned}
P\left(\left|\widehat{\lambda}_b - \lambda_0\right| > K\right) &= P\left(\left|\widehat{T}_b - T_b^0\right| > TK\right) \\
&\leq P\left(\sup_{|T_b - T_b^0| > TK} |g_e(T_b)| \geq \inf_{|T_b - T_b^0| > TK} \left|T_b - T_b^0\right| d(T_b)\right) \\
&\leq P\left(\sup_{p+2 \leq T_b \leq T-p-2} |g_e(T_b)| \geq TK \inf_{|T_b - T_b^0| > TK} d(T_b)\right) \\
&= P\left(d_T^{-1} \sup_{p+2 \leq T_b \leq T-p-2} |g_e(T_b)| \geq K\right). \qquad \text{(S.72)}
\end{aligned}
$$

We can write the first term of $g_e(T_b)$ as

$$
2\left(\delta^0\right)' (Z_0' M Z_2) (Z_2' M Z_2)^{-1/2} (Z_2' M Z_2)^{-1/2} Z_2 M e. \qquad \text{(S.73)}
$$

For the stochastic regressors, Theorem S.D.5 implies that for any $3 \leq j \leq p+2$, $(Z_2 e)_{j,1} / \sqrt{h} = O_p(1)$ and for any $3 \leq i \leq q+p+2$, $(Xe)_{i,1} / \sqrt{h} = O_p(1)$, since these estimates include a positive fraction of the data. We can use the above expression for $X'X$ to verify that $Z_2' M Z_2$ and $Z_0' M Z_2$ are $O_p(1)$. Then,

$$
\sup_{T_b} (Z_0' M Z_2) (Z_2' M Z_2)^{-1} (Z_2' M Z_0) \leq Z_0' M Z_0 = O_p(1),
$$

by Lemma S.D.3. Next, note that the first two elements of the vector $X'e$ and $Z_2' e$ are $O_p\left(h^{1/2}\right)$ [recall (S.69)]. By Assumption 2.1-(iii) and the inequality

$$
\sup_{T_b} \left\|(Z_2' M Z_2)^{-1/2} Z_2 M e\right\| \leq \sup_{T_b} \left\|(Z_2' M Z_2)^{-1/2}\right\| \sup_{T_b} \|Z_2 M e\|,
$$

we have that $(Z_2' M Z_2)^{-1/2} Z_2 M e$ is $O_p(h^{1/2})$ uniformly in T_b since the last $q+p$ (resp., p) elements of $X'e$ (resp., $Z_2' e$) are $o_p(1)$ locally uniformly in time. Therefore, uniformly over $p+2 \leq T_b \leq T-p-2$, the overall expression in (S.73) is $O_p(h^{1/2})$. As for the second term of (S.10), $Z_0' M e = O_p(h^{1/2})$. The first term in (S.11) is uniformly negligible and so is the last. Therefore, combining these results we can show that $\sup_{T_b} |g_e(T_b)| = O_p(\sqrt{h})$. Using Lemma S.D.11 below, we have $P(d_T^{-1} \sup_{p+2 \leq T_b \leq T-p-2} |g_e(T_b)| \geq K) \leq \varepsilon$, which shows that $\widehat{\lambda}_b \xrightarrow{P} \lambda_0$. □

Lemma S.D.11. *Let $d_B = \inf_{|T_b - T_b^0| > TB} T d(T_b)$. There exists a $\kappa > 0$ and for every $\varepsilon > 0$, there exists a $B < \infty$ such that $P(d_B \geq \kappa) \leq 1 - \varepsilon$.*

Proof. Assuming $N_b \leq N_b^0$ and following the same steps as in Lemma S.D.6 (but replacing R by $\overline{R}$)

$$
\begin{aligned}
T d(T_b) &\geq T\left(\delta_Z^0\right)' \overline{R}' \frac{X_\Delta' X_\Delta}{T_b^0 - T_b} (X_2' X_2)^{-1} (X_0' X_0) \overline{R}\left(\delta_Z^0\right) \\
&= \left(\delta_Z^0\right)' \overline{R}' \frac{X_\Delta' X_\Delta}{B} (X_2' X_2)^{-1} (X_0' X_0) \overline{R}\left(\delta_Z^0\right).
\end{aligned}
$$

Under Assumption 2.1-(iii) and in view of (S.68), $X_\Delta' X_\Delta$ is positive definite: for the $p \times p$ lower-right sub-block apply Lemma S.D.3 as in the proof of Lemma S.D.6, whereas for the remaining elements of $X_\Delta' X_\Delta$ the result follows from the convergence of approximations to Riemann sums. Note that $X_2' X_2$ and

$X_0'X_0$ are $O_p(1)$. It follows that

$$Td(T_b) \geq \left(\delta_Z^0\right)' \overline{R}' \frac{X_\Delta' X_\Delta}{N} \left(X_2'X_2\right)^{-1} \left(X_0'X_0\right) \overline{R}\delta_Z^0 \geq \kappa > 0.$$

The result follows choosing $B > 0$ such that $P(d_B \geq \kappa)$ is larger than $1-\varepsilon$. $\square$

Proof of part (ii) of Proposition S.C.1. We introduce again

$$\mathbf{D}_{K,T} = \left\{T_b : \, N\eta \leq N_b \leq N(1-\eta), \left|N_b^0 - N_b\right| > KT^{-1}\right\},$$

and observe that it is enough to show that $P(\sup_{T_b \in \mathbf{D}_{K,T}} Q_T(T_b) \geq Q_T(T_b^0)) < \varepsilon$, or

$$P\left(\sup_{T_b \in \mathbf{D}_{K,T}} h^{-1} g_e(T_b) \geq \inf_{T_b \in \mathbf{D}_{K,T}} h^{-1} \left|T_b - T_b^0\right| d(T_b)\right) < \varepsilon. \tag{S.74}$$

By Lemma S.D.1,

$$\inf_{T_b \in \mathbf{D}_{K,T}} d(T_b) \geq \inf_{T_b \in \mathbf{D}_{K,T}} \left(\delta_Z^0\right)' \overline{R}' \frac{X_\Delta' X_\Delta}{T_b^0 - T_b} \left(X_2'X_2\right)^{-1} \left(X_0'X_0\right) \overline{R}\delta_Z^0.$$

For the $(q+p)\times(q+p)$ lower right sub-block of $X_\Delta' X_\Delta$ the arguments of Proposition 3.2 apply: $(h(T_b^0 - T_b))^{-1}$ $[X_\Delta' X_\Delta]_{\{\cdot,(q+p)\times(q+p)\}}$ is bounded away from zero for all $T_b \in \mathbf{D}_{K,T}$ by choosing K large enough (recall $|T_b^0 - T_b| > K$), where $[A]_{\{\cdot, i\times j\}}$ is the $i \times j$ lower right sub-block of A. Furthermore, this approximation is uniform in T_b by Assumption 3.1. It remains to deal with the upper left sub-block of $X_\Delta' X_\Delta$. Consider its (1, 1)-th element. It is given by $\sum_{k=T_b+1}^{T_b^0} (h^{1/2})^2$. Thus $(h(T_b^0 - T_b))^{-1} \sum_{k=T_b+1}^{T_b^0} (h^{1/2})^2 > 0$. The same argument applies to the (2, 2)-th element of the upper left sub-block of $X_\Delta' X_\Delta$. The latter results imply that $\inf_{T_b \in \mathbf{D}_{K,T}} Td(T_b)$ is bounded away from zero. It remains to show that $\sup_{T_b \in \mathbf{D}_{K,T}} (h|T_b - T_b^0|)^{-1} g_e(T_b)$ is small when T is large. Recall that the terms Z_2 and Z_0 involve a positive fraction $N\eta$ of the data. We can apply Lemma S.D.3 to those elements which involve the stochastic regressors only, whereas the other terms are dealt with directly using the definition of $X'e$ in (S.69). Consider the first term of $g_e(T_b)$. Using the same steps which led to (S.19), we have

$$\begin{aligned}
&\left|2\left(\delta_Z^0\right)' \left(Z_0'MZ_2\right)\left(Z_2'MZ_2\right)^{-1} Z_2Me - 2\left(\delta_Z^0\right)'\left(Z_0'Me\right)\right| \\
&= \left|\left(\delta_Z^0\right)' Z_\Delta' Me\right| + \left|\left(\delta_Z^0\right)'\left(Z_\Delta' MZ_2\right)\left(Z_2'MZ_2\right)^{-1}\left(Z_2Me\right)\right|.
\end{aligned} \tag{S.75}$$

We can apply Lemma S.D.3 to the terms that do not involve $|N_b - N_b^0|$ but only stochastic regressors. Next consider the first term of

$$\begin{aligned}
\left(h\left(T_b^0 - T_b\right)\right)^{-1}\left(\delta_Z^0\right)'\left(Z_\Delta' MZ_2\right) = {} & \frac{\left(\delta_Z^0\right)'\left(Z_\Delta' Z_\Delta\right)}{h\left(T_b^0 - T_b\right)} \\
& - \left(\delta_Z^0\right)'\left(\frac{Z_\Delta' X_\Delta}{h\left(T_b^0 - T_b\right)}\left(X'X\right)^{-1} X'Z_2\right).
\end{aligned}$$

Applying the same manipulations as those used above for the $p \times p$ lower right sub-block of $Z_\Delta' Z_\Delta$, we have $(h(T_b^0 - T_b))^{-1}[Z_\Delta' Z_\Delta]_{\{\cdot, p\times p\}} = O_p(1)$, since there are $T_b^0 - T_b$ summands whose conditional first moments are each $O(h)$. The $O_p(1)$ result is uniform by Assumption 3.1. The same argument holds for the corresponding sub-block of $Z_\Delta' X_\Delta/(h(T_b^0 - T_b))$. Hence, as $h \downarrow 0$ the second term above is $O_p(1)$. Next, consider the upper left 2×2 block of $Z_\Delta' Z_\Delta$ (the same argument holds true for $Z_\Delta' X_\Delta$). Note that the predictable variable $Y_{(k-1)h}$ in the (2, 2)-th element can be treated as locally constant after multiplying by $h^{1/2}$ (recall $h^{1/2}Y_{(k-1)h} = \widetilde{Y}_{(k-1)h} = O_p(1)$ by Assumption S.D.2),

$$\sum_{k=T_b+1}^{T_b^0} \left(Y_{(k-1)h}h\right)^2 = \sum_{k=T_b+1}^{T_b^0} \left(\widetilde{Y}_{(k-1)h}h^{1/2}\right)^2 \leq C \sum_{k=T_b+1}^{T_b^0} h,$$

where $C = \sup_k |\widetilde{Y}^2_{(k-1)h}|$ is a fixed constant given the localization in Assumption S.D.2. Thus, when multiplied by $(h(T_b^0 - T_b))^{-1}$, the (2, 2)-th element of $Z'_\Delta Z_\Delta$ is $O_p(1)$. The same reasoning can be applied to the corresponding (1, 1)-th element. Next, let us consider the cross-products between the semimartingale regressors and the predictable regressors. Consider any $3 \leq j \leq p+2$,

$$\begin{aligned}\frac{1}{h\left(T_b^0 - T_b\right)} \sum_{k=T_b+1}^{T_b^0} z_{kh}^{(2)} z_{kh}^{(j)} &= \frac{1}{h\left(T_b^0 - T_b\right)} \sum_{k=T_b+1}^{T_b^0} \left(\widetilde{Y}_{(k-1)h}h^{1/2}\right) z_{kh}^{(j)} \\ &= \frac{1}{T_b^0 - T_b} \sum_{k=T_b+1}^{T_b^0} \widetilde{Y}_{(k-1)h} \frac{z_{kh}^{(j)}}{\sqrt{h}}.\end{aligned}$$

Since $z_{kh}^{(j)}/\sqrt{h}$ is i.n.d. with zero mean and finite variance and $\widetilde{Y}_{(k-1)h}$ is $O_p(1)$ by Assumption S.D.2, Assumption 3.1 implies that we can find a K large enough such that the right hand side is $O_p(1)$ uniformly in T_b. The same argument applies to $(Z'_\Delta Z_\Delta)_{1,j}$, $3 \leq j \leq p+2$. This shows that the term $(Z'_\Delta X_\Delta/(h(T_b^0 - T_b)))\,(X'X)^{-1} X'Z_2$ is bounded and so is $Z'_\Delta X_\Delta/(h(T_b^0 - T_b))$ using the same reasoning. Thus, $(h(T_b^0 - T_b))^{-1}\,(\delta_Z^0)'\,(Z'_\Delta M Z_2)$ is $O_p(1)$. By the same arguments as before, we can use Theorem S.D.5 to show that the second term of (S.75) is $O_p(h^{1/2})$ when multiplied by $(h(T_b^0 - T_b))^{-1}$ since the last term involves a positive fraction of the data. Now, expand the $(p+2)$-dimensional vector $Z'_\Delta Me$ as follows

$$\begin{aligned}\frac{Z'_\Delta Me}{h\left(T_b^0 - T_b\right)} &= \frac{1}{h\left(T_b^0 - T_b\right)} \sum_{k=T_b+1}^{T_b^0} z_{kh}e_{kh} \\ &\quad - \frac{1}{h\left(T_b^0 - T_b\right)} \left(\sum_{k=T_b+1}^{T_b^0} z_{kh}x'_{kh}\right) \left(X'X\right)^{-1} \left(X'e\right).\end{aligned}$$

The arguments for the last p elements are the same as above and yield [recall (S.20)]

$$\frac{[Z'_\Delta Me]_{\{\cdot,p\}}}{h\left(T_b^0 - T_b\right)} = o_p\left(K^{-1}\right) - O_p(1)\, O_p\left(h^{1/2}\right),$$

where we recall that by Assumption 2.2 $\Sigma_{Ze,N_b^0} = 0$. Note that the convergence is uniform over T_b by Lemma S.D.2. We now consider the first two elements of $Z'_\Delta e$:

$$\left|\sum_{k=T_b+1}^{T_b^0} z_{kh}^{(2)} e_{kh}\right| = \left|\sum_{k=T_b+1}^{T_b^0} h^{1/2}Y_{(k-1)h}h^{1/2}e_{kh}\right| \leq A \sum_{k=T_b+1}^{T_b^0} \left|\widetilde{Y}_{(k-1)h}h^{1/2}e_{kh}\right|,$$

for some positive $A < \infty$. Noting that $e_{kh}/\sqrt{h} \sim \text{i.n.d.}\mathscr{N}(0,\, \sigma^2_{e,k-1})$, we have

$$\left(h\left(T_b^0 - T_b\right)\right)^{-1} \sum_{k=T_b+1}^{T_b^0} z_{kh}^{(2)} e_{kh} \leq C \left(\left(T_b^0 - T_b\right)^{-1} \sum_{k=T_b+1}^{T_b^0} \left|e_{kh}/h^{1/2}\right|\right)$$

where $C = \sup_k |\widetilde{Y}_{(k-1)h}|$ is finite by Assumption S.D.2. Choose K large enough such that the probability that the right-hand side is larger than $B/3N$ is less than ε. For the first element of $Z'_\Delta e$ the argument is the same and thus

$$P\left(\left(h\left(T_b^0 - T_b\right)\right)^{-1}\sum_{k=T_b+1}^{T_b^0} z_{kh}^{(1)} e_{kh} > \frac{B}{3N}\right) \leq \varepsilon,$$

when K is large. For the last product in the second term of $Z'_\Delta Me/h$ the argument is easier. This component includes a positive fraction of data and thus

$$\sum_{k=1}^{T} x_{kh}^{(1)} e_{kh} = \sum_{k=1}^{T} h^{1/2} e_{kh} = h^{1/2} O_p(1), \tag{S.76}$$

using the result $\sum_{k=1}^{\lfloor t/h \rfloor} e_{kh} \overset{\text{u.c.p.}}{\Rightarrow} \int_0^t \sigma_{e,s} dW_{e,s}$. A similar argument applies to $x_{kh}^{(2)} e_{kh}$ by using in addition the localization Assumption S.D.2. Combining the above derivations, we have

$$\frac{1}{h\left(T_b^0 - T_b\right)} g_e(T_b) = \frac{1}{h\left(T_b^0 - T_b\right)} \left(\delta_Z^0\right)' 2 Z'_\Delta e + o_p(1). \tag{S.77}$$

In order to prove

$$P\left(\sup_{T_b \in \mathbf{D}_{K,T}} \left(h\left(T_b^0 - T_b\right)\right)^{-1} g_e(T_b) \geq \inf_{T_b \in \mathbf{D}_{K,T}} h^{-1} d(T_b)\right) < \varepsilon,$$

we can use (S.77). To this end, we shall find a $K > 0$, such that

$$\begin{aligned} P&\left(\sup_{T_b \leq T_b^0 - \frac{K}{N}} \left|\mu_\delta^0 \frac{2}{h}\left(T_b^0 - T_b\right)^{-1} \sum_{k=T_b+1}^{T_b^0} z_{kh}^{(1)} e_{kh}\right| > \frac{B}{3N}\right) \\ &\leq P\left(\sup_{T_b \leq T_b^0 - \frac{K}{N}} \left(T_b^0 - T_b\right)^{-1} \left|\sum_{k=T_b+1}^{T_b^0} \frac{e_{kh}}{\sqrt{h}}\right| > \frac{B}{6\left|\mu_\delta^0\right| N}\right) < \frac{\varepsilon}{3}. \end{aligned} \tag{S.78}$$

Recalling that $e_{kh}/h^{1/2} \sim \mathscr{N}(0, \sigma_{e,k-1}^2)$, the Hájek-Réiny inequality yields

$$P\left(\sup_{T_b \leq T_b^0 - \frac{K}{N}} \left(T_b^0 - T_b\right)^{-1} \left|\sum_{k=T_b+1}^{T_b^0} \frac{e_{kh}}{\sqrt{h}}\right| > \frac{B}{6\left|\mu_\delta^0\right| N}\right) \leq A \frac{36\left(\mu_\delta^0\right)^2 N^2}{B^2} \frac{1}{KN^{-1}}.$$

We can choose K sufficiently large such that the right-hand side is less than $\varepsilon/3$. The same bound holds for the second element of $Z'_\Delta e$. Next, by equation (S.22),

$$P\left(\sup_{T_b \leq T_b^0 - \frac{K}{N}} \frac{1}{h\left(T_b^0 - T_b\right)} \left\| 2\left(\delta_Z^0\right)' \sum_{k=T_b+1}^{T_b^0} \left[Z'_\Delta e\right]_{\{\cdot,p\}} \right\| > \frac{B}{3N}\right) < \frac{\varepsilon}{3},$$

since for each $j = 3, \ldots, p$, $\{z_{kh}^{(j)} e_{kh}/h\}$ is i.n.d. with finite variance, and thus the result is implied by the Hájek-Réiny inequality for large K. Using the latter results into (S.77), we have

$$P\left(\sup_{T_b \leq T_b^0 - \frac{K}{N}} \frac{1}{h\left(T_b^0 - T_b\right)} \left\| 2\left(\delta_Z^0\right)' \sum_{k=T_b+1}^{T_b^0} z_{kh} e_{kh} \right\| > \frac{B}{N}\right) < \varepsilon,$$

which verifies (S.74) and thus proves our claim. □

S.D.5.2 Proof of Theorem S.C.1

Part (i)-(ii) follows the same steps as in the proof of Proposition 4.1 part (i)-(ii) but using the results developed throughout the proof of part (i)-(ii) of Proposition S.C.1. As for part (iii), we begin with the following lemma, where again $\psi_h = h^{1-\kappa}$. Without loss of generality we set $B = 1$ in Assumption 4.1.

Lemma S.D.12. *Under Assumption 4.2 and S.D.2, uniformly in T_b,*

$$\left(Q_T\left(T_b\right) - Q_T\left(T_b^0\right)\right)/\psi_h = -\delta_h\left(Z_\Delta' Z_\Delta/\psi_h\right)\delta_h \pm 2\delta_h'\left(Z_\Delta'\widetilde{e}/\psi_h\right) + O_p\left(h^{3/4\wedge 1-\kappa/2}\right).$$

Proof. By the definition of $Q_T\left(T_b\right) - Q_T\left(T_b^0\right)$ and Lemma S.D.9,

$$\begin{aligned} &Q_T\left(T_b\right) - Q_T\left(T_b^0\right) \\ &\quad = -\delta_h'\left\{Z_\Delta' M Z_\Delta + \left(Z_\Delta' M Z_2\right)\left(Z_2' M Z_2\right)^{-1}\left(Z_2' M Z_\Delta\right)\right\}\delta_h \\ &\quad + g_e\left(T_b,\, \delta_h\right). \end{aligned} \tag{S.79}$$

We can expand the first term of (S.79) as follows

$$\delta_h' Z_\Delta' M Z_\Delta \delta_h = \delta_h' Z_\Delta' Z_\Delta \delta_h - \delta_h' A \delta_h, \tag{S.80}$$

where $A = Z_\Delta' X\left(X'X\right)^{-1} X' Z_\Delta$. We show that $\delta_h' A \delta_h$ is uniformly of higher order than $\delta_h' Z_\Delta' Z_\Delta \delta_h$. The cross-products between the semimartingale and the predictable regressors (i.e., the $p \times 2$ lower-left sub-block of $Z_\Delta' X$) are $o_p(1)$, as can be easily verified. Lemma S.D.10 provides the formal statement of the result for $Z_\Delta' Z_\Delta$. Hence, the result carries over to $Z_\Delta' X$ with no changes and by symmetry also to is the $2 \times p$ upper-right block. This allows us to treat the 2×2 upper-left block and the $p \times p$ lower-right block of elements, such as those in A separately. By Lemma S.D.3, $\left(X'X\right)^{-1} = O_p(1)$. Using Proposition 4.1-(ii), we let $N_b - N_b^0 = K\psi_h$. By the Burkhölder-Davis-Gundy inequality, we have standard estimates for local volatility so that

$$\left\|\mathbb{E}\left(\widehat{\Sigma}_{ZX}^{(i,j)}\left(T_b,\, T_b^0\right) - \Sigma_{ZX,\left(T_b^0-1\right)h}^{(i,j)} | \mathscr{F}_{\left(T_b^0-1\right)h}\right)\right\| \leq K h^{1/2},$$

with $3 \leq i \leq p+2$ and $3 \leq j \leq q+p+2$ which in turn implies $[Z_\Delta' X_\Delta]_{\{\cdot, p\times p\}} = O_p(1/(h(T_b^0 - T_b)))$. The same bound applies to the corresponding blocks of $Z_\Delta' Z_\Delta$ and $X_\Delta' Z_\Delta$. Now we focus on the (2, 2)-th element of A. First notice that

$$\left(Z_\Delta' X\right)_{2,2} = \sum_{k=T_b+1}^{T_b^0} z_{kh}^{(2)} x_{kh}^{(2)} = \sum_{k=T_b+1}^{T_b^0}\left(\widetilde{Y}_{(k-1)h}\right)^2 h.$$

By a localization argument (cf. Assumption S.D.2), $\widetilde{Y}_{(k-1)h}$ is bounded. Then, since the number of summands grows at a rate T^κ, we have $\left(Z_\Delta' X\right)_{2,2} = O_p(K h^{1-\kappa})$. The same proof can be used for $\left(Z_\Delta' X\right)_{1,1}$, which gives $\left(Z_\Delta' X\right)_{1,1} = O_p(K h^{1-\kappa})$. Thus, in view of (S.81), we conclude that (S.80) when divided by ψ_h is such that

$$\delta_h' Z_\Delta' M Z_\Delta \delta_h/\psi_h = \delta_h' Z_\Delta' Z_\Delta \delta_h/\psi_h - \delta_h' Z_\Delta' X\left(X'X\right)^{-1} X' Z_\Delta \delta_h/\psi_h$$

$$= \psi_h^{-1} \left(\delta^0\right)' Z'_\Delta Z_\Delta \delta^0 - \psi_h^{-1} h^{1/2} O_p \left(h^{2(1-\kappa)}\right). \tag{S.81}$$

For the second term of (S.79), we have

$$\begin{aligned} &\psi_h^{-1} h^{1/2} \left(\delta^0\right)' \left\{ \left(Z'_\Delta M Z_2\right) \left(Z'_2 M Z_2\right)^{-1} \left(Z'_2 M Z_\Delta\right) \right\} \delta^0 \\ &\quad = \psi_h^{-1} h^{1/2} \left\|\delta_0\right\|^2 O_p\left(\psi_h\right) O_p\left(1\right) O_p\left(\psi_h\right) \leq K \psi_h^{-1} h^{1/2} O_p\left(h^{2(1-\kappa)}\right) \end{aligned} \tag{S.82}$$

uniformly in T_b, which follows from applying the same reasoning used for $Z'_\Delta \left(I - M\right) Z_\Delta$ above to each of these three elements. Finally, consider the stochastic term $g_e\left(T_b,\, \delta_h\right)$. We have

$$\begin{aligned} g_e\left(T_b,\, \delta_h\right) = {} & 2\delta'_h \left(Z'_0 M Z_2\right) \left(Z'_2 M Z_2\right)^{-1} Z_2 M e - 2\delta'_h \left(Z'_0 M e\right) \\ & + e' M Z_2 \left(Z'_2 M Z_2\right)^{-1} Z_2 M e - e' M Z_0 \left(Z'_0 M Z_0\right)^{-1} Z'_0 M e. \end{aligned} \tag{S.83}$$

Recall (S.69), and $\sum_{k=T_b+1}^{T_b^0} x_{kh} e_{kh} = h^{-1/4} \sum_{T_b+1}^{T_b^0} x_{kh} \widetilde{e}_{kh}$. Introduce the following decomposition,

$$\left(X'e\right)_{2,1} = \sum_{k=1}^{T_b^0 - \lfloor T^\kappa \rfloor} x_{kh}^{(2)} \widetilde{e}_{kh} + h^{-1/4} \sum_{k=T_b^0 - \lfloor T^\kappa \rfloor + 1}^{T_b^0 + \lfloor T^\kappa \rfloor} x_{kh}^{(2)} \widetilde{e}_{kh} + \sum_{k=T_b^0 + \lfloor T^\kappa \rfloor + 1}^{T} x_{kh}^{(2)} \widetilde{e}_{kh},$$

where $\widetilde{e}_{kh} \sim$ i.n.d. $\mathscr{N}(0,\, \sigma_{e,k-1}^2 h)$. The first and third terms are $O_p(h^{1/2})$ in view of (S.76). The term in the middle is $h^{3/4} \sum_{k=T_b^0 - \lfloor T^\kappa \rfloor + 1}^{T_b^0 + \lfloor T^\kappa \rfloor} \widetilde{Y}_{(k-1)h} h^{-1/2} \widetilde{e}_{kh}$, which involves approximately $2T^\kappa$ summands. Since $\widetilde{Y}_{(k-1)h}$ is bounded by the localization procedure,

$$h^{3/4} \frac{T^{\kappa/2}}{T^{\kappa/2}} \sum_{k=T_b^0 - \lfloor T^\kappa \rfloor}^{T_b^0 + \lfloor T^\kappa \rfloor} \widetilde{Y}_{(k-1)h} \frac{\widetilde{e}_{kh}}{\sqrt{h}} = h^{3/4} T^{\kappa/2} O_p\left(1\right),$$

or

$$h^{-1/4} \sum_{k=T_b^0 - \lfloor T^\kappa \rfloor}^{T_b^0 + \lfloor T^\kappa \rfloor} x_{kh}^{(2)} \widetilde{e}_{kh} = h^{3/4-\kappa/2} O_p\left(1\right).$$

This implies that $\left(X'e\right)_{2,1}$ is $O_p(h^{1/2 \wedge 3/4 - \kappa/2})$. The same observation holds for $\left(X'e\right)_{1,1}$. Therefore, one follows the same steps as in the concluding part of the proof of Lemma 4.1 [cf. equation (S.54) and the derivations thereafter]. That is, for the first two terms of $g_e(T_b,\, \delta_h)$, using $Z'_0 M Z_2 = Z'_2 M Z_2 \pm Z'_\Delta M Z_2$, we have

$$\begin{aligned} &2h^{1/4} \left(\delta^0\right)' \left(Z'_0 M Z_2\right) \left(Z'_2 M Z_2\right)^{-1} Z_2 M e - 2h^{1/4} \left(\delta^0\right)' \left(Z'_0 M e\right) \\ &\quad = 2h^{1/4} \left(\delta^0\right)' Z'_\Delta M e \pm 2h^{1/4} \left(\delta^0\right)' Z'_\Delta M Z_2 \left(Z'_2 M Z_2\right)^{-1} Z'_2 M e. \end{aligned} \tag{S.84}$$

The last term above when multiplied by ψ_h^{-1} is such that

$$\psi_h^{-1} 2h^{1/4} \left(\delta^0\right)' Z'_\Delta M Z_2 \left(Z'_2 M Z_2\right)^{-1} Z'_2 M e = \left\|\delta^0\right\| O_p\left(1\right) O_p\left(h^{1 \wedge 5/4 - \kappa/2}\right),$$

where we have used the fact that $Z'_\Delta M Z_2 / \psi_h = O_p\left(1\right)$. For the first term of (S.84),

$$\begin{aligned} &2h^{1/4} \left(\delta^0\right)' Z'_\Delta M e / \psi_h \\ &\quad = 2h^{1/4} \left(\delta^0\right)' Z'_\Delta e / \psi_h - 2h^{1/4} \left(\delta^0\right)' Z'_\Delta X \left(X'X\right)^{-1} X'e / \psi_h \end{aligned}$$

$$= 2h^{1/4}\left(\delta^0\right)' Z'_\Delta e - 2\left(\delta^0\right)' O_p(1)\, O_p\left(h^{1\wedge 5/4-\kappa/2}\right).$$

As in the proof of Lemma 4.1, we can now use part (i) of the theorem so that the difference between the terms on the second line of $g_e(T_b,\,\delta_h)$ is negligible. That is, we can find a c_T sufficiently small such that,

$$\psi_h^{-1}\left[e'MZ_2\left(Z_2'MZ_2\right)^{-1}Z_2Me - e'MZ_0\left(Z_0'MZ_0\right)^{-1}Z_0'Me\right] = o_p(c_T h).$$

This leads to

$$\begin{aligned} g_e(T_b,\,\delta_h)/\psi_h &= 2h^{1/4}\left(\delta^0\right)' Z'_\Delta e/\psi_h + O_p\left(h^{3/4\wedge 1-\kappa/2}\right) \\ &\quad + \left\|\delta^0\right\| O_p\left(h^{3/4\wedge 1-\kappa/2}\right) + o_p(c_T h), \end{aligned}$$

for a sufficiently small c_T. This together with (S.81) and (S.82) yields,

$$\begin{aligned} \psi_h^{-1}\left(Q_T(T_b) - Q_T\left(T_b^0\right)\right) &= -\delta_h\left(Z'_\Delta Z_\Delta/\psi_h\right)\delta_h \\ &\quad \pm 2\delta_h'\left(Z'_\Delta e/\psi_h\right) + O_p\left(h^{3/4\wedge 1-\kappa/2}\right) + o_p\left(h^{1/2}\right), \end{aligned}$$

when T is large, where c_T is a sufficiently small number. This concludes the proof. □

Proof of part (iii) of Theorem S.C.1. We proceed as in the proof of Theorem 4.1 and, hence, some details are omitted. We again change the time scale $s \mapsto t \triangleq \psi_h^{-1}s$ on $\mathcal{D}(C)$ and observe that the re-parameterization θ_h and $\sigma_{h,t}$ does not alter the result of Lemma S.D.12. In addition, we have now,

$$\begin{aligned} dZ_{\psi,s}^{(1)} &= \psi_h^{-1/2}(ds)^{1/2} = (ds)^{1/2}, \\ dZ_{\psi,s}^{(2)} &= \psi_h^{-1/2}Y_{s-}ds = \psi_h^{-1/2}\widetilde{Y}_{s-}(ds)^{1/2} = \widetilde{Y}_{s-}(ds)^{1/2}, \end{aligned}$$

where the first equality in the second term above follows from $\widetilde{Y}_{(k-1)h} = h^{1/2}Y_{(k-1)h}$ on the old time scale. $N_b^0(v)$ varies on the time horizon $[N_b^0 - |v|\,,\, N_b^0 + |v|]$ as implied by $\mathcal{D}^*(C)$, as defined in Section 4. Again, in order to avoid clutter, we suppress the subscript ψ_h. We then have equation (S.60)-(S.61). Consider $T_b \leq T_b^0$ (i.e., $v \leq 0$). By Lemma S.D.12, there exists a $\overline{T}$ such that for all $T > \overline{T}$,

$$\begin{aligned} \overline{Q}_T(\theta^*) &= -h^{-1/2}\delta_h' Z_\Delta' Z_\Delta \delta_h + h^{-1/2}2\delta_h' Z_\Delta' e + o_p(1) \\ &= -\left(\delta^0\right)'\left(\sum_{k=T_b+1}^{T_b^0} z_{kh}z_{kh}'\right)\delta^0 \\ &\quad + 2\left(\delta^0\right)'\left(h^{-1/2}\sum_{k=T_b+1}^{T_b^0} z_{kh}\widetilde{e}_{kh}\right) + o_p(1), \end{aligned}$$

and note that this relationship corresponds to (S.62). As in the proof of Theorem 4.1 it is convenient to associate to the continuous time index t in $\mathcal{D}^*$, a corresponding $\mathcal{D}^*$-specific index t_v. We then define the following functions which belong to $\mathbb{D}(\mathcal{D}^*,\,\mathbb{R})$,

$$J_{Z,h}(v) \triangleq \sum_{k=T_b(v)+1}^{T_b^0} z_{kh}z_{kh}', \qquad J_{e,h}(v) \triangleq \sum_{k=T_b(v)+1}^{T_b^0} z_{kh}\widetilde{e}_{kh},$$

for $(T_b(v)+1)\,h \leq t_v < (T_b(v)+2)\,h$. Recall that the lower limit of the summation is $T_b(v)+1 = T_b^0 + \lfloor v/h \rfloor$ $(v \leq 0)$ and thus the number of observations in each sum increases at rate $1/h$. We first note that the partial sums of cross-products between the predictable and stochastic semimartingale regressors is null because the drift processes are of higher order (recall Lemma S.D.10). Given the previous lemma we can decompose $\overline{Q}_T(\theta,\, v)$ as follows,

$$\begin{aligned}\overline{Q}_T(\theta,\, v) &= \left(\delta_p^0\right)' R_{1,h}(v)\,\delta_p^0 + \left(\delta_Z^0\right)' R_{2,h}(v)\,\delta_Z^0 \\ &\quad + 2\left(\delta^0\right)'\left(\frac{1}{\sqrt{h}}\sum_{k=T_b+1}^{T_b^0} z_{kh}\widetilde{e}_{kh}\right),\end{aligned} \tag{S.85}$$

where

$$R_{1,h}(v) \triangleq \sum_{k=T_b(v)+1}^{T_b^0}\begin{bmatrix} h & Y_{(k-1)h}h^{3/2} \\ Y_{(k-1)h}h^{3/2} & \left(Y_{(k-1)h}h\right)^2\end{bmatrix}, \qquad R_{2,h}(v) \triangleq \left[Z_\Delta' Z_\Delta\right]_{\{\cdot,p\times p\}},$$

and δ^0 has been partitioned accordingly; that is, $\delta_p^0 = (\mu_\delta^0,\, \alpha_\delta^0)'$ is the vector of parameters associated with the predictable regressors whereas δ_Z^0 is the vector of parameters associated with the stochastic martingale regressors in Z. By standard results for convergence of Riemann sums,

$$\left(\delta_p^0\right)' R_{1,h}(v)\,\delta_p^0 \overset{\text{u.c.p.}}{\Rightarrow} \left(\delta_p^0\right)'\begin{bmatrix} N_b^0 - N_b & \int_{N_b^0+v}^{N_b^0}\widetilde{Y}_s ds \\ \int_{N_b^0+v}^{N_b^0}\widetilde{Y}_s ds & \int_{N_b^0+v}^{N_b^0}\widetilde{Y}_s^2 ds\end{bmatrix}\delta_p^0. \tag{S.86}$$

Next, since $Z_t^{(j)}$ $(j = 3, \ldots, p+2)$ is a continuous Itô semimartingale, we have by Theorem 3.3.1 in Jacod and Protter (2012),

$$R_{2,h}(v) \overset{\text{u.c.p.}}{\Rightarrow} \langle Z_\Delta,\, Z_\Delta\rangle(v)\,. \tag{S.87}$$

We now turn to examine the asymptotic behavior of the second term in (S.85) on $\mathcal{D}^*$. We use the following steps. First, we present a stable central limit theorem for each component of $Z_\Delta' e$. Second, we show the joint convergence stably in law to a continuous Gaussian process, and finally we verify tightness of the sequence of processes, which in turn yields the stable convergence under the uniform metric. We begin with the second element of $Z_\Delta' e$,

$$\frac{1}{\sqrt{h}}\sum_{k=T_b(v)+1}^{T_b^0}\alpha_\delta^0 z_{kh}^{(2)}\widetilde{e}_{kh} = \frac{1}{\sqrt{h}}\sum_{k=T_b(v)+1}^{T_b^0}\alpha_\delta^0\left(Y_{(k-1)h}h\right)\widetilde{e}_{kh},$$

and using $\widetilde{Y}_{(k-1)h} = h^{1/2}Y_{(k-1)h}$ [recall that $\widetilde{Y}_{(k-1)h}$ is bounded by the localization Assumption S.D.2] we then have

$$\begin{aligned}\frac{1}{\sqrt{h}}\sum_{k=T_b(v)+1}^{T_b^0}\alpha_\delta^0\left(Y_{(k-1)h}h\right)\widetilde{e}_{kh} &= \sum_{k=T_b(v)+1}^{T_b^0}\alpha_\delta^0\left(\widetilde{Y}_{(k-1)h}\right)\widetilde{e}_{kh} \\ &\overset{\text{u.c.p.}}{\Rightarrow} \int_{N_b^0+v}^{N_b^0}\alpha_\delta^0\widetilde{Y}_s dW_{e,s},\end{aligned}$$

which follows from the convergence of Riemann approximations for stochastic integrals [cf. Proposition

2.2.8 in Jacod and Protter (2012)]. For the first component, the argument is similar:

$$\frac{1}{\sqrt{h}}\sum_{k=T_b(v)+1}^{T_b^0}\mu_\delta^0 z_{kh}^{(1)}\widetilde{e}_{kh}\overset{\text{u.c.p.}}{\Rightarrow}\int_{N_b^0+v}^{N_b^0}\mu_\delta^0 dW_{e,s}. \tag{S.88}$$

Next, we consider the p-dimensional lower subvector of $Z'_\Delta e$, which can be written as

$$2\left(\delta_Z^0\right)'\left(\frac{1}{\sqrt{h}}\sum_{k=T_b(v)+1}^{T_b^0}\widetilde{z}_{kh}\widetilde{e}_{kh}\right), \tag{S.89}$$

where we have partitioned z_{kh} as $z_{kh}=\begin{bmatrix}h^{1/2} & Y_{(k-1)h}h & \widetilde{z}'_{kh}\end{bmatrix}'$. Then, note that the small-dispersion asymptotic re-parametrization implies that $\widetilde{z}_{kh}\widetilde{e}_{kh}$ corresponds to $z_{kh}\widetilde{e}_{kh}$ from Theorem 4.1. Hence, we shall apply the same arguments as in the proof of Theorem 4.1 since (S.89) is simply $2\left(\delta_Z^0\right)'$ times $\mathscr{W}_h(v)=h^{-1/2}J_{e,h}(v)$, where $J_{e,h}(v)\triangleq\sum_{k=T_b(v)+1}^{T_b^0}\widetilde{z}_{kh}\widetilde{e}$ with $(T_b(v)+1)h\leq t_v<(T_b(v)+2)h$. By Theorem 5.4.2 in Jacod and Protter (2012), $\mathscr{W}_h(v)\overset{\mathcal{L}\text{-s}}{\Rightarrow}\mathscr{W}_{Ze}(v)$. Since the convergence of the drift processes $R_{1,h}(v)$ and $R_{2,h}(v)$ occur in probability locally uniformly in time while $\mathscr{W}_h(v)$ converges stably in law to a continuous limit process, we have for each $(\theta,\,\cdot)$ a stable convergence in law under the uniform metric. This is a consequence of the property of stable convergence in law [cf. section VIII.5c in Jacod and Shiryaev (2003)]. Since the case $v>0$ is analogous, this proves the finite-dimensional convergence of the process $\overline{Q}_T(\theta,\,\cdot)$, for each θ. It remains to verify stochastic equicontinuity. As for the terms in $R_{1,h}(v)$, we can decompose

$$\left(\alpha_\delta\right)^2\left(\sum_{k=T_b(v)+1}^{T_b^0}\left(z_{kh}^{(2)}\right)^2-\left(\int_{N_b^0+v}^{N_b^0}\widetilde{Y}_s^2 ds\right)\right),$$

as

$$\overline{Q}_{6,T}(\theta,\,v)+\overline{Q}_{7,T}(\theta,\,v),$$

where

$$\overline{Q}_{6,T}(\theta,\,v)\triangleq\left(\alpha_\delta\right)^2\left(\sum_k\zeta_{2,h,k}^*\right),\quad\text{and}\quad\overline{Q}_{7,T}(\theta,\,v)\triangleq\left(\alpha_\delta\right)^2\left(\sum_k\zeta^{**}{}_{2,h,k}\right),$$

with

$$\begin{aligned}\zeta_{2,h,k}^*&\triangleq\left(z_{kh}^{(2)}\right)^2-\left(\int_{(k-1)h}^{kh}\widetilde{Y}_s^2 ds\right)-2\widetilde{Y}_{(k-1)h}\int_{(k-1)h}^{kh}\left(\widetilde{Y}_{(k-1)h}-\widetilde{Y}_s\right)ds\\&\quad+2\mathbb{E}\left[\widetilde{Y}_{(k-1)h}\left(\widetilde{Y}_{(k-1)h}\cdot h-\int_{(k-1)h}^{kh}\widetilde{Y}_s ds\right)\mid\mathscr{F}_{(k-1)h}\right]\\&\triangleq L_{1,h,k}+L_{2,h,k},\end{aligned}$$

and

$$\begin{aligned}\zeta_{2,h,k}^{**}&=2\widetilde{Y}_{(k-1)h}(\widetilde{Y}_{(k-1)h}h-\int_{(k-1)h}^{kh}\widetilde{Y}_s ds\\&\quad-\mathbb{E}\left[\left(\widetilde{Y}_{(k-1)h}h-\int_{(k-1)h}^{kh}\widetilde{Y}_s ds\right)\mid\mathscr{F}_{(k-1)h}\right]).\end{aligned}$$

Then, we have the following decomposition for

$$\overline{Q}_T^{\mathrm{c}}\left(\theta^*\right) \triangleq \overline{Q}_T\left(\theta^*\right) + \left(\delta^0\right)' \Lambda\left(v\right)\delta^0,$$

(if $v \leq 0$ and defined analogously for $v > 0$),

$$\overline{Q}_T^{\mathrm{c}}\left(\theta^*\right) = \sum_{r=1}^{9} \overline{Q}_{r,T}\left(\theta,\, v\right),$$

where $\overline{Q}_{r,T}\left(\theta,\, v\right)$, $r = 1, \ldots, 4$, are defined in (S.63) and $\overline{Q}_{5,T}\left(\theta,\, v\right) \triangleq \left(\mu_\delta\right)^2\left(\sum_k \zeta_{1,h,k}\right)$, $\overline{Q}_{8,T}\left(\theta,\, v\right) \triangleq \left(\mu_\delta\right)^2\left(h^{-1/2}\sum_k \xi_{1,h,k}\right)$, $\overline{Q}_{9,T}\left(\theta,\, v\right) \triangleq (\alpha_\delta)^2(h^{-1/2}\sum_k \xi_{2,h,k})$ where $\zeta_{1,h,k} \triangleq (z_{kh}^{(1)})^2 - h$, $\xi_{1,h,k} \triangleq h^{1/2}\widetilde{e}_{kh}$ and $\xi_{2,h,k} \triangleq (\widetilde{Y}_{(k-1)h}h^{1/2})\widetilde{e}_{kh}$. Moreover, recall that $\sum_k$ stands for $\sum_{T_{b(v)+1}}^{T_b^0}$ for $N_b\left(v\right) \in \mathcal{D}^*\left(C\right)$. Let us consider $\overline{Q}_{6,T}\left(\theta,\, v\right)$ first. For $s \in \left[\left(k-1\right)h,\, kh\right]$, by the Burkhölder-Davis-Gundy inequality

$$\left|\mathbb{E}\left[\widetilde{Y}_{(k-1)h}\left(\widetilde{Y}_{(k-1)h} - \widetilde{Y}_s\right) | \mathscr{F}_{(k-1)h}\right]\right| \leq Kh,$$

from which we can deduce that, using a maximal inequality for any $r > 1$,

$$\left[\mathbb{E}\left(\sup_{(\theta,\, v)}\left|\left(\alpha_\delta\right)^2\sum_k L_{2,h,k}\right|\right)^r\right]^{1/r} \leq K_r\left(\sup_{(\theta,\, v)}\left(\alpha_\delta\right)^{2r}\sum_k h^r\right)^{1/r} = K_r h^{\frac{r-1}{r}}. \tag{S.90}$$

By a Taylor series expansion for the mapping $f : y \rightarrow y^2$, and $s \in \left[\left(k-1\right)h,\, kh\right]$,

$$\mathbb{E}\left|\widetilde{Y}_{(k-1)h}^2 - \widetilde{Y}_s^2 - 2\widetilde{Y}_{(k-1)h}\left(\widetilde{Y}_{(k-1)h} - \widetilde{Y}_s\right)\right| \leq K\mathbb{E}\left[\left(\widetilde{Y}_{(k-1)h} - \widetilde{Y}_s\right)^2\right] \leq Kh,$$

where the second inequality follows from the Burkhölder-Davis-Gundy inequality. Thus, using a maximal inequality as in (S.90), we have for $r > 1$

$$\left[\mathbb{E}\left(\sup_{(\theta,\, v)}\left|\left(\alpha_\delta\right)^2\sum_k L_{1,h,k}\right|\right)^r\right]^{1/r} = K_r h^{\frac{r-1}{r}}. \tag{S.91}$$

(S.90) and (S.91) imply that $\overline{Q}_{6,T}\left(\cdot,\, \cdot\right)$ is stochastically equicontinuous. Next, note that $\overline{Q}_{7,T}\left(\theta,\, v\right)$ is a sum of martingale differences times $h^{1/2}$ (recall the definition of $\Delta_h\widetilde{V}_k = h^{1/2}\Delta_h V_k(\nu,\, \delta_{Z,1},\, \delta_{Z,2})$). Therefore by Assumption S.D.2, for any $0 \leq s < t \leq N$, $V_t - V_s = O_p\left(1\right)$ uniformly and therefore,

$$\sup_{(\theta,\, v)}\left|\overline{Q}_{7,T}\left(\theta,\, v\right)\right| \leq KO_p\left(h^{1/2}\right). \tag{S.92}$$

Given (S.86) and (S.90)-(S.92), we deduce that

$$\sup_{(\theta,\, v)}\left\{\left|\overline{Q}_{6,T}\left(\theta,\, v\right)\right| + \left|\overline{Q}_{7,T}\left(\theta,\, v\right)\right|\right\} = o_p\left(1\right).$$

As for the term involving $R_{1,h}\left(v\right)$, it is easy to see that $\sup_{(\theta,\, v)}|\overline{Q}\left(\theta,\, v\right)| \rightarrow 0$. Next, we can use some of the results in the proof of Theorem 4.1. In particular, the asymptotic stochastic equicontinuity of the sequence of processes $\{2\left(\delta_Z\right)' \mathscr{W}_h\left(v\right)\}$ follows from the same property as those applied to $\{\overline{Q}_{3,T}\left(\theta,\, v\right)\}$ and $\{\overline{Q}_{4,T}\left(\theta,\, v\right)\}$. The stochastic equicontinuity of

$$\left(\delta_Z\right)'\left(R_{2,h}\left(\theta,\, v\right) - \langle Z_\Delta,\, Z_\Delta\rangle\left(v\right)\right)\delta_Z,$$

also follows from the same proof. Recall $\overline{Q}_{1,T}(\theta,\, v)+\overline{Q}_{2,T}(\theta,\, v)$ as defined in (S.63). Thus, stochastic equicontinuity follows from (S.65) and the equation right before that. Next, let us consider $\overline{Q}_{9,T}(\theta,\, v)$. We use the alternative definition (ii) of stochastic equicontinuity in Andrews (1994). Consider any sequence $\{(\theta,\, v)\}$ and $\{(\bar{\theta},\, \bar{v})\}$ (we omit the dependence on h for simplicity). Assume $N_b \leq N_b^0 \leq \bar{N}_b$ (the other cases can be proven similarly) and let $Nd_h \triangleq \bar{N}_b - N_b$. Then,

$$\begin{aligned}\left|\overline{Q}_{9,T}(\theta,\, v)-\overline{Q}_{9,T}\left(\bar{\theta},\, \bar{v}\right)\right| &= \left|\alpha_\delta \sum_{k=T_b(v)+1}^{T_b^0} \widetilde{Y}_{(k-1)h}\widetilde{e}_{kh} - \bar{\alpha}_\delta \sum_{k=T_b^0}^{T_b(\bar{v})} \widetilde{Y}_{(k-1)h}\widetilde{e}_{kh}\right| \\ &\leq |\alpha_\delta| \left|\sum_{k=T_b(v)+1}^{T_b^0} \widetilde{Y}_{(k-1)h}\widetilde{e}_{kh}\right| \\ &\quad + |\bar{\alpha}_\delta| \left|\sum_{k=T_b^0}^{T_b(\bar{v})} \widetilde{Y}_{(k-1)h}\widetilde{e}_{kh}\right|.\end{aligned} \tag{S.93}$$

The second term is such that, by the Burkhölder-Davis-Gundy inequality for any $r\geq 1$,

$$\begin{aligned}&\mathbb{E}\left[\sup_{0\leq u\leq d_h}\left|\sum_{k=T_b^0}^{T_b^0+\lfloor Nu/h\rfloor} \widetilde{Y}_{(k-1)h}\widetilde{e}_{kh}\right|^r \mid \mathscr{F}_{N_b^0}\right] \\ &\leq K_r\left(Nd_h\right)^{r/2}\mathbb{E}\left[\frac{1}{Nd_h}\left(\sum_{k=T_b^0}^{T_b^0+\lfloor Nd_h/h\rfloor}\int_{(k-1)h}^{kh}\left(\widetilde{Y}_s\right)^2 ds\right)^{r/2} \mid \mathscr{F}_{N_b^0}\right] \leq K_r d_h^{r/2}.\end{aligned}$$

By the law of iterated expectations, and using the property that $d_h \downarrow 0$ in probability, we can find a T large enough such that for any $B>0$

$$\left(\mathbb{E}\left[\sup_{0\leq u\leq d_h}\left|\sum_{k=T_b^0}^{T_b^0+\lfloor Nu/h\rfloor} \widetilde{Y}_{(k-1)h}\widetilde{e}_{kh}\right|^r \mid \mathscr{F}_{N_b^0}\right]\right)^{1/r} \leq K_r d_h^{1/2} P\left(Nd_h > B\right) \rightarrow 0.$$

The argument for the first term in (S.93) is analogous. By Markov's inequality and combining the above steps, we have that for any $\varepsilon>0$ and $\eta>0$ there exists some $\overline{T}$ such that for $T>\overline{T}$,

$$P\left(\left|\overline{Q}_{9,T}(\theta,\, v)-\overline{Q}_{9,T}\left(\bar{\theta},\, \bar{v}\right)\right|>\eta\right)<\varepsilon.$$

Thus, the sequence $\{\overline{Q}_{9,T}(\cdot,\, \cdot)\}$ is stochastically equicontinuous. Noting that the same proof can be repeated for $\overline{Q}_{8,T}(\cdot,\, \cdot)$, we conclude that the sequence of processes $\{\overline{Q}_T^{c}(\theta^*),\, T\geq 1\}$ in (S.85) is stochastically equicontinuous. Furthermore, by (S.86) and (S.87) we obtain,

$$\left(\delta_p^0\right)' R_{1,h}(\theta,\, v)\,\delta_p^0 + \left(\delta_Z^0\right)'\left(R_{2,h}((\theta,\, v))\right)\delta_Z^0 \overset{\text{u.c.p.}}{\Rightarrow} \left(\delta^0\right)'\Lambda(v)\,\delta^0.$$

This suffices to guarantee the $\mathscr{G}$-stable convergence in law of the process $\{\overline{Q}_T(\cdot,\, \cdot),\, T\geq 1\}$ towards a process $\mathscr{W}(\cdot)$ with drift $\Lambda(\cdot)$ which, conditional on $\mathscr{G}$, is a two-sided Gaussian martingale process with covariance matrix given in (S.7). By definition, $\mathcal{D}^*(C)$ is compact and $Th(\widehat{\lambda}_{b,\pi}-\lambda_0)=O_p(1)$, which together with the fact that the limit process is a continuous Gaussian process enable one to deduce the main assertion from the continuous mapping theorem for the argmax functional. □

S.D.5.3 Proof of Proposition S.C.2

We begin with a few lemmas. Let $\widetilde{Y}_t^* \triangleq \widetilde{Y}_{\lfloor t/h \rfloor h}$. The first result states that the observed process $\{\widetilde{Y}_t^*\}$ converges to the non-stochastic process $\{\widetilde{Y}_t^0\}$ defined in (S.5) as $h \downarrow 0$. Assumption S.D.2 is maintained throughout and the constant $K > 0$ may vary from line to line.

Lemma S.D.13. *As $h \downarrow 0$, $\sup_{0 \le t \le N} |\widetilde{Y}_t^* - \widetilde{Y}_t^0| = o_p(1)$.*

Proof. Let us introduce a parameter γ_h with the property $\gamma_h \downarrow 0$ and $h^{1/2}/\gamma_h \to B$ where $B < \infty$. By construction, for $t < N_b^0$,

$$\widetilde{Y}_t - \widetilde{Y}_t^0 = \int_0^t \alpha_1^0 \left(\widetilde{Y}_s - \widetilde{Y}_s^0 \right) ds + B\gamma_h \left(\nu^0 \right)' D_t$$
$$+ B\gamma_h \left(\delta_{Z,1}^0 \right)' \int_0^t dZ_s + B\gamma_h \int_0^t \sigma_{e,s} dW_{e,s}.$$

We can use Cauchy-Schwarz's inequality, so that

$$\begin{aligned}
\left| \widetilde{Y}_t - \widetilde{Y}_t^0 \right|^2 &\le 2K \Bigg[\left| \int_0^t \alpha_1 \left(\widetilde{Y}_s - \widetilde{Y}_s^0 \right) ds \right|^2 \\
&\quad + \left(\left| \nu^{0\prime} D_t \right|^2 + \left| \delta_{Z,1}^{0\prime} \int_0^t dZ_s \right|^2 + \left| \int_0^t \sigma_{e,s} dW_{e,s} \right|^2 \right) (B\gamma_h)^2 \\
&\le 2Kt \Bigg[\left| \alpha_1^0 \right|^2 \int_0^t \left| \widetilde{Y}_s - \widetilde{Y}_s^0 \right|^2 ds + \Bigg(\sup_{0 \le s \le t} \left| \nu^{0\prime} D_s \right|^2 \\
&\quad + \sup_{0 \le s \le t} \left| \delta_{Z,1}^{0\prime} \int_0^t dZ_s \right|^2 + \sup_{0 \le s \le t} \left| \int_0^s \sigma_{e,u} dW_{e,u} \right|^2 \Bigg) (B\gamma_h)^2 \Bigg].
\end{aligned}$$

By Gronwall's inequality,

$$\begin{aligned}
\left| \widetilde{Y}_t - \widetilde{Y}_t^0 \right|^2 &\le 2 (B\gamma_h)^2 C \exp \left(\int_0^t 2K^2 t ds \right) \\
&\le 2 (B\gamma_h)^2 C \exp \left(2K^2 t^2 \right),
\end{aligned}$$

where $C < \infty$ is a bound on the sum of the supremum terms in the last equation above. The bound follows from Assumption S.D.2. Then,

$$\sup_{0 \le t \le N} \left| \widetilde{Y}_t - \widetilde{Y}_t^0 \right| \le K\sqrt{2} B\gamma_h \exp \left(K^2 N^2 \right) \to 0,$$

as $h \downarrow 0$ (and so $\gamma_h \downarrow 0$). The assertion then follows from $\lfloor t/h \rfloor h \to t$ as $h \downarrow 0$. The case with $t \ge N_b^0$ is proved in a similar fashion. □

Lemma S.D.14. *As $h \downarrow 0$, uniformly in (μ_1, α_1), $(N/T) \sum_{k=1}^{T_b^0} (\mu_1 + \alpha_1 \widetilde{Y}_{(k-1)h}) \xrightarrow{P} \int_0^{N_b^0} (\mu_1 + \alpha_1 \widetilde{Y}_s^0) ds$.*

Proof. Note that

$$\begin{aligned}
&\sup_{\mu_1, \alpha_1} \left| \frac{N}{T} \sum_{k=1}^{T_b^0} \left(\mu_1 + \alpha_1 \widetilde{Y}_{(k-1)h} \right) - \int_0^{N\lambda_0} \left(\mu_1 + \alpha_1 \widetilde{Y}_s^0 \right) \right| \\
&= \sup_{\mu_1, \alpha_1} \left| \int_0^{N_b^0} \left(\mu_1 + \alpha_1 \widetilde{Y}_s^* \right) ds - \int_0^{N_b^0} \left(\mu_1 + \alpha_1 \widetilde{Y}_s^0 \right) ds \right|
\end{aligned}$$

$$\leq \sup_{\alpha_1} \int_0^{N_b^0} |\alpha_1| \left| \widetilde{Y}_s^* - \widetilde{Y}_s^0 \right| ds \leq K O_p (\gamma_h) \sup_{\alpha_1} |\alpha_1| ,$$

which goes to zero as $h \downarrow 0$ by Lemma S.D.13 (recall $h^{1/2}/\gamma_h \to B$) and by Assumption S.D.2. □

Lemma S.D.15. *For each $3 \leq j \leq p+2$ and each θ, as $h \downarrow 0$,*

$$\sum_{k=1}^{\lfloor N_b^0/h \rfloor} \left(\mu_1 + \alpha_1 \widetilde{Y}_{(k-1)h} \right) \delta_{Z,1}^{(j)} \Delta_h Z_k^{(j)} \xrightarrow{P} \int_0^{N\lambda_0} \left(\mu_1 + \alpha_1 \widetilde{Y}_{(k-1)h}^0 \right) dZ_s^{(j)}.$$

Proof. Note that

$$\sum_{k=1}^{\lfloor N_b^0/h \rfloor} \left(\mu_1 + \alpha_1 \widetilde{Y}_{(k-1)h} \right) \delta_{Z,1}^{(j)} \Delta_h Z_k^{(j)} = \int_0^{N_b^0} \left(\mu_1 + \alpha_1 \widetilde{Y}_s^* \right) dZ_s^{(j)}.$$

By Markov's inequality and the dominated convergence theorem, for every $\varepsilon > 0$ and every $\eta > 0$:

$$\begin{aligned} &P\left(\left| \int_0^{N_b^0} \alpha_1 \left(\widetilde{Y}_s^* - \widetilde{Y}_s^0 \right) \delta_{Z,1}^{(j)} dZ_s^{(j)} \right| > \eta \right) \\ &\leq \frac{\left(\sup_{0 \leq s \leq N} \sum_{r=1}^{p} \left(\sigma_{Z,s}^{(j,r)} \right)^2 \right)^{1/2}}{\eta} |\alpha_1| \left| \delta_{Z,1}^{(j)} \right| \left(\int_0^{N_b^0} \mathbb{E} \left[\left(\widetilde{Y}_s^* - \widetilde{Y}_s^0 \right)^2 \right] ds \right)^{1/2} , \end{aligned}$$

which goes to zero as $h \downarrow 0$ in view of Lemma S.D.13 and Assumption S.D.2. □

Lemma S.D.16. *As $h \downarrow 0$, uniformly in μ_1, α_1,*

$$\sum_{k=1}^{T_b^0} \left(\mu_1 + \alpha_1 \widetilde{Y}_{(k-1)h} \right) \left(\widetilde{Y}_{kh} - \widetilde{Y}_{(k-1)h} - \left(\mu_1^0 + \alpha_1^0 \widetilde{Y}_{(k-1)h} \right) h \right) \xrightarrow{P} 0.$$

Proof. By definition [recall the notation in (S.4)],

$$\widetilde{Y}_{kh} - \widetilde{Y}_{(k-1)h} = \int_{(k-1)h}^{kh} \left(\mu_1^0 + \alpha_1^0 \widetilde{Y}_s \right) ds + \Delta_h \widetilde{V}_k \left(\nu^0,\, \delta_{Z,1}^0,\, \delta_{Z,2}^0 \right).$$

Then,

$$\begin{aligned} &\sum_{k=1}^{T_b^0} \left(\mu_1 + \alpha_1 \widetilde{Y}_{(k-1)h} \right) \left(\widetilde{Y}_{kh} - \widetilde{Y}_{(k-1)h} - \left(\mu_1^0 + \alpha_1^0 \widetilde{Y}_{(k-1)h} \right) h \right) \\ &\quad = \sum_{k=1}^{T_b^0} \int_{(k-1)h}^{kh} \left(\mu_1 + \alpha_1 \widetilde{Y}_{(k-1)h} \right) \left(\mu_1^0 + \alpha_1^0 \widetilde{Y}_s - \left(\mu_1^0 + \alpha_1^0 \widetilde{Y}_{(k-1)h} \right) \right) \\ &\quad\quad + \sum_{k=1}^{T_b^0} \int_{(k-1)h}^{kh} \left(\mu_1 + \alpha_1 \widetilde{Y}_{(k-1)h} \right) \Delta_h \widetilde{V}_k \left(\nu^0,\, \delta_{Z,1}^0,\, \delta_{Z,2}^0 \right) \\ &\quad = \int_0^{N_b^0} \left(\mu_1 + \alpha_1 \widetilde{Y}_{(k-1)h}^* \right) \left(\alpha_1^0 \left(\widetilde{Y}_s - \widetilde{Y}_{(k-1)h}^* \right) \right) ds \\ &\quad\quad + B \gamma_h \int_0^{N_b^0} \left(\mu_1 + \alpha_1 \widetilde{Y}_s^* \right) dV_s. \end{aligned}$$

For the first term on the right-hand side of the last equation

$$\begin{aligned}
&\sup_{\mu_1,\alpha_1}\left|\int_0^{N_b^0}\left(\mu_1+\alpha_1\widetilde{Y}_s^*\right)\left(\alpha_1^0\left(\widetilde{Y}_s-\widetilde{Y}_s^*\right)\right)ds\right| \\
&\quad\le\left|\alpha_1^0\right|\left|\int_0^{N_b^0}\sup_{\mu_1,\alpha_1}\left(\mu_1+\alpha_1\widetilde{Y}_s^*\right)\left(\widetilde{Y}_s-\widetilde{Y}_s^0+\widetilde{Y}_s^0-\widetilde{Y}_s^*\right)ds\right| \\
&\quad\le\left|\alpha_1^0\right|K\left(\int_0^{N_b^0}\sup_{0\le s\le N_b^0}\left|\widetilde{Y}_s-\widetilde{Y}_s^0\right|+\sup_{0\le s\le N_b^0}\left|\widetilde{Y}_s^0-\widetilde{Y}_s^*\right|ds\right),
\end{aligned}$$

which is $o_p(1)$ as $h\downarrow 0$ from Lemma S.D.13 and Assumption S.D.2. Next, consider the vector of regressors Z, and note that for any $3\le j\le p+2$,

$$\begin{aligned}
&B\gamma_h\sup_{\mu_1,\alpha_1}\left|\int_0^{N_b^0}\left(\mu_1+\alpha_1\widetilde{Y}_s^*\right)dZ_s^{(j)}\right| \\
&\quad\le B\gamma_h\sup_{\mu_1,\alpha_1}\left|\int_0^{N_b^0}\left(\mu_1+\alpha_1\widetilde{Y}_s^*\right)\sum_{r=1}^p\sigma_{Z,s}^{(j,r)}dW_Z^{(r)}\right|.
\end{aligned}$$

Let

$$R_{j,h}=R_{j,h}\left(\mu_1,\,\alpha_1\right)\triangleq\int_0^{N_b^0}B\gamma_h\left(\mu_1+\alpha_1\widetilde{Y}_s^*\right)\sum_{r=1}^p\sigma_{Z,s}^{(j,r)}dW_Z^{(r)},$$

(we index R_j by h because $\widetilde{Y}_s^*$ depends on h). Then, we want to show that, for every $\varepsilon>0$ and $K>0$,

$$P\left(\sup_{\mu_1,\alpha_1}\left|R_{j,h}\left(\mu_1,\,\alpha_1\right)\right|>K\right)\le\varepsilon. \tag{S.94}$$

In view of Chebyshev's inequality and the Itô's isometry,

$$\begin{aligned}
P\left(\left|R_{j,h}\right|>K\right)&\le\left(\frac{B\gamma_h}{K}\right)^2\mathbb{E}\left[\left|\int_0^{N_b^0}\left(R_{j,h}/\left(B\gamma_h\right)\right)\right|^2\right], \\
&\le\left[\sup_{0\le s\le N}\sum_{r=1}^p\left(\sigma_{Z,s}^{(j,r)}\right)^2\right]\left(\frac{B\gamma_h}{K}\right)^2\int_0^{N_b^0}\mathbb{E}\left[\left|\mu_1+\alpha_1\widetilde{Y}_s^*\right|^2ds\right],
\end{aligned}$$

so that by the boundness of the processes (cf. Assumption S.D.2) and the compactness of Θ_0, we have for some $A<\infty$,

$$P\left(\left|R_{j,h}\right|>K\right)\le A\left[\sup_{0\le s\le T}\sum_{r=1}^p\left(\sigma_{Z,s}^{(j,r)}\right)^2\right]\left(\frac{B\gamma_h}{K}\right)^2\to 0, \tag{S.95}$$

since $\gamma_h\downarrow 0$. This demonstrates pointwise convergence. It remains to show the stochastic equicontinuity of the sequence of processes $\{R_{j,h}(\cdot)\}$. Choose $2m>p$ and note that standard estimates for continuous Itô semimartingales result in $\mathbb{E}[|R_{j,h}|^{2m}]\le K$ which follows using the same steps that led to (S.95) with the Burkhölder-Davis-Gundy inequality in place of the Itô's isometry. Let $g(\widetilde{Y}_s^*,\,\widetilde{\theta})\triangleq\mu_{1,1}+\alpha_{1,1}\widetilde{Y}_s^*$, $\widetilde{\theta}_1\triangleq(\mu_{1,1},\alpha_{1,1})'$ and $\widetilde{\theta}_1\triangleq(\mu_{2,1},\alpha_{2,1})'$. For any $\widetilde{\theta}_1$, $\widetilde{\theta}_2$, first use the Burkhölder-Davis-Gundy inequality to yield,

$$
\begin{aligned}
&\mathbb{E}\left[\left|R_{j,h}\left(\widetilde{\theta}_2\right)-R_{j,h}\left(\widetilde{\theta}_1\right)\right|^{2m}\right]\\
&\leq\left(B\gamma_h\right)^{2m}K_m\left[\sup_{0\leq s\leq N}\sum_{r=1}^{p}\left(\sigma_{Z,s}^{(j,r)}\right)^2\right]^m\\
&\quad\times\mathbb{E}\left[\left(\int_0^{N_b^0}\left(g\left(\widetilde{Y}_s^*,\,\widetilde{\theta}_2\right)-g\left(\widetilde{Y}_s^*,\,\widetilde{\theta}_1\right)\right)^2ds\right)^m\right]\\
&\leq\left(B\gamma_h\right)^{2m}K_m\left[\sup_{0\leq s\leq N}\sum_{r=1}^{p}\left(\sigma_{Z,s}^{(j,r)}\right)^2\right]^m\\
&\quad\times\mathbb{E}\left[\left(\int_0^{N_b^0}\left(\left(\mu_{1,2}-\mu_{1,1}\right)+\left(\alpha_{1,2}-\alpha_{1,1}\right)\widetilde{Y}_s^*\right)^2ds\right)^m\right]
\end{aligned}
$$

$$
\begin{aligned}
&\leq\left(B\gamma_h\right)^{2m}K_m\left[\sup_{0\leq s\leq N}\sum_{r=1}^{p}\left(\sigma_{Z,s}^{(j,r)}\right)^2\right]^m\\
&\quad\times\mathbb{E}\left[\left(\int_0^{N_b^0}\left(\left(\mu_{1,2}-\mu_{1,1}\right)+\left(\alpha_{1,2}-\alpha_{1,1}\right)C\right)^2ds\right)^m\right]\\
&\leq\left(B\gamma_h\right)^{2m}K_m\mathbb{E}\left[\left(\int_0^{N_b^0}\left(2\left(\mu_{1,2}-\mu_{1,1}\right)^2+2C\left(\alpha_{1,2}-\alpha_{1,1}\right)^2\right)ds\right)^m\right]\\
&\leq 2^m\left(B\gamma_h\right)^{2m}K_m\left\|2\left(\widetilde{\theta}_2-\widetilde{\theta}_1\right)\right\|^{2m}\left(\int_0^{N_b^0}ds\right)^m\\
&\quad+2^m\left(B\gamma_h\right)^{2m}K\left(\widetilde{\theta}_1,\,\widetilde{\theta}_2,\,m,\,C\right),
\end{aligned}
\tag{S.96}
$$

where $C=\sup_{s\geq0}|\widetilde{Y}_s^*|$, $K(\widetilde{\theta}_1,\,\widetilde{\theta}_2,\,m,\,C)$ is some constant that depends on its arguments and we have used the fact that $(a+b)^2\leq 2a^2+2b^2$. Thus, since $\gamma_h\downarrow 0$, the mapping $R_{j,h}(\cdot)$ satisfies a Lipschitz-type condition [cf. Section 2 in Andrews (1992)]. This is sufficient for the asymptotic stochastic equicontiuity of $\{R_{j,h}(\cdot)\}$. Therefore, using Theorem 20 in Appendix I of Ibragimov and Has'minskiĭ (1981), (S.95) and (S.96) yield (S.94). Since the same result can be shown to remain valid for each term in the stochastic element $\Delta_h V_k(\nu,\,\delta_{Z,1},\,\delta_{Z,2})$, this establishes the claim. □

Proof of Proposition S.C.2. To avoid clutter, we prove the case for which the true parameters are $(\mu_1^0,\,\alpha_1^0)'$. The extension to parameters being local-to-zero is straightforward. The least-squares estimates of $(\mu_1^0,\,\alpha_1^0)'$ are given by,

$$
\widehat{\mu}_1\widehat{N}_b=\widetilde{Y}_{\widehat{N}_b}-\widetilde{Y}_0-\widehat{\alpha}_1h\sum_{k=1}^{\widehat{T}_b}\widetilde{Y}_{(k-1)h}
\tag{S.97}
$$

$$
\begin{aligned}
\widehat{\alpha}_1=&\frac{\sum_{k=1}^{\widehat{T}_b}\left(\widetilde{Y}_{kh}-\widetilde{Y}_{(k-1)h}\right)\widetilde{Y}_{(k-1)h}}{h\sum_{k=1}^{\widehat{T}_b}\widetilde{Y}_{(k-1)h}^2-\widehat{N}_b^{-1}\left(h\sum_{k=1}^{\widehat{T}_b}\widetilde{Y}_{(k-1)h}\right)^2}\\
&-\frac{\widehat{N}_b^{-1}\left(\widetilde{Y}_{\widehat{N}_b}-\widetilde{Y}_0\right)h\sum_{k=1}^{\widehat{T}_b}\widetilde{Y}_{(k-1)h}}{h\sum_{k=1}^{\widehat{T}_b}\widetilde{Y}_{(k-1)h}^2-\widehat{N}_b^{-1}\left(h\sum_{k=1}^{\widehat{T}_b}\widetilde{Y}_{(k-1)h}\right)^2}.
\end{aligned}
\tag{S.98}
$$

Then, assuming $\widehat{T}_b < T_b^0$,

$$\widehat{\alpha}_1 = \frac{\sum_{k=1}^{\widehat{T}_b}\left(\mu_1^0 h + \alpha_1^0 \widetilde{Y}_{(k-1)h} h + \Delta_h \widetilde{V}_{h,k}\right)\widetilde{Y}_{(k-1)h}}{h\sum_{k=1}^{\widehat{T}_b}\widetilde{Y}_{(k-1)h}^2 - \widehat{N}_b^{-1}\left(h\sum_{k=1}^{\widehat{T}_b}\widetilde{Y}_{(k-1)h}\right)^2}$$
$$-\frac{\left(\mu_1^0 + \alpha_1^0 \widehat{N}_b^{-1}\sum_{k=1}^{\widehat{T}_b}\widetilde{Y}_{(k-1)h}h + \widehat{N}_b^{-1}B\gamma_h\left(V_{\widehat{N}_b} - V_0\right)\right)h\sum_{k=1}^{\widehat{T}_b}\widetilde{Y}_{(k-1)h}}{h\sum_{k=1}^{\widehat{T}_b}\widetilde{Y}_{(k-1)h}^2 - \widehat{N}_b^{-1}\left(h\sum_{k=1}^{\widehat{T}_b}\widetilde{Y}_{(k-1)h}\right)^2}$$
$$+ o_p(1),$$

and thus

$$\widehat{\alpha}_1 = \frac{\sum_{k=1}^{T_b^0}\left(\mu_1^0 h + \alpha_1^0 \widetilde{Y}_{(k-1)h} h + \Delta_h \widetilde{V}_k\right)\widetilde{Y}_{(k-1)h}}{h\sum_{k=1}^{\widehat{T}_b}\widetilde{Y}_{(k-1)h}^2 - \widehat{N}_b^{-1}\left(h\sum_{k=1}^{\widehat{T}_b}\widetilde{Y}_{(k-1)h}\right)^2}$$
$$-\frac{\left(\mu_1^0 + \alpha_1^0 \widehat{N}_b^{-1}\sum_{k=1}^{T_b^0}\widetilde{Y}_{(k-1)h}h + \widehat{N}_b^{-1}B\gamma_h\left(V_{N_b^0} - V_0\right)\right)h\sum_{k=1}^{T_b^0}\widetilde{Y}_{(k-1)h}}{h\sum_{k=1}^{\widehat{T}_b}\widetilde{Y}_{(k-1)h}^2 - \widehat{N}_b^{-1}\left(h\sum_{k=1}^{\widehat{T}_b}\widetilde{Y}_{(k-1)h}\right)^2}$$
$$\times h\sum_{k=1}^{T_b^0}\widetilde{Y}_{(k-1)h}$$
$$-\frac{\sum_{k=\widehat{T}_b+1}^{T_b^0}\left(\mu_1^0 h + \alpha_1^0 \widetilde{Y}_{(k-1)h} h + \Delta_h \widetilde{V}_k\right)\widetilde{Y}_{(k-1)h}}{h\sum_{k=1}^{\widehat{T}_b}\widetilde{Y}_{(k-1)h}^2 - \widehat{N}_b^{-1}\left(h\sum_{k=1}^{\widehat{T}_b}\widetilde{Y}_{(k-1)h}\right)^2}$$
$$+\frac{\widehat{N}_b^{-1}\left(\sum_{k=\widehat{T}_b+1}^{T_b^0}\mu_1^0 h + \alpha_1^0\sum_{k=\widehat{T}_b+1}^{T_b^0}\widetilde{Y}_{(k-1)h}h + B\gamma_h\left(V_{N_b^0} - V_{\widehat{N}_b}\right)\right)}{h\sum_{k=1}^{\widehat{T}_b}\widetilde{Y}_{(k-1)h}^2 - \widehat{N}_b^{-1}\left(h\sum_{k=1}^{\widehat{T}_b}\widetilde{Y}_{(k-1)h}\right)^2}$$
$$\times h\sum_{k=\widehat{T}_b+1}^{T_b^0}\widetilde{Y}_{(k-1)h}.$$

By part (ii) of Theorem S.C.1, $N_b^0 - \widehat{N}_b = O_p(h^{1-\kappa})$, and thus it is easy to see that the third and fourth terms converge to zero in probability at a rate slower than $h^{1-\kappa}$. For the first and second terms, recalling that $\Delta_h \widetilde{V}_{h,k} = h^{1/2}\Delta V_{h,k}$ from (S.4), we have by ordinary convergence of approximations to Riemann sums, Lemma S.D.14 and the continuity of probability limits,

$$\alpha_1^0\sum_{k=1}^{T_b^0}\widetilde{Y}_{(k-1)h}h \xrightarrow{P} \alpha_1^0\int_0^{N_b^0}\widetilde{Y}_s ds, \qquad \sum_{k=1}^{T_b^0}\mu_1^0 h \xrightarrow{P} \mu_1^0\int_0^{N_b^0} ds,$$

and by Lemma S.D.15, $\sum_{k=1}^{T_b^0}\widetilde{Y}_{(k-1)h}\Delta_h\widetilde{V}_k \xrightarrow{P} 0$. Thus, we deduce that

$$\widehat{\alpha}_1 = \alpha_1^0 + O_p(B\gamma_h). \tag{S.99}$$

Using (S.99) into (S.97),

$$\begin{aligned}\widehat{\mu}_1 \widehat{N}_b &= \widetilde{Y}_{\widehat{N}_b} - \widetilde{Y}_0 - \alpha_1^0 h \sum_{k=1}^{\widehat{T}_b} \widetilde{Y}_{(k-1)h} - O_p\left(B\gamma_h\right), \\ &= \widetilde{Y}_{\widehat{N}_b} - \widetilde{Y}_0 - \alpha_1^0 h \sum_{k=1}^{T_b^0} \widetilde{Y}_{(k-1)h} - \alpha_1^0 h \sum_{k=\widehat{T}_b+1}^{T_b^0} \widetilde{Y}_{(k-1)h} - o_p\left(1\right).\end{aligned}$$

By part (ii) of Theorem S.C.1, the number of terms in the second sum above increases at rate T^{κ} and thus, $\alpha_1^0 h \sum_{k=\widehat{T}_b+1}^{T_b^0} \widetilde{Y}_{(k-1)h} = KO_p(h^{1-\kappa})$, where we have also used standard estimates for the drift arising from the Burkhölder-Davis-Gundy inequality. This gives

$$\widehat{\mu}_1 \widehat{N}_b = \widetilde{Y}_{N_b^0} - \widetilde{Y}_0 - \alpha_1^0 \int_0^{N_b^0} \widetilde{Y}_s ds - \alpha_1^0 O_p\left(h^{1-\kappa}\right) - o_p\left(1\right).$$

Noting that

$$\widetilde{Y}_{N_b^0} - \widetilde{Y}_0 = \mu_1^0 N_b^0 + \alpha_1^0 \int_0^{N_b^0} \widetilde{Y}_s ds + O_p\left(B\gamma_h\right)\left(V_{N_b^0} - V_0\right),$$

we have $\widehat{\mu}_1 N_b^0 = \mu_1^0 N_b^0 + O_p\left(B\gamma_h\right)\left(V_{N_b^0} - V_0\right)$, which yields

$$\widehat{\mu}_1 = \mu_1^0 + O_p\left(B\gamma_h\right). \tag{S.100}$$

Thus, as $h \downarrow 0$, $\widehat{\mu}_1$ is consistent for μ_1^0. The case where $\widehat{T}_b > T_b^0$ can be treated in the same fashion and is omitted. The consistency proof for $(\widehat{\mu}_2, \widehat{\alpha}_2)'$ is analogous and also omitted. The second step is to construct the least-squares residuals and scaling them up. The residuals are constructed as follows,

$$\widehat{u}_{kh} = \begin{cases} h^{-1/2}\left(\Delta_h \widetilde{Y}_k - \widehat{\mu}_1 \widetilde{x}_{kh}^{(1)} - \widehat{\alpha}_1 \widetilde{x}_{kh}^{(2)}\right), & k \leq \widehat{T}_b \\ h^{-1/2}\left(\Delta_h \widetilde{Y}_k - \widehat{\mu}_2 \widetilde{x}_{kh}^{(1)} - \widehat{\alpha}_2 \widetilde{x}_{kh}^{(2)}\right), & k > \widehat{T}_b, \end{cases}$$

where $\widetilde{x}_{kh}^{(1)} = h$ and $\widetilde{x}_{kh}^{(2)} = \widetilde{Y}_{(k-1)h} h$. This yields, for $k \leq T_b^0 \leq \widehat{T}_b$,

$$\widehat{u}_{kh} = h^{-1/2}\left(\mu_1^0 h + \alpha_1^0 \widetilde{Y}_{(k-1)h} h + B\gamma_h \Delta_h V_k - \widehat{\mu}_1 h - \widehat{\alpha}_1 \widetilde{Y}_{(k-1)h} h\right),$$

and using (S.99) and (S.100),

$$\begin{aligned}\widehat{u}_{kh} &= h^{-1/2}(\mu_1^0 h + \alpha_1^0 \widetilde{Y}_{(k-1)h} h + B\gamma_h \Delta_h V_k - \mu_1^0 h \\ &\quad - O_p\left(h^{3/2}\right) - \alpha_1^0 \widetilde{Y}_{(k-1)h} h - O_p\left(h^{3/2}\right)) \\ &= h^{-1/2} B\gamma \Delta_h V_k - O_p\left(h\right).\end{aligned} \tag{S.101}$$

Similarly, for $T_b^0 \leq \widehat{T}_b \leq k$,

$$\widehat{u}_{kh} = h^{-1/2} B\gamma_h \Delta_h V_k - O_p\left(h\right), \tag{S.102}$$

whereas for $\widehat{T}_b < k \leq T_b^0$,

$$\widehat{u}_{kh} = h^{-1/2}(\mu_1^0 h + \alpha_1^0 \widetilde{Y}_{(k-1)h} h + B\gamma_h \Delta_h V_k - \mu_2^0 h$$

$$
\begin{aligned}
&- O_p\left(h^{3/2}\right) - \alpha_2^0 \widetilde{Y}_{(k-1)h} h - O_p\left(h^{3/2}\right)) \\
&= h^{-1/2}\left(-\mu_\delta^0 h - \alpha_\delta^0 \widetilde{Y}_{(k-1)h} h + B\gamma_h \Delta_h V_k - O_p\left(h^{3/2}\right)\right) \\
&= -\mu_\delta^0 h^{1/2} - \alpha_\delta^0 \widetilde{Y}_{(k-1)h} h^{1/2} + h^{-1/2} B\gamma_h \Delta_h V_k - O_p\left(h\right). \qquad \text{(S.103)}
\end{aligned}
$$

Next, note that $\sum_{k=\widehat{T}_b+1}^{T_b^0} \mu_\delta^0 h^{1/2} \leq K h^{1/2-\kappa}$ and $\sum_{k=\widehat{T}_b+1}^{T_b^0} \alpha_\delta^0 \widetilde{Y}_{(k-1)h} h^{1/2} \leq K h^{1/2-\kappa}$ since by Theorem S.C.1-(ii) there are T^κ terms in each sum. Moreover, recall that $e_{kh} = \Delta_h e_k^* \sim \mathscr{N}(0,\, \sigma_{e,k-1}^2 h)$ and thus[7]

$$
\sum_{k=\widehat{T}_b+1}^{T_b^0} e_{kh} = \sqrt{h} \sum_{k=\widehat{T}_b+1}^{T_b^0} h^{-1/2} e_{kh} = h^{1/2-\kappa} o_p\left(1\right).
$$

Therefore, $\sum_{k=\widehat{T}_b+1}^{T_b^0} \widehat{u}_{kh} = K o_p(h^{1/2-\kappa})$. Since $\kappa \in (0,\, 1/2)$, this shows that the residuals $\widehat{u}_{kh}$ from equation (S.103) are asymptotically negligible. That is, asymptotically the estimator of $((\beta_S^0)',\, (\delta_{Z,1}^0)',\, (\delta_{Z,2}^0)')'$ minimizes (assuming $\widehat{T}_b \leq T_b^0$),

$$
\sum_{k=1}^{\widehat{T}_b} \left(\widehat{u}_{kh} - \widetilde{x}_{kh}' \beta_S\right)^2 + \sum_{k=T_b^0+1}^{T} \left(\widehat{u}_{kh} - \widetilde{x}_{kh}' \beta_S - \widetilde{z}_{0,kh}' \delta_S\right)^2 + o_p\left(1\right),
$$

where $X = [\widetilde{X}^{(1)},\, \beta^0 = \begin{bmatrix} \mu_1^0 & \alpha_1^0 & (\beta_S^0)' \end{bmatrix}'$, and Z_0 and δ_S^0 are partitioned accordingly. The subscript S indicates that these are the parameters of the stochastic semimartingale regressors. This is exactly the same regression model as in Proposition 3.3. Hence, the consistency result for the slope coefficients of the semimartingale regressors follows from the same proof. The following regression model estimated by least-squares provides consistent estimates for β_S^0 and δ_S^0: $\widehat{U} = \widetilde{X}\widehat{\beta}_S + \widehat{Z}_0 \widehat{\delta}_S + \text{residuals}$, where

$$
\widehat{Z}_0 = \begin{bmatrix}
\widetilde{z}_1^{(1)} & \cdots & \widetilde{z}_1^{(p)} \\
\vdots & \ddots & \vdots \\
\widetilde{z}_{\widehat{T}_b h}^{(1)} & \cdots & \widetilde{z}_{\widehat{T}_b h}^{(p)} \\
\widetilde{z}_{(T_b^0+1)h}^{(1)} & \cdots & \widetilde{z}_{(T_b^0+1)h}^{(p)} \\
\vdots & \ddots & \vdots \\
\widetilde{z}_N^{(1)} & \cdots & \widetilde{z}_N^{(p)}
\end{bmatrix},
$$

and $\widehat{U} = (\widehat{u}_{kh};\, k = 1, \ldots,\, \widehat{T}_b,\, T_b^0 + 1, \ldots,\, N)$. Therefore, using (S.101) and (S.102), we have

$$
\begin{aligned}
h^{-1/2} \begin{bmatrix} \widehat{\beta}_S - \beta^0 \\ \widehat{\delta}_S - \delta^0 \end{bmatrix} = {} & \begin{bmatrix} \widetilde{X}'\widetilde{X} & \widetilde{X}'\widehat{Z}_0 \\ \widehat{Z}_0'\widetilde{X} & \widehat{Z}_0'\widehat{Z}_0 \end{bmatrix}^{-1} \\
& \times h^{-1/2} \begin{bmatrix} \widetilde{X}'e & \widetilde{X}'\left(Z_0 - \widehat{Z}_0\right)\delta^0 + \widetilde{X}'AO_p\left(h\right) \\ \widehat{Z}_0'e & \widehat{Z}_0'\left(Z_0 - \widehat{Z}_0\right)\delta^0 + \widehat{Z}_0'AO_p\left(h\right) \end{bmatrix},
\end{aligned}
$$

for some matrix $A = O_p\left(1\right)$. It then follows by the same proof as in Proposition 3.3 that

$$
\begin{bmatrix} \widetilde{X}'\widetilde{X} & \widetilde{X}'\widehat{Z}_0 \\ \widehat{Z}_0'\widetilde{X} & \widehat{Z}_0'\widehat{Z}_0 \end{bmatrix}^{-1} \widetilde{X}'AO_p\left(h^{1/2}\right) = o_p\left(1\right), \qquad \text{(S.104)}
$$

[7]The same bound holds for the corresponding sum involving the other terms in $\Delta_h V_k$.

and

$$\begin{bmatrix} \widetilde{X}'\widetilde{X} & \widetilde{X}'\widehat{Z}_0 \\ \widehat{Z}_0'\widetilde{X} & \widehat{Z}_0'\widehat{Z}_0 \end{bmatrix}^{-1} \frac{1}{h^{1/2}}\widetilde{X}'\left(Z_0-\widehat{Z}_0\right)\delta^0 = O_p(1)\,o_p(1) = o_p(1). \tag{S.105}$$

The same arguments can be used for $\widehat{Z}_0'(Z_0-\widehat{Z}_0)\delta^0$ and $\widehat{Z}_0'AO_p(h)$. Therefore, in view of (S.99) and (S.100), we obtain $\widehat{\mu}_1=\mu_1^0+o_p(1)$ and $\widehat{\alpha}_1=\alpha_1^0+o_p(1)$, respectively, whereas (S.104) and (S.105) imply $\widehat{\beta}_S=\beta_S^0+o_p(1)$ and $\widehat{\delta}_S=\delta_S^0+o_p(1)$, respectively. Under the setting where the magnitude of the shifts is local to zero, we observe that by Proposition 4.1, $\widehat{N}_b-\widehat{N}_b^0=O_p(h^{1-\kappa})$ and one can follow the same steps that led to (S.99) and (S.100) and proceed as above. The final result is $\widehat{\theta}=\theta^0+o_p(1)$, which completes the proof. □

S.D.5.4 Negligibility of the Drift Term

Recall Lemma S.D.10 and apply the same proof as in Section S.D.4.10. Of course, the negligibility only applies to the drift processes $\mu_{\cdot,t}$ from (2.3) (i.e., only the drift processes of the semimartingale regressors) and not to μ_1^0, μ_2^0, α_1^0 or α_2^0. The steps are omitted since they are the same.

S.E Additional Simulations Results about HDR Confidence Sets

We continue with the analysis of finite-sample from Section 7. We consider discrete-time DGPs of the form

$$y_t = D_t'\nu^0 + Z_t'\beta^0 + Z_t'\delta_Z^0\mathbf{1}_{\{t>T_b^0\}} + e_t, \qquad t=1,\ldots,T, \tag{S.1}$$

with $T=100$ and, without loss of generality, $\nu^0=0$ (except for M5-M6, M8-M9). We consider eight versions of (S.1): M3 involves a break in the simultaneous mean and variance of an $i.i.d.$ series with $Z_t=1$ for all t, D_t absent, and $e_t=(1+\mathbf{1}_{\{t>T_b^0\}})u_t$ with $u_t\sim i.i.d.\,\mathscr{N}(0,1)$; M4 is the same as M1 but with stationary Gaussian AR(1) disturbances $e_t=0.3e_{t-1}+u_t$, $u_t\sim i.i.d.\,\mathscr{N}(0,0.49)$; M5 is a partial structural change model with $D_t=1$ for all t, $\nu^0=1$ and $Z_t=0.5Z_t+u_t$ with $u_t\sim i.i.d.\,\mathscr{N}(0,0.75)$ independent of $e_t\sim i.i.d.\,\mathscr{N}(0,1)$; M6 is similar to M5 but with $u_t\sim i.i.d.\,\mathscr{N}(0,1)$ and heteroskedastic disturbances given by $e_t=v_t|Z_t|$ where v_t is a sequence of $i.i.d.\,\mathscr{N}(0,1)$ random variables independent of $\{Z_t\}$; M7 is the same as M4 but with u_t drawn from a t_υ distribution with $\upsilon=5$ degrees of freedom; M8 is a model with a lagged dependent variable with $D_t=y_{t-1}$, $Z_t=1$, $e_t\sim i.i.d.\,\mathscr{N}(0,0.49)$, $\nu^0=0.3$ and $Z_t'\delta_Z^0\mathbf{1}_{\{t>T_b^0\}}$ is replaced by $Z_t'(1-\nu^0)\,\delta_Z^0\mathbf{1}_{\{t>T_b^0\}}$; M9 has FIGARCH(1,d,1) errors given by $e_t=\sigma_t u_t$, $u_t\sim i.i.d.\,\mathscr{N}(0,1)$ and $\sigma_t=0.1+(1-0.2L(1-L)^d)e_t^2$ where $d=0.6$ is the order of differencing and L the lag operator, $D_t=1$, $\nu^0=1$ and $Z_t\sim i.i.d.\,\mathscr{N}(1,1.44)$ independent of e_t. M10 is similar to M6 but with an ARFIMA(0.3, d, 0) regressor Z_t with order of differencing $d=0.5$, $\mathrm{Var}(Z_t)=1$ and $e_t\sim\mathscr{N}(0,1)$ independent of $\{Z_t\}$. We set $\beta^0=1$ in all models, except in M8 where $\beta^0=0$. The Results are reported in Table 5-12.

Table 5: Small-sample coverage rate and length of the confidence set for model M3

		$\delta_Z^0 = 0.3$		$\delta_Z^0 = 0.6$		$\delta_Z^0 = 1$		$\delta_Z^0 = 2$	
		Cov.	Lgth.	Cov.	Lgth.	Cov.	Lgth.	Cov.	Lgth.
$\lambda_0 = 0.5$	HDR	0.970	86.65	0.937	76.29	0.901	55.59	0.934	26.11
	Bai (1997)	0.854	70.60	0.843	58.27	0.857	40.70	0.923	14.24
	$\widehat{U}_T$.neq	0.961	88.95	0.961	80.33	0.961	61.15	0.964	32.16
	ILR	0.989	92.53	0.985	84.06	0.977	58.05	0.958	12.31
$\lambda_0 = 0.35$	HDR	0.976	89.81	0.961	83.26	0.935	64.87	0.934	26.11
	Bai (1997)	0.823	69.86	0.822	55.87	0.844	38.91	0.932	14.24
	$\widehat{U}_T$.neq	0.963	89.84	0.963	82.26	0.961	65.87	0.964	32.16
	ILR	0.990	93.48	0.985	88.69	0.982	68.23	0.977	15.45
$\lambda_0 = 0.2$	HDR	0.978	90.39	0.975	85.89	0.934	70.05	0.957	29.63
	Bai (1997)	0.782	70.24	0.805	56.37	0.831	37.66	0.928	14.80
	$\widehat{U}_T$.neq	0.968	91.11	0.968	87.62	0.972	78.17	0.967	46.24
	ILR	0.980	93.32	0.981	91.60	0.978	81.60	0.981	22.60

The model is $y_t = \beta^0 + \delta_Z^0 \mathbf{1}_{\{t > \lfloor T\lambda_0 \rfloor\}} + e_t$, $e_t = \left(1 + \mathbf{1}_{\{t > \lfloor T\lambda_0 \rfloor\}}\right) u_t$, $u_t \sim i.i.d.\, \mathscr{N}(0,\, 1)$, $T = 100$. The notes of Table 2 apply.

Table 6: Small-sample coverage rate and length of the confidence set for model M4

		$\delta_Z^0 = 0.3$		$\delta_Z^0 = 0.6$		$\delta_Z^0 = 1$		$\delta_Z^0 = 2$	
		Cov.	Lgth.	Cov.	Lgth.	Cov.	Lgth.	Cov.	Lgth.
$\lambda_0 = 0.5$	HDR	0.904	72.44	0.901	57.37	0.919	29.70	0.971	5.85
	Bai (1997)	0.833	66.34	0.834	41.32	0.895	18.63	0.969	5.49
	$\widehat{U}_T$.eq	0.958	87.16	0.968	71.47	0.958	45.82	0.957	28.01
	ILR	0.932	79.38	0.944	53.48	0.966	21.98	0.993	4.87
$\lambda_0 = 0.35$	HDR	0.910	70.98	0.902	53.88	0.917	28.07	0.973	5.99
	Bai (1997)	0.849	65.13	0.840	40.43	0.900	18.69	0.974	5.49
	$\widehat{U}_T$.eq	0.960	87.46	0.961	72.79	0.962	46.44	0.961	28.03
	ILR	0.942	80.94	0.946	55.20	0.965	23.55	0.993	4.93
$\lambda_0 = 0.2$	HDR	0.905	72.26	0.913	50.61	0.933	25.07	0.973	6.35
	Bai (1997)	0.829	65.56	0.899	41.42	0.932	19.62	0.966	5.55
	$\widehat{U}_T$.eq	0.962	88.77	0.968	78.61	0.963	57.87	0.965	29.88
	ILR	0.938	83.24	0.951	63.66	0.972	28.94	0.994	5.16

The model is $y_t = \beta^0 + \delta_Z^0 \mathbf{1}_{\{t > \lfloor T\lambda_0 \rfloor\}} + e_t$, $e_t = 0.3 e_{t-1} + u_t$, $u_t \sim i.i.d.\, \mathscr{N}(0,\, 0.49)$, $T = 100$. The notes of Table 2 apply.

Table 7: Small-sample coverage rate and length of the confidence set for model M5

		$\delta_Z^0 = 0.3$		$\delta_Z^0 = 0.6$		$\delta_Z^0 = 1$		$\delta_Z^0 = 2$	
		Cov.	Lgth.	Cov.	Lgth.	Cov.	Lgth.	Cov.	Lgth.
$\lambda_0 = 0.5$	HDR	0.915	77.14	0.912	61.71	0.910	30.64	0.912	7.15
	Bai (1997)	0.805	65.94	0.821	44.07	0.850	20.71	0.887	5.96
	$\widehat{U}_T$.eq	0.950	85.23	0.951	67.40	0.951	39.87	0.955	17.46
	ILR	0.961	84.37	0.966	59.94	0.977	26.09	0.986	7.14
$\lambda_0 = 0.35$	HDR	0.915	75.53	0.911	58.88	0.905	29.77	0.912	7.27
	Bai (1997)	0.821	64.69	0.826	42.93	0.849	20.77	0.888	5.99
	$\widehat{U}_T$.eq	0.948	85.48	0.948	68.95	0.948	41.40	0.954	17.57
	ILR	0.959	84.67	0.964	61.55	0.973	27.70	0.987	7.13
$\lambda_0 = 0.2$	HDR	0.911	74.46	0.931	56.22	0.935	29.22	0.929	7.85
	Bai (1997)	0.820	64.06	0.870	42.86	0.896	22.11	0.887	6.16
	$\widehat{U}_T$.eq	0.952	86.80	0.956	75.20	0.952	51.99	0.952	19.92
	ILR	0.961	86.03	0.964	68.69	0.978	36.34	0.985	7.51

The model is $y_t = \nu^0 + Z_t\beta^0 + Z_t\delta_Z^0\mathbf{1}_{\{t>\lfloor T\lambda_0\rfloor\}} + e_t$, $X_t = 0.5X_{t-1} + u_t$, $u_t \sim i.i.d.\,\mathscr{N}(0,\,0.75)$, $e_t \sim$ i.i.d.$\mathscr{N}(0,\,1)$, $T = 100$. The notes of Table 2 apply.

Table 8: Small-sample coverage rate and length of the confidence set for model M6

		$\delta_Z^0 = 0.3$		$\delta_Z^0 = 0.6$		$\delta_Z^0 = 1$		$\delta_Z^0 = 2$	
		Cov.	Lgth.	Cov.	Lgth.	Cov.	Lgth.	Cov.	Lgth.
$\lambda_0 = 0.5$	HDR	0.920	77.03	0.923	70.69	0.930	60.02	0.969	35.03
	Bai (1997)	0.690	56.73	0.716	41.63	0.783	27.53	0.885	12.70
	$\widehat{U}_T$.eq	0.962	87.76	0.962	78.32	0.962	63.80	0.962	40.82
	ILR	0.790	71.07	0.805	59.66	0.824	40.78	0.909	11.63
$\lambda_0 = 0.35$	HDR	0.928	76.41	0.925	68.21	0.933	56.17	0.964	31.73
	Bai (1997)	0.691	55.18	0.720	40.25	0.757	26.90	0.883	12.62
	$\widehat{U}_T$.eq	0.953	87.76	0.953	78.55	0.953	64.81	0.953	41.98
	ILR	0.795	71.34	0.804	60.48	0.832	30.42	0.903	10.78
$\lambda_0 = 0.2$	HDR	0.915	75.86	0.919	66.79	0.926	52.50	0.957	27.46
	Bai (1997)	0.707	55.03	0.770	39.77	0.828	26.82	0.901	12.68
	$\widehat{U}_T$.eq	0.951	88.48	0.952	82.09	0.954	71.84	0.950	50.72
	ILR	0.795	72.01	0.809	62.75	0.829	45.18	0.913	12.62

The model is $y_t = \nu^0 + Z_t\beta^0 + Z_t\delta_Z^0\mathbf{1}_{\{t>\lfloor T\lambda_0\rfloor\}} + e_t$, $e_t = v_t\,|Z_t|$, $v_t \sim i.i.d.\,\mathscr{N}(0,\,1)$, $Z_t = 0.5Z_{t-1} + u_t$, $u_t \sim i.i.d.\,\mathscr{N}(0,\,1)$ $T = 100$. The notes of Table 2 apply.

Table 9: Small-sample coverage rate and length of the confidence set for model M7

		$\delta_Z^0 = 0.3$		$\delta_Z^0 = 0.6$		$\delta_Z^0 = 1$		$\delta_Z^0 = 2$	
		Cov.	Lgth.	Cov.	Lgth.	Cov.	Lgth.	Cov.	Lgth.
$\lambda_0 = 0.5$	HDR	0.918	75.64	0.910	67.46	0.931	48.54	0.957	12.50
	Bai (1997)	0.834	70.13	0.824	52.16	0.861	28.69	0.948	8.45
	$\widehat{U}_T$.eq	0.959	88.62	0.959	78.87	0.959	58.60	0.952	30.15
	ILR	0.969	86.75	0.959	67.91	0.967	34.13	0.995	9.17
$\lambda_0 = 0.35$	HDR	0.926	74.78	0.914	64.86	0.924	45.69	0.956	12.25
	Bai (1997)	0.851	69.35	0.847	51.17	0.878	28.59	0.944	8.47
	$\widehat{U}_T$.eq	0.964	88.82	0.960	79.74	0.964	60.26	0.964	30.64
	ILR	0.972	88.69	0.975	73.95	0.981	39.08	0.992	9.08
$\lambda_0 = 0.2$	HDR	0.909	78.12	0.921	61.87	0.933	40.66	0.961	11.70
	Bai (1997)	0.824	65.23	0.867	51.35	0.915	29.83	0.955	8.70
	$\widehat{U}_T$.eq	0.961	89.71	0.960	83.68	0.961	69.25	0.960	35.78
	ILR	0.966	91.48	0.971	82.78	0.984	51.93	0.995	10.87

The model is $y_t = \beta^0 + \delta_Z^0 \mathbf{1}_{\{t > \lfloor T\lambda_0 \rfloor\}} + e_t$, $e_t = 0.3e_{t-1} + u_t$, $u_t \sim i.i.d.\, \mathrm{t}_v$, $v = 5$, $T = 100$. The notes of Table 2 apply.

Table 10: Small-sample coverage rate and length of the confidence set for model M8

		$\delta_Z^0 = 0.3$		$\delta_Z^0 = 0.6$		$\delta_Z^0 = 1$		$\delta_Z^0 = 2$	
		Cov.	Lgth.	Cov.	Lgth.	Cov.	Lgth.	Cov.	Lgth.
$\lambda_0 = 0.5$	HDR	0.918	75.08	0.913	60.44	0.931	32.30	0.965	6.34
	Bai (1997)	0.778	60.94	0.815	38.14	0.885	17.29	0.949	5.34
	$\widehat{U}_T$.eq	0.949	84.56	0.950	67.64	0.953	42.95	0.950	30.25
	ILR	0.943	83.69	0.946	63.24	0.956	32.85	0.982	10.49
$\lambda_0 = 0.35$	HDR	0.919	74.16	0.916	58.53	0.931	32.10	0.965	6.48
	Bai (1997)	0.799	60.25	0.814	37.94	0.872	17.49	0.952	5.35
	$\widehat{U}_T$.eq	0.951	85.01	0.948	69.14	0.957	48.40	0.949	30.31
	ILR	0.946	84.12	0.944	63.99	0.960	33.45	0.977	8.71
$\lambda_0 = 0.2$	HDR	0.912	73.43	0.929	56.18	0.949	31.23	0.965	6.96
	Bai (1997)	0.795	59.43	0.864	38.17	0.910	18.52	0.954	5.34
	$\widehat{U}_T$.eq	0.950	86.94	0.951	76.52	0.946	55.72	0.947	38.80
	ILR	0.945	83.94	0.953	63.55	0.963	32.41	0.982	15.01

The model is $y_t = \delta_Z^0 \left(1 - \nu^0\right) \mathbf{1}_{\{t > \lfloor T\lambda_0 \rfloor\}} + \nu^0 y_{t-1} + e_t$, $e_t \sim i.i.d.\, \mathscr{N}\left(0,\, 0.49\right)$, $\nu^0 = 0.3$, $T = 100$. The notes of Table 2 apply.

Table 11: Small-sample coverage rate and length of the confidence sets for model M9

		$\delta_Z^0 = 0.3$		$\delta_Z^0 = 0.6$		$\delta_Z^0 = 1$		$\delta_Z^0 = 2$	
		Cov.	Lgth.	Cov.	Lgth.	Cov.	Lgth.	Cov.	Lgth.
$\lambda_0 = 0.5$	HDR	0.903	61.09	0.927	31.14	0930	18.33	0.930	9.10
	Bai (1997)	0.791	37.86	0.831	17.73	0.855	10.43	0.868	5.30
	$\widehat{U}_T$.eq	0.947	65.23	0.947	39.76	0.947	28.82	0.947	20.36
	ILR	0.909	72.62	0.946	45.06	0.962	23.97	0.978	9.34
$\lambda_0 = 0.35$	HDR	0.904	60.58	0.918	30.96	0.904	18.16	0.928	0.34
	Bai (1997)	0.791	37.70	0.829	18.04	0.852	10.61	0.870	5.34
	$\widehat{U}_T$.eq	0.942	66.27	0.942	40.63	0.942	29.39	0.942	20.67
	ILR	0.922	72.20	0.947	45.27	0.959	24.93	0.973	8.55
$\lambda_0 = 0.2$	HDR	0.920	61.37	0.946	31.00	0.942	20.44	0.944	9.04
	Bai (1997)	0.791	39.23	0.841	19.28	0.876	11.99	0.886	6.16
	$\widehat{U}_T$.eq	0.934	71.42	0.931	47.53	0.934	34.12	0934	24.06
	ILR	0.920	72.68	0.935	49.61	0.959	27.90	0.972	10.01

The model is $y_t = \nu^0 + Z_t\beta^0 + Z_t\delta_Z^0\mathbf{1}_{\{t>\lfloor T\lambda_0\rfloor\}} + e_t$, $Z_t \sim i.i.d.\,\mathscr{N}\,(1,\,1.44)\,,\,\{e_t\}$ follows a FIGARCH(1,0.6,1) process and $T = 100$. The notes of Table 2 apply.

Table 12: Small-sample coverage rate and length of the confidence set for model M10

		$\delta_Z^0 = 0.3$		$\delta_Z^0 = 0.6$		$\delta_Z^0 = 1$		$\delta_Z^0 = 2$	
		Cov.	Lgth.	Cov.	Lgth.	Cov.	Lgth.	Cov.	Lgth.
$\lambda_0 = 0.5$	HDR	0.952	74.84	0.930	36.02	0.921	13.11	0.916	4.34
	Bai (1997)	0.809	45.33	0.844	17.11	0.864	8.27	0.883	3.61
	$\widehat{U}_T$.eq	0.959	72.69	0.959	39.81	0.959	24.25	0.959	14.79
	ILR	0.929	83.23	0.951	69.67	0.971	44.40	0.987	10.44
$\lambda_0 = 0.35$	HDR	0.934	73.08	0.937	35.37	0.923	13.68	0.920	4.55
	Bai (1997)	0.821	45.70	0.838	17.78	0.867	8.53	0.889	3.71
	$\widehat{U}_T$.eq	0.964	76.14	0.964	44.61	0.965	27.33	0.964	15.84
	ILR	0.934	81.32	0.959	62.98	0.977	34.38	0.984	9.12
$\lambda_0 = 0.2$	HDR	0.941	71.46	0.959	59.03	0.950	15.39	0.919	5.03
	Bai (1997)	0.818	47.82	0.872	20.44	0.878	9.60	0.873	3.92
	$\widehat{U}_T$.eq	0.971	82.40	0.971	59.03	0.971	39.02	0.972	20.42
	ILR	0.928	83.26	0.952	70.03	0.964	42.65	0.982	10.30

The model is $y_t = \nu^0 + Z_t\beta^0 + Z_t\delta_Z^0\mathbf{1}_{\{t>\lfloor T\lambda_0\rfloor\}} + e_t$, $e_t \sim i.i.d.\,\mathscr{N}\,(0,\,1)\,,\,Z_t \sim \text{ARFIMA}\,(0.3,\,0.6,\,0)\,,\,T = 100$. The notes of Table 2 apply.

References

Andrews, D.W.K., 1992. Generic Uniform Convergence. Econometric Theory 8, 241–257.

Andrews, D.W.K., 1994. Empirical Process Methods in Econometrics, in: Engle, R.F., McFadden, D.L. (Eds.), Handbook of Econometrics. Amsterdam: Elsevier Science. volume 4. chapter 37, pp. 2247–2294.

Bai, J., 1997. Estimation of a Change-Point in Multiple Regression Models. The Review of Economics and Statistics 79, 551–563.

Bai, J., Perron, P., 1998. Estimating and Testing Linear Models with Multiple Structural Changes. Econometrica 66, 47–78.

Barndorff-Nielsen, O.E., Shephard, N., 2004. Econometric Analysis of Realised Covariation: High Frequency Based Covariance, Regression and Correlation in Financial Economics. Econometrica 72, 885–925.

Casini, A., Perron, P., 2020. Continuous Record Asymptotics for Sructural Change Models. Extended working paper arXiv preprint arXiv:1803.10881 .

Casini, A., Perron, P., 2021. Continuous Record Asymptotics for Change-Point Models. arXiv preprint arXiv:1803.10881 .

Christopeit, N., 1986. Quasi-Least-Squares Estimation in Semimartingale Regression Models. Stochastics 16, 255–278.

Galtchouk, L., Konev, V., 2001. On Sequential Estimation of Parameters in Semimartingale Regression Models with Continuous Time Parameter. Annals of Statistics 29, 1508–1536.

Ibragimov, A., Has'minskiĭ, R.Z., 1980. On the estimates of the signal, its derivatives and the point of maximum for gaussian observations. Theory of Probability and its Applications 25.

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

Jacod, J., Protter, P., 2012. Discretization of Processes. Berlin Heidelberg: Springer.

Jacod, J., Rosenbaum, M., 2013. Quarticity and Other Functionals of Volatility: Efficient Estimation. The Annals of Statistics 41, 1462–1484.

Jacod, J., Shiryaev, A.N., 2003. Limit Theorems for Stochastic Processes. Berlin Heidelberg: Springer-Verlag.

Jiang, L., Wang, X., Yu, S., 2018. New Distribution Theory for the Estimation of Structural Break Point in Mean. Journal of Econometrics 205, 156–176.

Lai, T.L., Wei, C.Z., 1983. Asymptotic Properties of General Autoregressive Models and Strong Consistency of Least-Squares Estimates of Their Parameters. Journal of Multivariate Analysis 13, 1–23.

Laredo, C.F., 1990. A Sufficient Condition for Asymptotic Sufficiency of Incomplete Observations of a Diffusion Process. The Annals of Statistics 18, 1158–1171.

Li, J., Todorov, V., Tauchen, G., 2017. Adaptive estimation of continuous-time regression models using high-frequency data. Journal of Econometrics 200, 36–47.

Mel'nikov, A.V., Novikov, A.A., 1988. Sequential Inference with Fixed Accuracy for Semimartingales. Theory of Probability and its Applications 33, 480–494.

Sørensen, M., Uchida, M., 2003. Small-diffusion Asymptotics for Discretely Sampled Stochastic Differential Equations. Bernoulli 9, 1051–1069.