Continuous Record Asymptotics for Structural Change Models*

Alessandro Casini† Pierre Perron‡

Boston University Boston University

30th November 2017

First Version: 28th October 2015

Abstract

For a partial structural change in a linear regression model with a single break, 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 $T \rightarrow \infty$ with $h \downarrow 0$ while keeping the time span N fixed. We impose very mild regularity conditions on an underlying continuous-time model assumed to generate the data. 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, semi-martingale.

---

*We wish to thank Mark Podolskij for helpful suggestions. We also thank Christian Gouriéroux and Viktor Todorov for related discussions on some aspects of this project. We are grateful to Iván Fernández-Val, Hiro Kiado, Zhongjun Qu as well as seminar participants at Boston University 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. Casini gratefully acknowledges partial financial support from the Bank of Italy.

†Corresponding author at: Department of Economics, Boston University, 270 Bay State Rodad, Boston, MA 02215, US. Email: acasini@bu.edu.

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

1 Introduction

Parameter instability in linear regression models is a common problem and more so when the span of the data is large. In the context of a partial structural change in 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 asymptotic framework whereby T increases by shrinking the time interval h to zero while keeping time span N fixed. We impose very mild conditions on an underlying continuous-time model assumed to generate the data, basically continuous Itô semimartingales. Using an infill asymptotic setting, the uncertainty about the unknown parameters is assessed from the sample paths of the processes, which differs from the standard large-N asymptotics whereby it is assessed from features of the distributions or moments of the processes. This allows us to impose mild pathwise regularity conditions and to avoid any ergodic or weak-dependence assumption. Our setting includes most linear models considered in the structural change literature based on large-N asymptotics, which essentially involve processes satisfying some form of mixing conditions.

An extensive amount of research addressed structural change 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), Bai and Perron (1998) and Perron and Qu (2006). Qu and Perron (2007) generalized this work by considering multivariate regressions. Extensions to models with endogenous regressors were considered by Perron and Yamamoto (2014) [see also Hall, Han, and Boldea (2010)], though Perron and Yamamoto (2015) argue that standard least-squares methods are still applicable, and indeed preferable, in such cases. Notable also are the contributions on testing for structural changes by Hawkins (1977), Picard (1985), Kim and Siegmund (1989), Andrews (1993), Horváth (1993), Andrews and Ploberger (1994) and Bai and Perron (1998), among others. See the reviews of Csörgő and Horváth (1997), Perron (2006) and references therein. In this literature, the resulting large-N limit theory for the estimate of the break date depends on the exact distribution of the regressors and disturbances. Therefore, a so-called shrinkage asymptotic theory was adopted whereby the magnitude of the shift converges to zero as T increases, which leads to a limit distribution invariant to the distributions of the regressors and errors.

We study a general change-point problem under a continuous record asymptotic framework and develop inference procedures based on the derived asymptotic distribution. As $h \downarrow 0$, identification of the break point translates to the detection of a change in the slope coefficients for the continuous local martingale part of locally square-integrable semimartingales. 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. The properties of this limit theory, in particular how the magnitude of the shift and how the span versus the sample size affect the precision of the break date estimate are then discussed. The knowledge of such features of the distribution of the estimator is important from a theoretical perspective and cannot be gained from the classical large-N asymptotics. It is also very useful to provide guidelines as to the proper method to use to construct confidence sets.

Our continuous record limit distribution 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 sample 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 at T increases and, given the mixing assumptions imposed, only the neighborhood around the break date remains relevant. Furthermore, the domain of the extremum depends on the position of the break point N_b^0 and therefore 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 in small samples, 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. These results are natural consequences of our continuous record asymptotic theory, which indicate that the time span, location and magnitude of the break and statistical properties of the errors and regressors all jointly play a primary role in shaping the limit distribution of the break date estimator. For example, when the shift magnitude is small, the probability density displays three modes. As the shift magnitude increases, this tri-modality vanishes.[1]

Furthermore, unless the magnitude of the break is large, the asymptotic distribution is symmetric only if both: (i) the break date is located at mid-sample, (ii) the distribution of the errors and regressors do not differ "too much" across regimes. Given the fixed-span setting, our limit theory treats the volatility of the regressors and errors as random quantities. We thus use the concept of stable convergence in distribution. As for the impact of the sample size relative to the span of the data on the precision of the estimate, we find that the span plays a more pronounced role. We also show, via simulations, that our continuous record asymptotics provides good approximations

[1]In work that we became aware of after the first draft of this paper, Jiang, Wang, and Yu (forthcoming) studied the finite-sample bias of a break point estimator based on maximum likelihood for a simple univariate diffusion with constant volatility and a change-point in the drift. They also find that the span can be important and the distribution can be asymmetric. We comment on the differences in the Appendix.

to the finite-sample distributions of the estimate of the break date.

Our continuous record asymptotic theory is not limited to providing a better approximation to the finite-sample distribution. It can also be exploited to address the problem of conducting inference about the break date. This issue has received considerable attention. Besides the original asymptotic arguments used by Bai (1997) and Bai and Perron (1998), Elliott and Müller (2007) proposed to invert Nyblom's (1990) statistic, while Eo and Morley (2015) introduced a procedure based on the likelihood-ratio statistic of Qu and Perron (2007). The latter methods were mainly motivated by the fact that the empirical coverage rates of the confidence intervals obtained from Bai's (1997) method are below the nominal level with small breaks. The method of Elliott and Müller (2007) delivers the most accurate coverage rates, though at the expense of increased average lengths of the confidence sets especially with large breaks [cf. Chang and Perron (forthcoming)]. What is still missing is a method that, uniformly over break magnitudes, achieves both accurate coverage rates and satisfactory average lengths of the confidence sets for a wide range of data-generating processes. Given the peculiar properties of the continuous record asymptotic distribution, we propose an inference method which is rather non-standard and relates to Bayesian analyses. We use the concept of Highest Density Region to construct confidence sets for the break date. Our method is simple to implement and has a frequentist interpretation.

The simulation analysis conducted indicates that our approach has two notable properties. First, it provides adequate empirical coverage rates over all data-generating mechanisms considered and, importantly, for any size and/or location of the break, a notoriously difficult problem. Second, the lengths of the confidence sets are always shorter than those obtained using Elliott and Müller's (2007) approach. Often, the reduction in length is substantial and increases with the size of the break. Also, our method performs markedly better when lagged dependent variables are present in the model. Compared to Bai's (1997) method, our approach yields better coverage rates, especially when the magnitude of the break is small. With large breaks, the two methods are basically equivalent. Of particular interest is the fact that our confidence set can be the union of disjoint intervals. This is illustrated in Section 6.

The paper is organized as follows. Section 2 introduces the model, the estimation method and extensions to predictable processes. 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. Simulation results about its adequacy are reported in Section 7. Section 8 provides brief concluding remarks. Additional details and some proofs for the main results are included in an appendix. The Supplement contains most of the proofs as well as additional materials.

2 Model and Assumptions

Section 2.1 introduces the benchmark model of interest, the main assumptions, the estimation method and the relation of our setup with the traditional large-N asymptotic framework. In Section 2.2 we extend the benchmark model to include predictable processes. The following notations are used throughout. Recall the relation $N = Th$. We shall use $T \to \infty$ and $h \downarrow 0$ interchangeably. All vectors are column vectors. For two vectors a and b, we write $a \leq b$ if the inequality holds component-wise. We denote the transpose of a matrix A by A' and the (i, j) elements of A by $A^{(i,j)}$. For a sequence of matrices $\{A_T\}$, we write $A_T = o_p(1)$ if each of its elements is $o_p(1)$ and likewise for $O_p(1)$. $\mathbb{R}$ denotes the set of real numbers. 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 and for a set A, the indicator function of A is denoted by $\mathbf{1}_A$. The symbol $\otimes$ denotes the product of σ-fields. 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 $\overset{\mathcal{L}-\mathrm{s}}{\Rightarrow}$ 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. Finally, note that in general N is not identified and could be normalized to one. However, we keep a generic N throughout to allow a better intuitive understanding of the results.

2.1 The Benchmark Model

We consider the following partial structural change model with a single break point:

$$
\begin{aligned}
Y_t &= D_t'\pi^0 + Z_t'\delta_1^0 + e_t, &&\left(t = 0, 1, \ldots, T_b^0\right) \\
Y_t &= D_t'\pi^0 + Z_t'\delta_2^0 + e_t, &&\left(t = T_b^0 + 1, \ldots, T\right),
\end{aligned}
\tag{2.1}
$$

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 π^0, δ_1^0 and δ_2^0 are unknown with $\delta_1^0 \neq \delta_2^0$. Our main purpose is to develop inference methods for the unknown break 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 processes $\{D_s, Z_s, e_s\}_{s\geq 0}$ are continuous-time processes, defined on a filtered probability space $\left(\Omega, \mathscr{F}, (\mathscr{F}_s)_{s\geq 0}, P\right)$, where s can be interpreted as the continuous-time index. We observe realizations of $\{Y_s, D_s, Z_s\}$ at discrete points of time. Below, we impose very minimal "pathwise" assumptions on these continuous-time stochastic processes which imply mild restrictions on the observed discrete-time counterparts. We discuss what these assumptions imply for our model and the distributional properties of the errors and regressors.

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 $\{{}_hY_{kh},\ {}_hD_{kh},\ {}_hZ_{kh};\ k=0,\ 1,\ldots,\ T=N/h\}$. ${}_hD_{kh}\in\mathbb{R}^q$ 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 Section 2.2. Recall the Doob-Meyer decomposition [cf. Doob (1953) and Meyer (1967)][2] from which it follows that any locally-integrable semimartingle process can be decomposed into a "predictable" and a "martingale" part. The discretized processes ${}_hD_{kh}$ and ${}_hZ_{kh}$ are assumed to be adapted to the increasing and right-continuous filtration $\{\mathscr{F}_t\}_{t\geq 0}$. For any process X we denote its "increments" by $\Delta_hX_k=X_{kh}-X_{(k-1)h}$. For $k=1,\ldots,T$, let $\Delta_hD_k\triangleq\mu_{D,k}h+\Delta_hM_{D,k}$ and $\Delta_hZ_k\triangleq\mu_{Z,k}h+\Delta_hM_{Z,k}$ where the "drifts" $\mu_{D,t}\in\mathbb{R}^q$, $\mu_{Z,t}\in\mathbb{R}^p$ are $\mathscr{F}_{t-h}$-measurable (exact assumptions will be given below), 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., $\mathbb{E}\left(\Delta_hM_{D,t}\Delta_hM'_{D,t}|\,\mathscr{F}_{t-h}\right)=\Sigma_{D,t-h}\Delta t$ and $\mathbb{E}\left(\Delta_hM_{Z,t}\Delta_hM'_{Z,t}|\,\mathscr{F}_{t-h}\right)=\Sigma_{Z,t-h}\Delta t$ (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, model (2.1) can be expressed as

$$\Delta_hY_k\triangleq\begin{cases}(\Delta_hD_k)'\,\pi^0+(\Delta_hZ_k)'\,\delta^0_{Z,1}+\Delta_he^*_k, & (k=1,\ldots,\lfloor T\lambda_0\rfloor)\\ (\Delta_hD_k)'\,\pi^0+(\Delta_hZ_k)'\,\delta^0_{Z,2}+\Delta_he^*_k, & (k=\lfloor T\lambda_0\rfloor+1,\ldots,\ T)\end{cases}, \tag{2.2}$$

where the error process $\{\Delta_he^*_t,\ \mathscr{F}_t\}$ is a continuous local martingale difference sequence with conditional variance $\mathbb{E}\left[(\Delta_he^*_t)^2\,|\,\mathscr{F}_{t-h}\right]=\sigma^2_{e,t-h}\Delta t$ 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

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

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.

[2]A treatment of the probabilistic material can be found in Aït-Sahalia and Jacod (2014), Karatzas and Shreve (1996), Protter (2005), Jacod and Shiryaev (2003) and Jacod and Protter (2012). For measure theoretical aspects we refer to Billingsley (1995).

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 $\int_0^t\Sigma_{D,s}^{(j,j)}ds<\infty$ $(j=1,\ldots,q)$ and $\int_0^t\Sigma_{Z,s}^{(j,j)}ds<\infty$ $(j=1,\ldots,p)$. Furthermore, for every $j=1,\ldots,q$, $r=1,\ldots,p$, and $k=1,\ldots,T$, $h^{-1}\int_{(k-1)h}^{kh}\Sigma_{D,s}^{(j,j)}ds$ and $h^{-1}\int_{(k-1)h}^{kh}\Sigma_{Z,s}^{(r,r)}ds$ are bounded away from zero and infinity, uniformly in k and h; (iv) e_t^* is such that $e_t^*\triangleq\int_0^t\sigma_{e,s}dW_{e,s}$ 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 $t\geq 0$.*

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), Li, Todorov, and Tauchen (2017) and Li and Xiu (2016)]. Part (iii) imposes restrictions on the regressors which require them to have finite integrated covariance. The second part of condition (iii) means that the process $\Sigma_{\cdot,t}^{(j,j)}$ is bounded away from zero and infinity on any bounded time interval. Part (iv) specifies the error term to be contemporaneously uncorrelated with the regressors. We also rule out jump processes. This is a natural restriction to impose since it essentially implies that the structural change in our model arises from the shift in the parameter $\delta_{Z,1}^0$ after T_b^0 only. Hence, our results are not expected to provide good approximations for applications involving high-frequency data for which jumps are likely to be important. Our intended scope is for models involving data sampled at, say, the daily or lower frequencies. Since this is an important point we restate it as a separate assumption:

Assumption 2.2. *D, Z, e and $\Sigma^0\triangleq\{\Sigma_{\cdot,t},\,\sigma_{e,t}\}_{t\geq0}$ have P-a.s. continuous sample paths.*

The assumption above implies that the variables in our model are diffusion processes if one further assumes that the volatilities are deterministic. We shall not impose the latter condition. As a consequence, the processes D_t and Z_t belong to the class of continuous Itô semimartingales with stochastic volatility. Our choice of modeling volatility as a latent factor is justified on multiple grounds. First, a setting in which the variance process is stochastic seems to be more appropriate for the development of a fixed-span asymptotic experiment since sampling uncertainty cannot be averaged out with a limited span of data. Second, some estimates will follow a mixed Gaussian distribution asymptotically, which may lead to better approximations. Third, it does not impose any substantial impediment for the development of our theoretical results. Fourth, such results will be valid under general conditions on the variance processes, e.g., nonstationarity and long-memory.

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. We shall extend model (2.2) to allow for predictable processes (e.g., a constant and/or lagged dependent variable) in a separate section.

Assumption 2.3. $N_b^0 = \lfloor N\lambda_0 \rfloor$ *for some* $\lambda_0 \in (0,\, 1)$.

Assumption 2.3 dictates the asymptotic framework adopted and implies that the change-point occurs at the observation-index $T_b^0 = \lfloor T\lambda_0 \rfloor$, where $T_b^0 = \lfloor N_b^0/h \rfloor$. Our framework requires us to distinguish between the actual break date N_b^0 and the index of the observation associated with the break point, T_b^0. From a practical perspective, the assumption states that the change-point is bounded away from the starting and end points. It implies that the pre- and post-break segments of the sample remain fixed whereas the usual assumption under the large-N asymptotics implies that the time horizons before and after the break date grow proportionately. This, along with the usual mixing assumptions imply that only a small neighborhood around the true break date is relevant asymptotically, thereby ruling out the possibility for the long-span asymptotics to discern features simply caused by the location of the break. As opposed to the large-N asymptotics, the continuous record asymptotic framework preserves information about the data span and the location of the break. This feature is empirically relevant; simulations reported in Elliott and Müller (2007) suggests that the location of the break affects the properties of its estimate in small samples. We show below that our theory reproduces these small-sample features and provide accurate approximations to the finite-sample distributions.

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

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

where the true parameter $\theta^0 = \left((\beta^0)'\,,\, (\delta^0)'\right)'$ 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.

The final step is to write the model in matrix format which will be useful for the derivations. Let $Y = (y_h, \ldots, y_{Th})'$, $X = (x_h, \ldots, x_{Th})'$, $e = (e_h, \ldots, e_{Th})'$, $X_1 = (x_h, \ldots, x_{T_b h}, 0, \ldots, 0)'$, $X_2 = \left(0, \ldots, 0, x_{(T_b+1)h}, \ldots, x_{Th}\right)'$ and $X_0 = \left(0, \ldots, 0, x_{\left(T_b^0+1\right)h}, \ldots, x_{Th}\right)'$. 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: $Y = X\beta^0 + Z_0\delta^0 + e$. 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^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 or h. The estimate of the break fraction is then $\widehat{\lambda}_b = \widehat{T}_b/T$.

2.2 The Extended Model with Predictable Processes

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. We consider the following extended model:

$$\Delta_h Y_k \triangleq \begin{cases} \mu_{1,h}h + \alpha_{1,h}Y_{(k-1)h} + (\Delta_h D_k)'\pi^0 + (\Delta_h Z_k)'\delta_{Z,1}^0 + \Delta_h e_k^*, & (k = 1, \ldots, \lfloor T\lambda_0 \rfloor) \\ \mu_{2,h}h + \alpha_{2,h}Y_{(k-1)h} + (\Delta_h D_k)'\pi^0 + (\Delta_h Z_k)'\delta_{Z,2}^0 + \Delta_h e_k^*, & (k = \lfloor T\lambda_0 \rfloor + 1, \ldots, T) \end{cases} \tag{2.5}$$

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., $\mu_{1,h} \triangleq \mu_1^0 h^{-1/2}$, $\mu_{2,h} \triangleq \mu_2^0 h^{-1/2}$, $\mu_{\delta,h} \triangleq \mu_{2,h} - \mu_{1,h}$, $\alpha_{1,h} \triangleq \alpha_1^0 h^{-1/2}$, $\alpha_{2,h} \triangleq \alpha_2^0 h^{-1/2}$ 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 (1981), 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 $\mu_\cdot^0$ and $\alpha_\cdot^0$, the continuous-time model is then equivalent to

$$\begin{aligned} Y_t = Y_0 &+ \int_0^t \left(\mu_1^0 + \mu_\delta^0 \mathbf{1}_{\left\{s > N_b^0\right\}}\right) ds + \int_0^t \left(\alpha_1^0 + \alpha_\delta^0 \mathbf{1}_{\left\{s > N_b^0\right\}}\right) Y_s ds \\ &+ D_t'\pi^0 + \int_0^t \left(\delta_{Z,1}^0 + \delta^0 \mathbf{1}_{\left\{s > N_b^0\right\}}\right)' dZ_s + e_t^*, \end{aligned} \tag{2.6}$$

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, at the expense of additional technical details of no substance.

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), Bai and Perron (1998) and also Perron and Qu (2006) who relaxed the conditions used. 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. In order to procced, we impose the following identification conditions.

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.*

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 $(T_b^0,\, \theta^0)$.*

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 2.1-2.3 and 3.1-3.2, for any $\varepsilon > 0$ and $K > 0$, and all large T, $P\left(\left|\widehat{\lambda}_b - \lambda_0\right| > K\right) < \varepsilon$.*

Proposition 3.2. *Under Assumption 2.1-2.3 and 3.1-3.2 for any $\varepsilon > 0$, there exists a $K > 0$ such that for all large T, $P\left(T\left|\widehat{\lambda}_b - \lambda_0\right| > K\right) = P\left(\left|\widehat{T}_b - T_b^0\right| > K\right) < \varepsilon$.*

We have the same T-convergence rate as under large-N asymptotics. Let $\theta^0 = \left((\beta^0)',\, (\delta_1^0)',\, (\delta_2^0)'\right)'$. 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 $\mathrm{argmin}_{\beta \in \mathbb{R}^{p+q}, \delta \in \mathbb{R}^p} \left\| Y - X\beta - \widehat{Z}_2 \delta \right\|^2$, where we use $T_b = \widehat{T}_b$ in the construction of $\widehat{Z}_2$. Then we have the following result, akin to an extension of corresponding results in Section 3 of Barndorff-Nielsen and Shephard (2004). As a matter of notation, let $\Sigma^* \triangleq \{\mu_{\cdot,t},\, \Sigma_{\cdot,t},\, \sigma_{e,t}\}_{t \geq 0}$ and denote expectation taken with respect to Σ^* by $\mathbb{E}^*$.

Proposition 3.3. *Under Assumption 2.1-2.3 and 3.1-3.2, we have as $T \to \infty$ (N fixed), conditionally on Σ^*,* $\left(\sqrt{T/N}\left(\widehat{\beta}-\beta^0\right),\, \sqrt{T/N}\left(\widehat{\delta}-\delta^0\right)\right)' \xrightarrow{d} \mathscr{M}\mathscr{N}\left(0,\, V\right)$ *where $\mathscr{M}\mathscr{N}$ denotes a mixed Gaussian distribution, with*

$$V \triangleq \overline{V}^{-1} \lim_{T\to\infty} T \begin{bmatrix} \sum_{k=1}^{T} \mathbb{E}^*\left(x_{kh}x_{kh}' e_{kh}^2\right) & \sum_{k=T_b^0}^{T} \mathbb{E}^*\left(x_{kh}z_{kh}' e_{kh}^2\right) \\ \sum_{k=T_b^0}^{T} \mathbb{E}^*\left(x_{kh}z_{kh}' e_{kh}^2\right) & \sum_{k=T_b^0}^{T} \mathbb{E}^*\left(z_{kh}z_{kh}' e_{kh}^2\right) \end{bmatrix} \overline{V}^{-1},$$

and

$$\overline{V} \triangleq \lim_{T\to\infty} \begin{bmatrix} \sum_{k=1}^{T} \mathbb{E}^*\left(x_{kh}x_{kh}'\right) & \sum_{k=T_b^0}^{T} \mathbb{E}^*\left(x_{kh}z_{kh}'\right) \\ \sum_{k=T_b^0}^{T} \mathbb{E}^*\left(x_{kh}x_{kh}'\right) & \sum_{k=T_b^0}^{T} \mathbb{E}^*\left(z_{kh}z_{kh}'\right) \end{bmatrix}.$$

The limit law of the regression parameters is mixed Gaussian, where the variance matrix V is stochastic. Hence, the theorem is also useful because it approximates a setting where the uncertainty about the break date transmits to a limit law for the regression parameters that has heavier tails than the Gaussian law; this turns out to be often the case in practice. Under the assumption of deterministic variances, the limit law would be a normal variate.

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. Section 4.1 presents the main theoretical results. The features of the asymptotic distribution obtained are discussed in Section 4.2.

4.1 Main Theoretical Results

We first discuss the main arguments of our derivation. 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, $\widehat{N}_b - N_b^0 = O_p\left(T^{-1}\right)$, i.e., $\widehat{N}_b$ is in a shrinking neighborhood of N_b^0, which, however, shrinks too fast and impedes the development of a feasible limit theory. Hence, we rescale time and work with the objective function in a small neighborhood of the true break date under this "new" time scale. We begin with the following assumption which specifies that i) we use a shrinking condition on δ^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 necessary in our context in order to accurately approximate the change-point problem. In addition, it also leads to a limit distribution that is influenced by parameters that can be consistently estimated, so that a feasible method of inference can be conducted.

Assumption 4.1. *Let $\delta_h = \delta^0 h^{1/4}$ and assume that for all $t \in \left(N_b^0 - \epsilon,\, N_b^0 + \epsilon\right)$, with $\epsilon \downarrow 0$ and $T^{1-\kappa}\epsilon \to B < \infty$, $0 < \kappa < 1/2$, $\mathbb{E}\left[\left(\Delta_h e_t^*\right)^2 \mid \mathscr{F}_{t-h}\right] = \sigma^2_{h,t-h}\Delta t$ P-a.s., where $\sigma_{h,t} \triangleq \sigma_h \sigma_{e,t}$, $\sigma_h \triangleq \overline{\sigma} h^{-1/4}$ and $\overline{\sigma} \triangleq \int_0^N \sigma^2_{e,s} ds$.*

The vector of scaled true parameters is $\theta_h \triangleq \left(\left(\beta^0\right)',\, \delta_h'\right)'$. Define

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

We shall refer to $\{\Delta_h \widetilde{e}_t,\, \mathscr{F}_t\}$ as the normalized residual process. Under this framework, the rate of convergence 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 a 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.

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

Consider the set $\mathcal{D}(C) \triangleq \left\{N_b :\, N_b \in \left\{N_b^0 + Ch^{1-\kappa}\right\},\, |C| < \infty\right\}$, on the original time scale. Let $Z_\Delta \triangleq \left(0, \ldots, 0,\, z_{(T_b+1)h}, \ldots, z_{T_b^0 h},\, 0, \ldots, 0\right)$ if $T_b < T_b^0$ and $Z_\Delta \triangleq \left(0, \ldots, 0,\, z_{\left(T_b^0+1\right)h}, \ldots, z_{T_b h},\, 0, \ldots, 0\right)$ if $T_b > T_b^0$; also set $\psi_h \triangleq h^{1-k}$. The following lemma will be needed in the derivations.

Lemma 4.1. *Under Assumption 2.1-2.3, 3.1-3.2 and 4.1, uniformly in T_b,*

$$\left(Q_T\left(T_b\right) - Q_T\left(T_b^0\right)\right)/\psi_h = -\delta_h\left(Z_\Delta' Z_\Delta/\psi_h\right)\delta_h + 2\delta_h'\left(Z_\Delta' e/\psi_h\right)\operatorname{sgn}\left(T_b^0 - T_b\right) + o_p\left(h^{1/2}\right). \tag{4.2}$$

Lemma 4.1 shows that only the terms involving the regressors whose parameters are allowed to shift have a first-order effect on the asymptotic analysis. For brevity, we use the notation $\pm$ in place of $\text{sgn}\left(T_b^0 - T_b\right)$ hereafter.

The conditional first moment of the centered criterion function $Q_T\left(T_b\right) - Q_T\left(T_b^0\right)$ is of order $O\left(h^{1-\kappa}\right)$, i.e., "oscillates" rapidly as $h \downarrow 0$. Hence, in order to approximate the behavior of $\left\{\widehat{T}_b - T_b^0\right\}$ we 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 $\mathcal{D}\left(C\right)$, respectively. We then have $\left|R_C - L_C\right| = O\left(Ch^{1-\kappa}\right)$. Now, take the vanishingly small interval $\left[L_C,\, R_C\right]$ on the original time scale, and stretch it into a time interval $\left[T^{1-\kappa}L_C,\, T^{1-\kappa}R_C\right]$ on a new "fast time scale". Since the criterion function is scaled by ψ_h^{-1}, all scaled processes are $O_p\left(1\right)$. Now, let $N_b\left(v\right) = N_b^0 - vh^{1-\kappa}$, $v \in \left[-C,\, C\right]$. Using Lemma 4.1 and Assumption 4.1 (see the appendix),

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

In addition, in view of (2.3), we let $dZ_{\psi,s} = \psi_h^{-1/2}\sigma_{Z,s}dW_{Z,s}$ for $s \in \left[N_b^0 - vh^{1-\kappa},\, N_b^0 + vh^{1-\kappa}\right]$. 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 $t \in \mathcal{D}^*\left(C\right)$, where $\mathcal{D}^*\left(C\right) \triangleq \left\{t:\, t \in \left[N_b^0 + v\left\|\delta^0\right\|^2/\overline{\sigma}^2\right],\, |v| \leq C\right\}$. Therefore,

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

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

Our analysis is local; we examine the limiting behavior of the centered and rescaled criterion function process in a neighborhood $\mathcal{D}^*(C)$ 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 $(Q_T(T_b(v)) - Q_T(T_b^0))/h^{1/2}$ 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

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

where $\theta^* \triangleq (\theta_h', v)'$ with $T_b(v) \triangleq T_b^0 + \lfloor v/h \rfloor$ and $T_b(v)$ is the time index on the "fast time scale". When v varies, $T_b(v)$ visits all integers between 1 and T, with the normalizations $T_b(v) = 1$ if $T_b(v) \leq 1$ and $T_b(v) = T$ if $T_b(v) \geq T$. On the old time scale $N_b(u) = N_b^0 + u$ with $v \to \psi_h^{-1}u$, so that $N_b(u)$ is in a vanishing neighborhood of N_b^0. On $\mathcal{D}^*(C)$, we index the process $Q_T(\theta_h, T_b(v)) - Q_T(\theta^0, T_b^0)$ 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; the optimization problem is not affected by the change of time scale. In fact, by Proposition 4.1, $u = Th\left(\widehat{\lambda}_b - \lambda_0\right) = KO_p(h^{1-\kappa})$ on the old time scale; whereas on the new "fast time scale", $v = Th\left(\widehat{\lambda}_b - \lambda_0\right) = O_p(1)$. The maximization problem is not changed because v/h can take any value in $\mathbb{R}$. The process $Q_T(\theta_h, T_b(v)) - Q_T(\theta^0, T_b^0)$ is thus analyzed on a fixed horizon since v now varies over $\left[-N_b^0/\left(\|\delta^0\|^{-2}\overline{\sigma}^2\right), (N - N_b^0)/\left(\|\delta^0\|^{-2}\overline{\sigma}^2\right)\right]$. Hence, redefine

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

Note that $\mathcal{D}^*(C)$ is compact. Let $\mathbb{D}(\mathcal{D}^*(C), \mathbb{R})$ denote the space of all *càdlàg* functions from $\mathcal{D}^*(C)$ into $\mathbb{R}$. Endow this space with the Skorokhod topology and note that $\mathbb{D}(\mathcal{D}^*(C), \mathbb{R})$ is a Polish space. The faster rate of convergence of $\widehat{\lambda}_b$ established in Proposition 4.1-(ii) combined with the $\sqrt{T}$-rate for the regression parameters allow us to apply the continuous mapping theorem for the argmax functional [cf. Kim and Pollard (1990)]. 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 $\overline{Q}_T(\theta^*)$. These asymptotic results mainly rely 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 $\overline{Q}_T(\theta^*)$ is, according to Lemma 4.1, the same as the limit law of

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

where $\overset{d}{\equiv}$ denotes (first order) equivalence in law, $\widetilde{e}_{kh} \triangleq h^{1/4}e_{kh}$ and since (approximately) $e_{kh} \sim i.n.d.\,\mathscr{N}\left(0, \sigma_{h,k-1}^2 h\right)$, $\sigma_{h,k} = \sigma_h\sigma_{e,k}$ then $\widetilde{e}_{kh} \sim i.n.d.\,\mathscr{N}\left(0, \sigma_{e,k-1}^2 h\right)$. Hence, the limit law of

$\overline{Q}_T(\theta^*)$ is, to first-order, equivalent to the law of

$$-\left(\delta^0\right)'\left(Z_\Delta' Z_\Delta\right)\delta^0 \pm 2\left(\delta^0\right)'\left(h^{-1/2} Z_\Delta' \widetilde{e}\right). \tag{4.3}$$

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 $\overline{Q}_T(\theta^*)$ account for the discrepancy between the deterministic and stochastic component in (4.3).

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 full details are relegated to the Appendix. Part of the proof involves showing the stable convergence in distribution [cf. Rényi (1963) and Aldous and Eagleson (1978)] toward an $\mathscr{F}$-conditionally two-sided Gaussian process. The limiting process is realized on a extension of the original probability space and we relegate this description to Section A.1 in the Appendix.

Theorem 4.1. *Under Assumption 2.1-2.3, 3.1-3.2 and 4.1, and under the "fast time scale",*

$$N\left(\widehat{\lambda}_b-\lambda_0\right) \overset{\mathcal{L}-\mathrm{s}}{\Rightarrow} \underset{v\in\left[-N_b^0/\left(\|\delta^0\|^{-2}\overline{\sigma}^2\right),\,\left(N-N_b^0\right)/\left(\|\delta^0\|^{-2}\overline{\sigma}^2\right)\right]}{\operatorname{argmax}} \left\{-\left(\delta^0\right)'\left\langle Z_\Delta,\, Z_\Delta\right\rangle(v)\,\delta^0+2\left(\delta^0\right)'\mathscr{W}(v)\right\}, \tag{4.4}$$

where $\langle Z_\Delta,\, Z_\Delta\rangle(v)$ is the predictable quadratic variation process of Z_Δ. The process $\mathscr{W}(v)$ is, conditionally on the σ-field $\mathscr{F}$, a two-sided centered Gaussian martingale with independent increments and variances given in Section A.1 in the Appendix.

Note that the theorem is in accordance with Proposition 4.1 because it holds under the new "fast time scale". The theoretical results from Section 3 and Theorem 4.1 allow one to draw the following features. Under fixed shifts, the fractional break date is super-consistent—i.e., $N\left(\widehat{\lambda}_b-\lambda_0\right)=O_p\left(h/\left\|\delta^0\right\|^2\right)$—whereas the estimate of the break point $\widehat{T}_b$ is not even consistent. Simply letting the magnitude of the shift shrink to zero does not result in a useful approximation when the shifts are small. When one augments the shrinking shifts assumption with locally increasing variances (cf. Assumption 4.1), the rate of convergence of $\widehat{\lambda}_b$ becomes slower. Further, through a change of time scale, Theorem 4.1 suggests that the span of the data is more important than the sample size when shifts are small. Further work will report on formalizing the complex relationships between the sampling frequency, sample size, span of the data and shift magnitude. For example, we can show that $\widehat{T}_b$ is itself consistent if the shift is large, i.e., $\delta\to\infty$, even if the sample size and the sampling frequency are fixed.

For comparison purposes, recall that the classical large-N limiting distribution is related to the location of the maximum of a two-sided Wiener process over the interval $(-\infty,\,\infty)$. Its probability density is symmetric for the case of stationary regimes and has thicker tails and a higher peak than the density of a Gaussian variate. In contrast, the limiting 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 $\left[-N_b^0/\left(\|\delta^0\|^{-2}\,\overline{\sigma}^2\right),\,(N-N_b^0)/\left(\|\delta^0\|^{-2}\,\overline{\sigma}^2\right)\right]$. Notably, this domain depends on the true value of the break point 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.

Additional relevant remarks follow; more details are provided in Section 4.2. The size of the shift plays a key role in determining the density of the asymptotic distribution. More precisely, for an appropriately defined "signal-to-noise" ratio, the density displays interesting properties which change when this quantity 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. This complies with simulation evidence pointing out that, for instance, 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 and this is clearly an 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 time horizon $[0,\,N]$ is fixed and thus we can distinguish between the statistical content of the segments $[0,\,N_b^0]$ and $[N_b^0,\,N]$. In contrast, this is not feasible under the classical shrinkage large-N asymptotics because both the pre- and post-break segments increase to infinity proportionately and mixing conditions are imposed so that the only relevant information is a neighborhood around the true break date. As for the relative impact of time span N versus sample size T on the precision of the estimator, the time span plays a key role and has a more pronounced impact relative to the sample size. We shall see in the next section that the asymptotic distribution derived under a continuous record provides an accurate approximation to the finite-sample distribution and the approximation is remarkably better than that resulting from the classical shrinkage large-N asymptotics [cf. Bai (1997) and Yao (1987)]. Details on how to simulate the limiting distribution in Theorem 4.1 are given in Section A.4.

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

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, 3.1-3.2 and 4.1-4.2, and under the "fast time scale",*

$$\frac{\left((\delta^0)'\langle Z,\, Z\rangle_1\delta^0\right)^2}{(\delta^0)'\Omega_{\mathscr{W},1}\delta^0} N\left(\widehat{\lambda}_b - \lambda_0\right) \Rightarrow \underset{s\in\left[-\frac{N_b^0}{\|\delta^0\|^{-2}\overline{\sigma}^2}\frac{\left((\delta^0)'\langle Z,\,Z\rangle_1\delta^0\right)^2}{(\delta^0)'\Omega_{\mathscr{W},1}(\delta^0)},\, \frac{N-N_b^0}{\|\delta^0\|^{-2}\overline{\sigma}^2}\frac{\left((\delta^0)'\langle Z,\,Z\rangle_1\delta^0\right)^2}{(\delta^0)'\Omega_{\mathscr{W},1}\delta^0}\right]}{\operatorname{argmax}} \mathscr{V}(s). \tag{4.5}$$

Unlike the asymptotic distribution derived under classical long-span asymptotics, the probability density function of the argmax process in 4.5 is not available in closed form. Furthermore, the limiting distribution depends on unknown quantities. We first discuss the probabilistic properties of the infeasible density and then explain how we 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.

4.2 Infeasible Density of the Asymptotic Distribution

An important parameter is $\rho \triangleq \left((\delta^0)'\langle Z,\, Z\rangle_1\delta^0\right)^2/\left((\delta^0)'\Omega_{\mathscr{W},1}\delta^0\right)$. We plot the probability density functions of the infeasible distribution of $\rho N\left(\widehat{\lambda}_b - \lambda_0\right)$ for $N = 100$, as given in Theorem 4.2, and compare it with the corresponding distribution in Bai (1997). We first consider cases for which the first and second moments of regressors and errors do not vary "too much" across regimes; cases that satisfy:

$$\frac{1}{\varpi} \leq \frac{\rho}{\xi_1},\, \frac{\rho}{\xi_2} \leq \varpi, \qquad \xi_1 = \frac{(\delta^0)'\langle Z,\, Z\rangle_2\delta^0}{(\delta^0)'\langle Z,\, Z\rangle_1\delta^0}, \qquad \xi_2 = \left(\frac{(\delta^0)'\Omega_{\mathscr{W},2}\delta^0}{(\delta^0)'\Omega_{\mathscr{W},1}\delta^0}\right), \tag{4.6}$$

for some number ϖ (see below). Then $\mathscr{V}(s)$ in (4.5) becomes

$$\mathscr{V}(s) = \begin{cases} -\frac{|s|}{2} + W_1^*(-s), & \text{if } s < 0 \\ -\frac{|s|}{2}\xi_1 + \sqrt{\xi_2}W_2^*(s), & \text{if } s \geq 0. \end{cases}$$

We consider the case of "nearly stationary regimes" where we set $\varpi = 1.5$ so that the degree of heterogeneity across regimes is restricted. Figure 1 displays the density of $\rho\left(\widehat{T}_b - T_0\right)$ for $\lambda_0 = 0.3,\ 0.5,\ 0.7$ and for a low signal-to-noise ratio $\rho^2 = 0.2$. In addition, we set $\xi_1 = \xi_2 = 1$ so that each regime has the same distribution. We also plot the density of Bai's (1997) large-N shrinkage asymptotic distribution [see also Yao (1987)]. The corresponding plots when $\rho^2 = 0.3,\ 0.5,\ 0.8$ are reported in Figure 2-4. Note that the restrictions in (4.6) imply that a high value of ρ corresponds to a large shift size δ^0.

Several interesting observations appear at the outset. First, 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. Second, it has thicker tails and a much higher peak than the density of a standard normal variate. None of these features are shared by the density derived under a continuous record. When the true break date 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 ($\rho^2 = 0.2,\ 0.3$), the density features three modes. The highest mode is not at the true break date $\lambda_0 = 0.5$ when the signal is very low ($\rho^2 = 0.2$). When the break date is not at mid-sample, the density is asymmetric despite having homogenous regimes. This tri-modality vanishes as the signal-to-noise ratio increases ($\rho^2 = 0.8$) (Figure 4, middle panel). When $\rho^2 = 0.3$ and the true break date is not at mid sample ($\lambda_0 = 0.3$ and $\lambda_0 = 0.7$; left and right panel, respectively, Figure 2) the density is asymmetric; for values of λ_0 less (larger) than 0.5, the probability density is right (left) skewed. Such feature is more apparent when the signal-to-noise ratio is low ($\rho^2 = 0.2,\ 0.3$ and 0.5; Figure 1-3, side panels). When the signal is low and λ_0 is less (larger) than 0.5, the probability density has highest mode at values that correspond to $\widehat{\lambda}_b$ 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 ($\rho^2 = 0.5,\ 0.8\,,\ 1.5$) the highest mode is centered at a value which corresponds to $\widehat{\lambda}_b$ being close to λ_0 (cf. Figure 3-4, side panels). Indeed, the density is still asymmetric when $\lambda_0 = 0.3$ and $\lambda_0 = 0.7$ if $\rho^2 = 0.5$.[3]

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 ρ 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.

The supplementary document contains an extended description of the features of the asymp-

[3] 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.

totic distribution where we consider many other cases for λ_0 and ρ^2 as well as an analysis of cases allowing differences between the distribution of errors and regressors in the pre- and post-break regimes (referred to as "non-stationary regimes"). In the latter case, we show that even if the signal-to-noise ratio is moderately high the continuous record asymptotic distribution is asymmetric even when the break occurs at mid-sample. This is in stark contrasts to the "nearly stationary" scenario since the density was shown to be always symmetric no matter the value taken by ρ if $\lambda_0 = 0.5$. This means that the asymptotic distribution attributes different weights to the informational content of the two regimes since they possess heterogeneous characteristics.

The results for the densities under the two different scenarios, "nearly stationary" versus "non-stationary regimes", allows one to deduce the following feature. The asymptotic distribution of $\widehat{T}_b$ is symmetric if both (i) the break date is located at exactly mid-sample, and (ii) the distributions of the errors and regressors do not differ "too much" across regimes. This holds unless the break magnitude is very large in which case the density is symmetric.

5 Approximation to the Finite-Sample 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 break date with a feasible version of the continuous record asymptotic distribution obtained with plug-in estimates. We obtain the finite-sample distribution of $\rho\left(\widehat{T}_b - T_b^0\right)$ based on 100,000 simulations from the following model:

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

where $Z_t = 0.5Z_{t-1} + u_t$ with $u_t \sim i.i.d.\, \mathscr{N}(0,\, 1)$ independent of $e_t \sim i.i.d.\, \mathscr{N}(0,\, \sigma_e^2)$, $\sigma_e^2 = 1$, $\pi^0 = 1$, $Z_0 = 0$, $D_t = 1$ for all t, and $T = 100$. We set $T_b^0 = \lfloor T\lambda_0 \rfloor$ with $\lambda_0 = 0.3, 0.5, 0.7$ and consider different break sizes $\delta^0 = 0.2, 0.3, 0.5, 1$. 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, $\left\|\delta^0\right\|^{-2} \overline{\sigma}^2$, ξ_1, ξ_2 and ρ at their true values. The feasible counterparts are constructed with plug-in estimates of ξ_1, ξ_2, ρ and $\left(N_b^0 \left\|\delta^0\right\|^2 / \overline{\sigma}^2\right)\rho$. 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. The estimates $\widehat{\xi}_1$ and $\widehat{\xi}_2$ 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. Use is made of the fact that the quadratic variation $\langle Z, Z \rangle_1$ 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 $T \sum_{k=1}^{\widehat{T}_b} \widehat{e}_{kh}^2 z_{kh} z'_{kh} / \left(\widehat{\lambda}_b \right)$. The argument for $\lambda_0 \left\| \delta^0 \right\|^2 \overline{\sigma}^{-2} \rho$ 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 $\widehat{\xi}_i \xrightarrow{p} \widehat{\xi}_i$ $(i = 1, 2)$, the corresponding approximations for $\widehat{\vartheta}$ and $\widehat{\rho}$ are given by $\widehat{\vartheta}/h \xrightarrow{p} \vartheta$ and $\widehat{\rho}/h \xrightarrow{p} \rho$. To derive the latter two results we used that on the "fast time scale", Assumption 4.1 implies that the errors have higher volatilities and thus the squared residual $\widehat{e}_{kh}^2$ needs to be multiplied by the factor $h^{1/2}$. Then, $h^{1/2} \sum_{k=1}^{T} \widehat{e}_{kh}^2 \xrightarrow{p} \overline{\sigma}^2$. However, before taking the limit as $T \to \infty$ we can apply a change in variable which results in the extra factor h canceling from the latter two estimates. In addition, our estimates can also be shown to be valid under the standard large-N asymptotics with fixed shifts.

Proposition 5.1. *Under the conditions of Theorem 4.2, (4.5) 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 by using plug-in estimates. This allows feasible inference about the break date.

The results are presented in Figure 5-8 which also plot the classical shrinkage asymptotic distribution from Bai (1997). Here by signal-to-noise ratio we mean δ^0/σ_e which, given $\sigma_e^2 = 1$, equals the break size δ^0. We can summarize the results as follows. The finite-sample distribution shares all of the features characterizing the density of the infeasible continuous record distribution across all break magnitudes and break locations. Furthermore, the density of the feasible version of the continuous record asymptotic distribution provides a good approximation to the infeasible one and thus also to the finite-sample distribution. This holds for both stationary and non-stationary regimes. The latter case corresponds to the following modification of model (5.1) where we specify

$$Z_t = 0.5 Z_{t-1} + \sigma_{Z,t} e_{Z,t}, \qquad \sigma_{Z,t} = \begin{cases} 0.86, & t \leq T_b^0 \\ 1.20, & t > T_b^0 \end{cases}, \qquad \mathrm{Var}\left(e_t\right) = \begin{cases} 1, & t \leq T_b^0 \\ 2, & t > T_b^0 \end{cases},$$

so that the second moments of both the regressors and the errors roughly duplicates after T_b^0. Figure S-14-S-16 in the Supplement suggest interesting observations. First, the density of the finite-sample distribution is never symmetric even when $\lambda_0 = 0.5$. Second, it is always negatively skewed and the mode associated with the end sample point is higher than the mode associated with the starting sample point. Third, the density is never centered at the origin but slightly to the right of it. These features are easy to interpret. There is more variability in the post-break regime and it is more likely that the least-squares estimator overestimates the break date. The feasible density of the continuous record distribution provides a good approximation also in the case of non-stationary regimes. The supplementary material present additional results for a wide variety of models. In all cases, the feasible asymptotic distribution provides a good approximation to the finite-sample distribution.

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 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. When the distribution is unimodal and symmetric, e.g., for large break magnitudes, the HDR region coincides with the standard confidence interval symmetric about the estimate. All that is needed to implement the procedure is an estimate of the density function. Once estimable quantities are plugged-in as explained in Section 5, we derive the empirical counterpart of the limiting distribution. Choose some significance level $0 < \alpha < 1$ and let $\widehat{P}_{T_b}$ denote the empirical counterpart of the probability distribution of $\rho N\left(\widehat{\lambda}_b - \lambda_b^0\right)$ as defined in Theorem 4.2. Further, let $\widehat{p}_{T_b}$ denote the 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 is 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).

Definition 6.2. Confidence Sets for T_b^0 under a Continuous Record: Under Assumption 2.1-2.3, 3.1-3.2, 4.1-4.2 and under the "fast time scale", 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 (4.5) 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 in 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 9 reports the HDR of the feasible limiting distribution of $\rho\left(\widehat{T}_b - T_b^0\right)$ for a random draw from the model described by (5.1) with parameters $\pi^0 = 1$, $\beta^0 = 0$, unit second moments and autoregressive coefficient 0.6 for Z_t and $\sigma_e^2 = 1.2$ for the error term. We set $\lambda_0 = 0.35$, 0.5 and $\delta^0 = 0.3$, 0.8, 1.5. The sample size is $T = 100$ 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. In each plot, 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 $\rho\left(\widehat{T}_b - T_b^0\right)$ 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 $\delta^0 = 0.3$ 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 $C(\mathrm{cv}_{0.05}) = \{1, \ldots, 12\} \cup \{18, \ldots 100\}$. This includes the true break date T_b^0 and the overall length is 95 observations. Table 1 reports for each method the coverage rate and 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.

Figure 9 (middle panel) reports results for a larger break size $\delta^0 = 0.8$. The multi-modality is no longer present. When $\lambda_0 = 0.35$, the estimated break date is $\widehat{T}_b = 25$ and the length of $C(\mathrm{cv}_{0.05})$ is 27 out of 100 observations given by $C(\mathrm{cv}_{0.05}) = \{12, \ldots, 38\}$. Relative to Elliott and Müller's (2007) confidence sets which always cover T_b^0 in this example, the set constructed using the HDR is about 30% shorter. Bai's (1997) method is again shorter than the other methods but it fails to cover the true value when $\lambda_0 = 0.35$. However, it does so when $\lambda_0 = 0.5$ and its length is 18. In the latter case (right plot), our method covers the true break date and the interval has almost the same length whereas Elliott and Müller's (2007) approach results in an overall length of 35. Our method still provides accurate coverage when raising the break size to $\delta^0 = 1.5$ as can be seen from the bottom panel. When $\lambda_0 = 0.35$ (left panel), the estimated break date is $\widehat{T}_b = 36$ and all three methods cover the true break date. The confidence interval from Bai's (1997) method results in the shortest length since it includes only 8 points whereas our confidence interval includes 9 points and Elliott and Müller's (2007) method includes 24 points.

This single simulation, by and large, anticipates the small-sample results in the Monte Carlo study reported in the next Section: Bai's (1997) method results in a coverage probability below the nominal level; our method provides accurate coverage rates and the average length of the confidence set is always shorter than with Elliott and Müller's (2007) method. It is evident that the confidence set for T_b^0 constructed using the HDR provides a useful summary of the underlying probability distribution of the break point estimator. For small break sizes, the HDR captures well the bi- or tri-modality of the density. As we raise the magnitude of the break, the HDR becomes a single interval around the estimated break point, which is a desirable property.

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), which essentially involves the inversion of the likelihood-ratio test of Qu and Perron (2007). We omit the technical details of these methods and refer to the original sources or Chang and Perron (forthcoming) for a review and comparisons. The current state of this literature can be summarized as follows. The empirical coverage rates of the confidence intervals obtained from Bai's (1997) method are often below the nominal level when the magnitude of the break is small. EM's approach is by far the best in terms of providing an exact coverage rate that is closest to the nominal level. However, the lengths of the confidence sets are larger relative to the other methods, often by a very wide margin. The lenghts can be very large (e.g., the whole sample) even when the size of the break is very large; e.g., in models with serially correlated errors or with lagged dependent variables. The ILR-based confidence sets display a coverage probability often above the nominal level and this results in an average length larger than with Bai's (1997) method; further, it has a poor coverage probability for all break sizes in models with heteroskedastic errors and autocorrelated regressors. These findings suggest that there does not exist a method that systematically provides over a wide range of empirically relevant models both good coverage probabilities and reasonable lengths of the confidence sets, especially one that has good properties for all break sizes, whether large or small.

The results to be reported suggest that our approach has two notable properties. First, it provides adequate empirical coverage probabilities over all DGPs considered for any size and/or location of the break in the sample. Second, the lengths of the confidence sets are always shorter than those obtained with EM's approach. Oftentimes, the decrease in length is substantial and more so as the size of the break increases. To have comparable coverage rates, we can compare the lengths of our confidence sets with those obtained using Bai's (1997) method only when the size of the break is moderate to large. For those cases, our method delivers confidence sets with lengths only slightly larger and they become equivalent as the size of the break increases. Also, our HDR method has, overall, better coverage rates and shorter lengths compared to ILR.

We consider discrete-time DGPs of the form

$$y_t = D_t'\pi^0 + Z_t'\beta^0 + Z_t'\delta^0\mathbf{1}_{\left\{t>T_b^0\right\}} + e_t, \qquad\qquad t = 1, \ldots, T, \tag{7.1}$$

with $T = 100$ and, without loss of generality, $\pi^0 = 0$ (except in M4-M5, M7-M9). We consider ten versions of (7.1): M1 involves a break in the mean of an *i.i.d.* series with $Z_t = 1$ for all t, D_t absent, and $e_t \sim i.i.d.\, \mathscr{N}(0,\, 1)$; M2 is the same as M1 but with a simultaneous break in the variance such that $e_t = \left(1 + \mathbf{1}_{\left\{t>T_b^0\right\}}\right) u_t$ with $u_t \sim i.i.d.\, \mathscr{N}(0,\, 1)$; M3 is the same as M1 but with stationary Gaussian AR(1) disturbances $e_t = 0.3e_{t-1} + u_t$, $u_t \sim i.i.d.\, \mathscr{N}(0,\, 0.49)$; M4 is a partial structural change model with $D_t = 1$ for all t, $\pi^0 = 1$ and $Z_t = 0.5Z_t + u_t$ with $u_t \sim i.i.d.\, \mathscr{N}(0,\, 0.75)$ independent of $e_t \sim i.i.d.\, \mathscr{N}(0,\, 1)$; M5 is similar to M4 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 $\{Z_t\}$; M6 is the same as M3 but with u_t drawn from a t_ν distribution with $\nu = 5$ degrees of freedom; M7 is a model with a lagged dependent variable with $D_t = y_{t-1}$, $Z_t = 1$, $e_t \sim i.i.d.\, \mathscr{N}(0,\, 0.49)$, $\pi^0 = 0.3$ and $Z_t'\delta^0\mathbf{1}_{\left\{t>T_b^0\right\}}$ is replaced by $Z_t'\left(1-\pi^0\right)\delta^0\mathbf{1}_{\left\{t>T_b^0\right\}}$; M8 is the same as M7 but with $\pi^0 = 0.8$ and $e_t \sim i.i.d.\, \mathscr{N}(0,\, 0.04)$; M9 has FIGARCH(1,d,1) errors given by $e_t = \sigma_t u_t$, $u_t \sim \mathscr{N}(0,\, 1)$ and $\sigma_t = 0.1 + \left(1 - 0.2L\left(1-L\right)^d\right) e_t^2$ where $d = 0.6$ is the order of differencing and L the lag operator, $D_t = 1$, $\pi^0 = 1$ and $Z_t \sim i.i.d.\, \mathscr{N}(1,\, 1.44)$ independent of e_t. M10 is similar to M5 but with ARFIMA(0.3, d, 0) regressor Z_t with order of differencing $d = 0.5$, $\mathrm{Var}\left(Z_t\right) = 1$ and $e_t \sim \mathscr{N}(0,\, 1)$ independent of $\{Z_t\}$. We set $\beta^0 = 1$ in all models, except in M7-M8 where $\beta^0 = 0$.

We use the appropriate limit distribution in each case when applying Bai's (1997), ILR and our method. When the errors are uncorrelated, we simply estimate variances. For model M3, in order to estimate the long-run variance we use for all methods Andrews and Monahan's (1992) AR(1) pre-whitened two-stage procedure to select the bandwidth with a quadratic spectral kernel. Except for M2, we report results for the statistic $\widehat{U}_T$.eq proposed by EM, which imposes homogeneity across regimes (the results are qualitatively similar using the version $\widehat{U}_T$.neq, which allows heterogeneity across regimes). The methods of Bai (1997), the HDR and ILR all require an estimate of the break date. We use the least-squares estimate obtained with a trimming parameter $\epsilon = 0.15$. When constructing the confidence set, we apply to all methods the same trimming corresponding to the degrees of freedom of the EM's (2007) statistic. This amounts to eliminating from consideration a few observations at the beginning and end of the sample, i.e., the number of parameters being estimated. We set the significance level at $\alpha = 0.05$, and the break occurs at date $\lfloor T\lambda_0 \rfloor$, where $\lambda_0 = 0.2,\ 0.35,\ 0.5$. The results are presented in Table 2-11 for DGP M1-M10, respectively. 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). The sup-Wald rejection probability provides a measure of the magnitude of the break relative to the noise; low (large) values indicating a small (large) break. For models with predictable processes we use the procedure two-step described in Section A.2.

Note that for M9-M10, one cannot apply the result of Theorem 4.2 and the associated method to obtain a feasible estimate of the distribution. Thus, we resort to the more general Theorem 4.1 which is valid under stochastic variances. The methods used to estimate the distribution is presented in Section A.4.

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. For example, for M3 with $\lambda_0 = 0.35$ and $\delta^0 = 0.6$, it is below 85% even though the sup-Wald test rejection rate is roughly 70%. When the size of the break is smaller, it systematically fails to cover the true break date with correct probability. These features evidently translate into lengths of the confidence intervals which are relatively shorter than with other methods, but given the differences in coverage rate such comparisons are meaningless. Only for moderate to large shifts is it legitimate to compare our method with that of Bai (1997), in which case the confidence intervals are similar in length; our HDR method delivers confidence sets slightly larger for medium-sized breaks (e.g., $\delta^0 = 1.5$) and the differences vanish as the size of the break increases. Overall, our HDR method and that of EM show accurate empirical coverage rates for all DGP considered. The ILR shows coverage rates systematically above the nominal level and, hence, an average length significantly longer than from Bai's (1997) and our HDR methods in some cases (e.g., M2-M4), at least when the magnitude of the shift is small or moderate. As opposed to our HDR method, the ILR displays poor coverage rates for all break sizes in M5 which includes heteroskedastic errors.

The coverage rates of EM's method are the most accurate, indeed very close to the nominal level. Turning to the comparisons of the length of the confidence sets, EM's method almost always displays confidence sets which are larger than those from the other approaches. For example, for M3 with $\lambda_0 = 0.2$ and $\delta^0 = 0.6$, the average length of EM's confidence set is 78.61 while that of the HDR method is 50.61. Such results are not particular to M3, but remain qualitatively the same across all models. When the size of the break is small to moderate, the lengths of the confidence sets obtained using our HDR method are shorter than those obtained using EM's with differences ranging from 5% to 40%. The fact that EM's method provides confidence sets that are larger becomes even more apparent as the size of the break increases. 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. For example, when a lagged dependent variable is present (cf. M8), with $\lambda_0 = 0.5$ and $\delta^0 = 2$, the average length our our HDR confidence set is 6.34 while that of EM is 30.25, a reduction in length of about 75%. The results for M8, a change in mean with a lagged dependent 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 $\delta^0 = 2$, the average length from the HDR method is 8.34 compared to 93.71 with EM's. This is in line with the results in Chang and Perron (forthcoming).

Finally, to show that our asymptotic results are valid and still provide good approximations with long-memory volatility, we consider M9 and M10. The results show that Bai's method is not robust in that its coverage probability is below the nominal level even when the break magnitude is large. In contrast, the HDR-based method performs well and the average length of the confidence set is significantly shorter than that with EM's or the ILR method, especially when the break is not at mid-sample for the latter.

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 Conclusions

We examined a partial structural change model under a continuous record asymptotic framework. We established the consistency, rate of convergence and asymptotic distribution of the least-squares estimator under very mild assumptions on an underlying continuous-time data generating process. Contrary to the traditional large-N asymptotics, our asymptotic theory is able to provide good approximations and explain the following features. With the time horizon $[0, N]$ fixed, we can account for the asymmetric informational content provided by the pre- and post-break samples. The time span, the location of the break and the properties of the errors and regressors all jointly play a primary role in shaping the limit distribution of the estimate of the break date. The latter 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 derived a feasible counterpart using consistent plug-in estimates and show that it provides accurate approximations to the finite-sample distributions. In particular, the asymptotic and finite-sample distributions are (i) never symmetric unless the break point is located at mid-sample and the regimes are stationary, and (ii) positively (resp., negatively) skewed if the break point occurs in the first (resp., second)-half of the sample. This holds true across different break magnitudes except for very large break sizes in which case the distribution is symmetric. We used our limit theory to construct confidence sets for the break date based on the concept of Highest Density Region. Our method is simple to implement and relies entirely on the derived feasible asymptotic distribution. Overall, it delivers accurate coverage probabilities and relatively short average lengths of the confidence sets across a variety of data-generating mechanisms. Importantly, it does so irrespective of the magnitude of the break, whether large or small, a notoriously difficult problem in the literature.

References

Aït-Sahalia, Y., and J. Jacod (2014): *High-Frequency Financial Econometrics*. Princeton University Press.

Aldous, D., and G. Eagleson (1978): "On Mixing and Stability of Limit Theorems," *Annals of Probability*, 6(2), 325–331.

Andrews, D. W. K. (1992): "Generic Uniform Convergence," *Econometric Theory*, 8(2), 241–257.

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

——— (1994): "Empirical Process Methods in Econometrics," in *Handbook of Econometrics*, ed. by R. F. Engle, and D. L. McFadden, vol. 4, chap. 37, pp. 2247–2294. Amsterdam: Elsevier Science.

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

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

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

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

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

——— (2006): "Multiple Structural Change Models: A Simulation Analysis," in *Econometric Theory and Practice: Frontiers of Analysis and Applied Research*, ed. by S. D. D. Corbea, and B. E. Hansen, pp. 212–237. Cambridge University Press.

Barndorff-Nielsen, O. E., and N. Shephard (2004): "Econometric Analysis of Realised Covariation: High Frequency Based Covariance, Regression and Correlation in Financial Economics," *Econometrica*, 72(3), 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(2), 183–208.

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

Carrasco, M., and X. Chen (2002): "Mixing and Moment Properties of Various GARCH and Stochastic Volatility Models," *Econometric Theory*, 18(1), 17–39.

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

Christopeit, N. (1986): "Quasi-Least-Squares Estimation in Semimartingale Regression Models," *Stochastics*, 16, 255–278.

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

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

Doob, J. L. (1953): *Stochastic Processes*. New York: John Wiley and Sons.

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

Elliott, G., and U. K. Müller (2006): "Efficient Tests for General Persistent Time Variation in Regression Coefficients," *Review of Economic Studies*, 73(4), 907–940.

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

Galtchouk, L., and V. Konev (2001): "On Sequential Estimation of Parameters in Semimartingale Regression Models with Continuous Time Parameter," *The Annals of Statistics*, 29(5), 1508–1536.

Genon-Catalot, V., T. Jeantheau, and C. Laredo (2000): "Stochastic Volatility Models as Hidden Markov Models and Statistical Applications," *Bernoulli*, 6(6), 1051–1079.

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

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

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

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

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

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

Jacod, J. (1994): "Limit of Random Measures Associated with the Increments of a Brownian Semimartingale," *Discussion Paper, Université de Paris VI.*

——— (1997): "On Continuous Conditional Gaussian Martingales and Stable Convergence in Law," in *Seminaire de Probabilitiés de Strasbourg*, no. 31, pp. 232–246.

Jacod, J., and P. Protter (1998): "Asymptotic Error Distributions for the Euler Method for Stochastic Differential Equations," *Annals of Probability*, 26(1), 267–307.

——— (2012): *Discretization of Processes*. Berlin Heidelberg: Springer.

Jacod, J., and M. Rosenbaum (2013): "Quarticity and Other Functionals of Volatility: Efficient Estimation," *The Annals of Statistics*, 41(3), 1462–1484.

Jacod, J., and A. N. Shiryaev (2003): *Limit Theorems for Stochastic Processes*. Berlin Heidelberg: Springer-Verlag.

Jiang, L., X. Wang, and S. Yu (forthcoming): "New Distribution Theory for the Estimation of Structural Break Point in Mean," *Journal of Econometrics*.

Karatzas, I., and S. Shreve (1996): *Brownian Motion and Stochastic Calculus*, vol. 113 of *Graduate Texts in Mathematics*. New York: Springer-Verlag, 2nd edn.

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

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

Lai, T. L., and C. Z. Wei (1983): "Asymptotic Properties of General Autoregressive Models and Strong Consistency of Least-Squares Estimates of Their Parameters," *Journal of Multivariate Analysis*, 13(1), 1–23.

Laredo, C. F. (1990): "A Sufficient Condition for Asymptotic Sufficiency of Incomplete Observations of a Diffusion Process," *The Annals of Statistics*, 18(3), 1158–1171.

Li, J., V. Todorov, and G. Tauchen (2017): "Adaptive Estimation of Continuous-Time Regression Models Using High-Frequency Data," *Journal of Econometrics*, 200(1), 36–47.

Li, J., and D. Xiu (2016): "Generalized Method of Integrated Moments for High-Frequency Data," *Econometrica*, 84(4), 1613–1633.

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

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

Mel'nikov, A. V., and A. A. Novikov (1988): "Sequential Inference with Fixed Accuracy for Semimartingales," *Theory of Probability and its Applications*, 33(3), 480–494.

Meyer, P. (1967): *Intégrales Stochastiques*. Berlin Heidelberg: Springer Verlag.

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

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

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

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

Perron, P., and Y. Yamamoto (2015): “Using OLS to Estimate and Test for Structural Changes in Models with Endogenous Regressors,” *Journal of Applied Econometrics*, 30(1), 119–144.

Perron, P., and X. Zhu (2005): “Structural Breaks with Deterministic and Stochastic Trends,” *Journal of Econometrics*, 129(1–2), 65–119.

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

Protter, P. (2005): *Stochastic Integration and Differential Equations*. Berlin Heidelberg: Springer-Verlag.

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

Rényi, A. (1963): “On Stable Sequences of Events,” *Sankya Series*, 25, 293–302.

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

Sørensen, M., and M. Uchida (2003): “Small-diffusion Asymptotics for Discretely Sampled Stochastic Differential Equations,” *Bernoulli*, 9(6), 1051–1069.

Yao, Y. (1987): “Approximating the Distribution of the ML Estimate of the Change-Point in a Sequence in Independent Random Variables,” *The Annals of Statistics*, 15, 1321–1328.

A Appendix

A.1 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 $\left(\widetilde{\Omega},\, \widetilde{\mathscr{F}},\, \widetilde{P}\right)$ of the original space $\left(\Omega,\, \mathscr{F},\, \{\mathscr{F}_t\}_{t\geq 0},\, P\right)$. We also consider another filtration $\left\{\widetilde{\mathscr{F}}_t\right\}_{t\geq 0}$ 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 $\left(\Omega^*,\, \mathscr{F}^*,\, (\mathscr{F}_t^*)_{t\geq 0},\, P^*\right)$ 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 $\left(\widetilde{\Omega},\, \widetilde{\mathscr{F}},\, \left\{\widetilde{\mathscr{F}}_t\right\}_{t\geq 0},\, \widetilde{P}\right)$, and furthermore, conditionally on $\mathscr{F}$, is a two-sided centered continuous Gaussian process with independent increments and (conditional) covariance

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

where $\Omega_{\mathscr{W},1}^{(u,j)}(v) = \int_{N_b^0+v}^{N_b^0}\Omega_{Ze,s}^{(u,j)}ds$ and $\Omega_{\mathscr{W},2}^{(u,j)}(v) = \int_{N_b^0}^{N_b^0+v}\Omega_{Ze,s}^{(u,j)}ds$. 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).

A.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 (2.5). Let $\beta^0 = \left(\mu_1^0,\, \alpha_1^0,\, (\pi^0)',\, \left(\delta_{Z,1}^0\right)'\right)'$, $\delta^0 = \left(\mu_\delta^0,\, \alpha_\delta^0,\, \left(\delta_{Z,2}^0 - \delta_{Z,1}^0\right)'\right)'$, $\left((\beta^0)',\, \left((\delta^0)'\right)\right)' \in \Theta_0$, and $x_{kh} = \left(\left(\mu_{1,h}/\mu_1^0\right)h,\, \left(\alpha_{1,h}/\alpha_1^0\right)Y_{(k-1)h}h,\, \Delta_h D'_k,\, \Delta_h Z'_k\right)$. In matrix format, the model is $Y = X\beta^0 + Z_0\delta^0 + e$, where now X is $T \times (p+q+2)$ and $Z_0 = X\overline{R}$, $\overline{R} \triangleq \left[(I_2,\, 0_{2\times p})',\, \left(0'_{(p+q)\times 2},\, R\right)\right]'$, with R as defined in Section 2.1. Natural estimates of β^0 and δ^0 minimize the following criterion function,

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

Hence, we define our LS estimator as the minimizer of the following approximation to (A.2):

$$h^{-1}\sum_{k=1}^{T}\left(\Delta_h Y_k - \mu_1^h h - \alpha_1^h Y_{(k-1)h} h - \pi' \Delta_h D_k - \delta'_{Z,1}\Delta_h Z_k \mathbf{1}\left\{k \leq T_b\right\} - \delta'_{Z,2}\Delta_h Z_k \mathbf{1}\left\{k > T_b\right\}\right)^2 .$$

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 $\Delta_h \widetilde{V}_k = h^{1/2}\Delta_h V_k\left(\pi^0,\, \delta^0_{Z,1},\, \delta^0_{Z,2}\right)$ where

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

The small-dispersion format of our model is then

$$\Delta_h \widetilde{Y}_k = \left(\mu_1^0 h + \alpha_1^0 \widetilde{Y}_{(k-1)h} h\right)\mathbf{1}\left\{k \leq T_b^0\right\} + \left(\mu_2^0 h + \alpha_2^0 \widetilde{Y}_{(k-1)h} h\right)\mathbf{1}\left\{k > T_b^0\right\} + \Delta_h \widetilde{V}_k\left(\pi^0,\, \delta^0_{Z,1},\, \delta^0_{Z,2}\right). \qquad \text{(A.3)}$$

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

$$d\widetilde{Y}_t^0/dt = \left(\mu_1^0 + \alpha_1^0 \widetilde{Y}_t^0\right)\mathbf{1}\left\{t \leq N_b^0\right\} + \left(\mu_2^0 + \alpha_2^0 \widetilde{Y}_t^0\right)\mathbf{1}\left\{t > N_b^0\right\}. \qquad \text{(A.4)}$$

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 $\widehat{T}_b \triangleq \arg\max_{T_b} Q_T\left(T_b\right)$, where

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

and the LS estimates of the regression parameters are

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

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 A.1. *Assumption 2.3 and 3.2 hold. Assumption 2.1, 2.2 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 A.1. *Consider model (2.5). Under Assumption A.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, $P\left(T\left|\widehat{\lambda}_b - \lambda_0\right| > K\left\Vert\delta^0\right\Vert^{-2}\overline{\sigma}^2\right) < \varepsilon$.*

Assumption A.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 $T^{1-\kappa}\epsilon \to B < \infty$, $0 < \kappa < 1/2$, $\mathbb{E}\left[(\Delta_h e_t^*)^2 \,|\, \mathscr{F}_{t-h}\right] = \sigma_{h,t}^2 \Delta t$ 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 $\left(\widetilde{\Omega},\, \widetilde{\mathscr{F}},\, \left\{\widetilde{\mathscr{F}}_t\right\}_{t\geq 0},\, \widetilde{P}\right)$ of the original filtered probability space as explained in Section A.1. 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 A.1. *Consider model (A.3). Under Assumption A.1-A.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, $P\left(T^{1-\kappa}\left|\widehat{\lambda}_b - \lambda_0\right| > K\left\Vert\delta^0\right\Vert^{-2}\overline{\sigma}^2\right) < \varepsilon$; (iii) under the "fast time scale",*

$$N\left(\widehat{\lambda}_b - \lambda_0\right) \overset{\mathcal{L}-\mathrm{s}}{\Rightarrow} \underset{v\in\left[-\frac{N_b^0}{\left\Vert\delta^0\right\Vert^{-2}\overline{\sigma}^2},\, \frac{N-N_b^0}{\left\Vert\delta^0\right\Vert^{-2}\overline{\sigma}^2}\right]}{\operatorname{argmax}} \left\{-\left(\delta^0\right)'\Lambda(v)\,\delta^0 + 2\left(\delta^0\right)'\mathscr{W}(v)\right\}, \qquad \text{(A.5)}$$

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

$$\Lambda(v) \triangleq \begin{cases} \Lambda_1(v), & \text{if } v \le 0 \\ \Lambda_2(v), & \text{if } v > 0 \end{cases}, \qquad \text{with} \quad \Lambda_1(v) \triangleq \begin{bmatrix} \int_{N_b^0+v}^{N_b^0} ds & \int_{N_b^0+v}^{N_b^0} \widetilde{Y}_s ds & 0_{1\times p} \\ \int_{N_b^0+v}^{N_b^0} \widetilde{Y}_s ds & \int_{N_b^0+v}^{N_b^0} \widetilde{Y}_s^2 ds & 0_{1\times p} \\ 0_{p\times 1} & 0_{p\times 1} & \langle Z, Z\rangle_1(v) \end{bmatrix},$$

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* $\left(Z_\Delta^{(u)}, Z_\Delta^{(j)}\right)$, $3 \le u, j \le p$ *and* $v \le 0$*. The process* $\mathscr{W}(v)$ *is, conditionally on* $\mathscr{G}$*, a two-sided centered Gaussian martingale with independent increments.*

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

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

where $\mathscr{W}_{Ze}^{(i)}(v) \triangleq \sum_{r=1}^{p} \int_{N_b^0+v}^{N_b^0} \sigma_{Ze,s}^{(i,r)} dW_s^{1*(r)}$ $(i = 1, \dots, p)$ 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

$$\widetilde{\mathbb{E}}\left(\mathscr{W}^{(u)}(v)\,\mathscr{W}^{(j)}(v)\right) = \Omega_{\mathscr{W}}^{(u,j)}(v) = \begin{cases} \Omega_{\mathscr{W},1}^{(u,j)}(v), & \text{if } v \le 0 \\ \Omega_{\mathscr{W},2}^{(u,j)}(v), & \text{if } v > 0 \end{cases}, \tag{A.6}$$

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 $1 \le u, j \le 2$, $u \ne j$; $\Omega_{\mathscr{W},1}^{(u,j)}(v) = 0$, if $u = 1, 2$, $j = 3, \dots, p$; $\Omega_{\mathscr{W},1}^{(u,j)}(v) = \int_{N_b^0+v}^{N_b^0} \Omega_{Ze,s}^{(u-2,j-2)} ds$ if $3 \le u, j \le p+2$; 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 A.3. $\Sigma^* = \{\mu_{\cdot,t}, \Sigma_{\cdot,t}, \sigma_{e,t}\}_{t\ge 0}$ *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

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

Corollary A.1. *Under Assumption A.1-A.3,*

$$\frac{\left(\left(\delta^{0}\right)'\Lambda_{1}\delta^{0}\right)^{2}}{\left(\delta^{0}\right)'\Omega_{\mathscr{W},1}\delta^{0}}N\left(\widehat{\lambda}_{b}-\lambda_{0}\right)\Rightarrow\underset{s\in\left[-\frac{N_{b}^{0}}{\|\delta^{0}\|^{-2}\overline{\sigma}^{2}}\frac{\left(\left(\delta^{0}\right)'\Lambda_{1}\delta^{0}\right)^{2}}{\left(\delta^{0}\right)'\Omega_{\mathscr{W},1}\delta^{0}},\,\frac{N-N_{b}^{0}}{\|\delta^{0}\|^{-2}\overline{\sigma}^{2}}\frac{\left(\left(\delta^{0}\right)'\Lambda_{1}\delta^{0}\right)^{2}}{\left(\delta^{0}\right)'\Omega_{\mathscr{W},1}\delta^{0}}\right]}{\operatorname{argmax}}\mathscr{V}\left(s\right). \tag{A.7}$$

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

$$\mathscr{V}_{\mathrm{sta}}\left(s\right)=\begin{cases}-\frac{|s|}{2}+W_{1}^{*}\left(s\right), & \text{if } s<0\\ -\frac{\left(\delta^{0}\right)'\Lambda_{2}\delta^{0}}{\left(\delta^{0}\right)'\Lambda_{1}\delta^{0}}\frac{|s|}{2}+\left(\frac{\left(\delta^{0}\right)'\Lambda_{2}\delta^{0}}{\left(\delta^{0}\right)'\Lambda_{1}\delta^{0}}\right)^{1/2}W_{2}^{*}\left(s\right), & \text{if } s\geq0\end{cases},\quad \mathscr{V}_{\mu,\mathrm{sta}}\left(s\right)=\begin{cases}-\frac{|s|}{2}+W_{1}^{*}\left(s\right), & \text{if } s<0\\ -\frac{|s|}{2}+W_{2}^{*}\left(s\right), & \text{if } s\geq0\end{cases}.$$

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

$$\frac{\left(\delta^{0}\right)'\Lambda_{1}\delta^{0}}{\overline{\sigma}^{2}}N\left(\widehat{\lambda}_{b}-\lambda_{0}\right)\Rightarrow\underset{s\in\left[-\frac{N_{b}^{0}}{\|\delta^{0}\|^{-2}\overline{\sigma}^{2}}\frac{\left(\delta^{0}\right)'\Lambda_{1}\delta^{0}}{\overline{\sigma}^{2}},\,\frac{N-N_{b}^{0}}{\|\delta^{0}\|^{-2}\overline{\sigma}^{2}}\frac{\left(\delta^{0}\right)'\Lambda_{1}\delta^{0}}{\overline{\sigma}^{2}}\right]}{\operatorname{argmax}}\mathscr{V}_{\mathrm{sta}}\left(s\right).$$

Corollary A.3. *Under Assumption A.1-A.3, with $\pi^{0}=0$, $\delta_{Z,i}^{0}=0$, and $\alpha_{i}^{0}=0$ for $i=1,\,2$:*

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

Remark A.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, Wang, and Yu (forthcoming) under a continuous-time setting in their Theorem 4.2-(b) which is similar to our Corollary A.3. Our limit theory differs in many respects, besides being obviously more general. Jiang, Wang, and Yu (forthcoming) 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 A.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 $h\downarrow0$ whereas Jiang, Wang, and Yu (forthcoming) 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 (A.3), we propose a two-step estimator. In fact, (A.4) essentially implies that asymptotically the evolution of the dependent variable is governed by a deterministic drift function given by $\mu_1^0 + \alpha_1^0 \widetilde{Y}_t^0$ (resp., $\mu_2^0 + \alpha_2^0 \widetilde{Y}_t^0$) if $t \leq N_b^0$ (resp., $t > N_b^0$). Thus, in a first step we construct least-squares estimates of μ_i^0 and α_i^0 $(i = 1, 2)$. 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 A.2. *Under Assumption A.1-A.2, as $h \downarrow 0$, $\widehat{\theta} \xrightarrow{P} \theta^0$.*

The consistency of the estimate $\widehat{\theta}$ is all that is needed to carry out our inference procedures about the break point T_b^0 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.

A.3 Proofs of Theorem 4.1-4.2

A.3.1 Proof of Theorem 4.1

Proof. Let us focus on the case $T_b(v) \leq T_b^0$ (i.e., $v \leq 0$). 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 $Q_T(T_b(v)) - Q_T(T_b^0)$ is determined by that of $-\delta_h' (Z_\Delta' Z_\Delta) \delta_h \pm 2\delta_h' (Z_\Delta' e)$. 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 $Q_T(T_b(u)) - Q_T(T_b^0)$ 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

$$dZ_{\psi,s} = \psi_h^{-1/2} \sigma_{Z,s} dW_{Z,s}, \qquad dW_{\psi,e,s} = \psi_h^{-1/2} \sigma_{e,s} dW_{e,s}, \qquad \text{with } s \in \mathcal{D}(C). \tag{A.8}$$

For $s \in [N_b^0 - Ch^{1-\kappa}, \, N_b^0 + Ch^{1-\kappa}]$ let $v = \psi_h^{-1}(N_b^0 - s)$, 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

$$dZ_{\psi,t} = \sigma_{Z,t} dW_{Z,t}, \qquad dW_{\psi,e,t} = \sigma_{e,t} dW_{e,t}, \qquad \text{with } t \in \mathcal{D}^*(C). \tag{A.9}$$

This can be used into the following quantities for $N_b(v) \in \mathcal{D}(C)$. 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 (A.8)-(A.9) is such that

$$\psi_h^{-1} Z'_\Delta Z_\Delta = \sum_{k=T_b^0+\lfloor v/h \rfloor}^{T_b^0} z_{kh} z'_{kh}, \qquad v \in \mathcal{D}^*(C). \tag{A.10}$$

Using the same argument:

$$\psi_h^{-1} Z'_\Delta \widetilde{e} = \sum_{k=T_b^0+\lfloor v/h \rfloor}^{T_b^0} z_{kh} \widetilde{e}_{kh}, \qquad v \in \mathcal{D}^*(C). \tag{A.11}$$

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

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

and thus, when multiplied by $h^{-1/2}$, we have $\overline{Q}_T(T_b) = -\left(\delta^0\right)' Z'_\Delta Z_\Delta \left(\delta^0\right) \pm 2\left(\delta^0\right)' \left(h^{-1/2} Z'_\Delta \widetilde{e}\right) + o_p(1)$, since on $\mathcal{D}^*(C)$, $e_{kh} \sim$ i.n.d. $\mathscr{N}\left(0,\, \sigma_{h,k-1}^2 h\right)$, $\sigma_{h,k} = O\left(h^{-1/4}\right) \sigma_{e,k}$ and $\widetilde{e}_{kh}$ is the normalized error [i.e., $\widetilde{e}_{kh} \sim$ i.n.d. $\mathscr{N}\left(0,\, \sigma_{e,k-1}^2 h\right)$] defined in (4.1). Hence, according to the re-parametrization introduced in the main text, we examine the behavior of

$$\overline{Q}_T(\theta^*) = -\left(\delta^0\right)' \left(\sum_{k=T_b+1}^{T_b^0} z_{kh} z'_{kh}\right) \delta^0 + 2\left(\delta^0\right)' \left(h^{-1/2} \sum_{k=T_b+1}^{T_b^0} z_{kh} \widetilde{e}_{kh}\right). \tag{A.12}$$

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 $\left\{\overline{Q}_T(\cdot),\, T \geq 1\right\}$. Let us associate to the continuous-time index t a corresponding $\mathcal{D}^*(C)$-specific index t_v. This means that each t_v identifies a distinct t in $\mathcal{D}^*(C)$ through v as define above. More specifically, for each $(\cdot,\, v) \in \mathcal{D}^*(C)$, define the new functions

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

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 $[Z,\, Z]_1(v) \triangleq [Z,\, Z]_{h\left\lfloor N_b^0/h \right\rfloor} -$

$[Z, Z]_{h\lfloor t_v/h \rfloor}$, and recall by Assumption 2.2 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 A.1. By Theorem 5.4.2 in Jacod and Protter (2012) we have $\mathscr{W}_h(v) \overset{\mathcal{L}-\mathrm{s}}{\Rightarrow} \mathscr{W}_1(v)$ 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: $\overline{Q}_T(\theta^*) \overset{\mathcal{L}_f-\mathrm{s}}{\to} -(\delta^0)' \langle Z, Z\rangle_1(v)\delta^0 + 2(\delta^0)' \mathscr{W}_1(v)$, where $\overset{\mathcal{L}_f-\mathrm{s}}{\to}$ signifies finite-dimensional stable convergence in law. Similarly, for $v > 0$, $\overline{Q}_T(\theta^*) \overset{\mathcal{L}_f-\mathrm{s}}{\to} -(\delta^0)' \langle Z, Z\rangle_2(v)\delta^0 + 2(\delta^0)' \mathscr{W}_2(v)$. Next, we verify the asymptotic stochastic equicontinuity of the sequence of processes $\left\{\overline{Q}_T(\cdot), T \geq 1\right\}$.[4] For $1 \leq i \leq p$, let $\zeta_{h,k}^{(i)} \triangleq z_{kh}^{(i)} \widetilde{e}_{kh}$, $\zeta_{h,k}^{*(i)} \triangleq \mathbb{E}\left[z_{kh}^{(i)} \widetilde{e}_{kh} | \mathscr{F}_{(k-1)h}\right]$, 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 $\overline{Q}_T^{\mathrm{c}}(\theta^*) \triangleq \overline{Q}_T(\theta^*) + (\delta^0)' \langle Z, Z\rangle_1(v)\delta^0$ (if $v \leq 0$, and defined analogously for $v > 0$),

$$\overline{Q}_T^{\mathrm{c}}(\theta^*) = \sum_{r=1}^{4} \overline{Q}_{r,T}(\theta^*), \tag{A.13}$$

where $\overline{Q}_{1,T}(\theta^*) \triangleq -(\delta^0)' \left(\sum_k \zeta^*{}_{Z,h,k}\right)\delta^0$, $\overline{Q}_{2,T}(\theta^*) \triangleq -(\delta^0)' \left(\sum_k \zeta^{**}{}_{Z,h,k}\right)\delta^0$, $\overline{Q}_{3,T}(\theta^*) \triangleq (\delta^0)' \left(h^{-1/2} \sum_k \zeta^*{}_{h,k}\right)$, and $\overline{Q}_{4,T}(\theta^*) \triangleq (\delta^0)' \left(h^{-1/2} \sum_k \zeta^{**}{}_{h,k}\right)$; where $\sum_k$ stands for $\sum_{k=T_b^0+\lfloor v/h \rfloor}^{T_b^0}$. We have

$$\sup_{(\theta, v) \in \mathcal{D}^*(C)} \left\| \overline{Q}_{3,T}(\theta^*) \right\| \leq K \left\| \delta^0 \right\| h^{-1/2} \sum_k \left\| \zeta_{h,k}^* \right\| \overset{P}{\to} 0, \tag{A.14}$$

which follows from Jacod and Rosenbaum (2013) given that $\Sigma_{Ze,k} = 0$ identically by Assumption 2.1-(iv). 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 (θ, v), $\left(\bar{\theta}, \bar{v}\right)$ 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

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

[4] 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.

$$\leq K\Big(\sum_{k=T_b^0+1}^{T_b(\bar{v})} \left\|\zeta_{Z,h,k}^*\right\| \left\|c_{p\times 1}\right\|^2 + \left\|\sum_{k=T_b^0+1}^{T_b^0+\lfloor \bar{v}/h\rfloor} \zeta_{Z,h,k}^* - \sum_{k=T_b^0+\lfloor v/h\rfloor}^{T_b^0} \zeta_{Z,h,k}^*\right\| \|\delta\|^2$$

$$\leq K\left(\left(pc^2\right)\sum_{k=T_b^0+1}^{T_b^0+\lfloor \bar{v}/h\rfloor} \left\|\zeta_{Z,h,k}^*\right\| + \sum_{k=T_b^0+\lfloor v/h\rfloor}^{T_b(\bar{v})} \left\|\zeta_{Z,h,k}^*\right\| \|\delta\|^2\right).$$

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

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

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

$$\left(\mathbb{E}\left[\sup_{(\theta,\, v)\in\mathcal{D}^*(C)} \left|\overline{Q}_{2,T}\left(\theta^*\right)\right|\right]^r\right)^{1/r} \leq K_r \left\|\delta^0\right\|^2 h^{-2/r} h \leq K_r h^{1-2/r} \to 0, \qquad \text{(A.15)}$$

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

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

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

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

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

$$\limsup_{h\downarrow 0} P\left[\sup_{\theta^*\in\Theta:\,\bar{\theta}^*\in\mathbf{B}(\tau,\,\theta^*)} \left|\overline{Q}_{4,T}(\theta^*) - \overline{Q}_{4,T}\left(\bar{\theta}^*\right)\right| > \eta\right], \tag{A.16}$$

can be made arbitrary less than $\varepsilon > 0$ as $h \downarrow 0$, by choosing τ small enough. Combining (A.14), (A.15) and (A.16), we conclude that the process $\left\{\overline{Q}_T(\theta, v), T \geq 1\right\}$ is asymptotically stochastic equicontinuous. Since the finite-dimensional convergence was demonstrated above, this suffices to guarantee the stable convergence in law of the process $\left\{\overline{Q}_T(\theta, v), T \geq 1\right\}$ toward a two-sided Gaussian limit process with drift $(\delta^0)'[Z, Z]_{\cdot}(\cdot)\,\delta^0$, having P-a.s. continuous sample paths with $\mathscr{F}$-conditional covariance matrix given in (A.1). Because $N\left(\widehat{\lambda}_b - \lambda_0\right) = O_p(1)$ 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 A.3.3, a result which shows the negligibility of the drift term, the proof of Theorem 4.1 is concluded. □

A.3.2 Proof of Theorem 4.2

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

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

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

$$\underset{v\in\left[-\frac{N_b^0}{\|\delta^0\|^{-2}\overline{\sigma}^2},\,\frac{N-N_b^0}{\|\delta^0\|^{-2}\overline{\sigma}^2}\right]}{\operatorname{argmax}} \mathscr{H}(v) \overset{d}{\equiv} \underset{s\in\left[-\frac{N_b^0}{\|\delta^0\|^{-2}\overline{\sigma}^2}\frac{\left((\delta^0)'\langle Z, Z\rangle_1\delta^0\right)^2}{(\delta^0)'\Omega_{\mathscr{W},1}(\delta^0)},\,\frac{N-N_b^0}{\|\delta^0\|^{-2}\overline{\sigma}^2}\frac{\left((\delta^0)'\langle Z, Z\rangle_1\delta^0\right)^2}{(\delta^0)'\Omega_{\mathscr{W},1}(\delta^0)}\right]}{\operatorname{argmax}} \mathscr{V}(s),$$

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 any function $f(s)$, $\arg\max_s cf(s) = \arg\max_s f(s)$. Thus,

$$\underset{v \in \left[-\frac{N_b^0}{\|\delta^0\|^{-2}\overline{\sigma}^2}, \frac{N-N_b^0}{\|\delta^0\|^{-2}\overline{\sigma}^2}\right]}{\text{argmax}} \mathscr{H}(v)$$

$$\overset{d}{\equiv} \underset{s \in \left[-\frac{N_b^0}{\|\delta^0\|^{-2}\overline{\sigma}^2} \frac{\left((\delta^0)'\langle Z, Z\rangle_1 \delta^0\right)^2}{(\delta^0)'\Omega_{\mathscr{W},1}(\delta^0)}, \frac{N-N_b^0}{\|\delta^0\|^{-2}\overline{\sigma}^2} \frac{\left((\delta^0)'\langle Z, Z\rangle_1 \delta^0\right)^2}{(\delta^0)'\Omega_{\mathscr{W},1}(\delta^0)}\right]}{\text{argmax}} \left(\frac{\left((\delta^0)'\langle Z, Z\rangle_1 \delta^0\right)^2}{(\delta^0)'\Omega_{\mathscr{W},1}(\delta^0)}\right)^{-1} \mathscr{V}(s),$$

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

$$\frac{\left((\delta^0)'\langle Z, Z\rangle_1 \delta^0\right)^2}{(\delta^0)'\Omega_{\mathscr{W},1}(\delta^0)} N\left(\widehat{\lambda}_b - \lambda_0\right) \Rightarrow \underset{s \in \left[-\frac{N_b^0}{\|\delta^0\|^{-2}\overline{\sigma}^2} \frac{\left((\delta^0)'\langle Z, Z\rangle_1 \delta^0\right)^2}{(\delta^0)'\Omega_{\mathscr{W},1}(\delta^0)}, \frac{N-N_b^0}{\|\delta^0\|^{-2}\overline{\sigma}^2} \frac{\left((\delta^0)'\langle Z, Z\rangle_1 \delta^0\right)^2}{(\delta^0)'\Omega_{\mathscr{W},1}(\delta^0)}\right]}{\text{argmax}} \mathscr{V}(s).$$

This concludes the proof. □

A.3.3 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

$$dZ_{\psi,s} = \psi_h^{-1/2}\mu_{Z,s}ds + \psi_h^{-1/2}\sigma_{Z,s}dW_{Z,s}, \qquad dW_{\psi,e,s} = \psi_h^{-1/2}\sigma_{e,s}dW_{e,s}, \qquad \text{with } s \in \mathcal{D}(C). \tag{A.17}$$

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 (A.9), $dZ_{\psi,t} = \psi_h^{\vartheta}\mu_{Z,t}dt + \sigma_{Z,t}dW_{Z,t}$ 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 (A.12) remains the same when $\mu_{\cdot,t}$ are nonzero. We set for any $1 \leq i \leq p$ and $1 \leq j \leq q+p$, $\mu_{Z,k}^{*(i)} \triangleq \int_{(k-1)h}^{kh}\mu_{Z,s}^{(i)}ds$, $\mu_{X,k}^{*(j)} \triangleq \int_{(k-1)h}^{kh}\mu_{X,s}^{(j)}ds$, $z_{0,kh}^{(i)} \triangleq \sum_{r=1}^{p}\int_{(k-1)h}^{kh}\sigma_{Z,s}^{(i,r)}dW_Z^{(r)}$ and $x_{0,kh}^{(j)} \triangleq \sum_{r=1}^{q+p}\int_{(k-1)h}^{kh}\sigma_{X,s}^{(j,r)}dW_X^{(r)}$. 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,

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

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.

A.4 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 $\widehat{N}_b$ (cf. Proposition 4.1). The ratio $\left\|\delta^0\right\|^2 / \overline{\sigma}^2$ is consistently estimated by $\left\|\widehat{\delta}\right\|^2 / \left(T^{-1} \sum_{k=1}^{T} \widehat{e}_{kh}^2\right)$ 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 $\left\{-\left(\delta^0\right)' \left\langle Z_\Delta,\, Z_\Delta\right\rangle (v)\, \delta^0 + 2\left(\delta^0\right)' \mathscr{W}(v)\right\}$. For $v \leq 0$, this can be consistently estimated by

$$-T^{1/2}\left[\left(\widehat{\delta}\right)' \left(\sum_{k=\widehat{T}_b+\lfloor v/h\rfloor}^{\widehat{T}_b} z_{kh} z_{kh}'\right) \widehat{\delta} - 2\widehat{\delta}' \widehat{\mathscr{W}_h}(v)\right], \tag{A.18}$$

where $\widehat{\mathscr{W}_h}$ is a simple-size dependent sequence of Gaussian processes whose marginal distribution is characterized by $h^{1/2} T \sum_{k=\widehat{T}_b^{\mathrm{LS}}+\lfloor v/h\rfloor}^{\widehat{T}_b^{\mathrm{LS}}} \widehat{e}_{kh}^2 z_{kh} z_{kh}'$ 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 estimator with (A.18) 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.

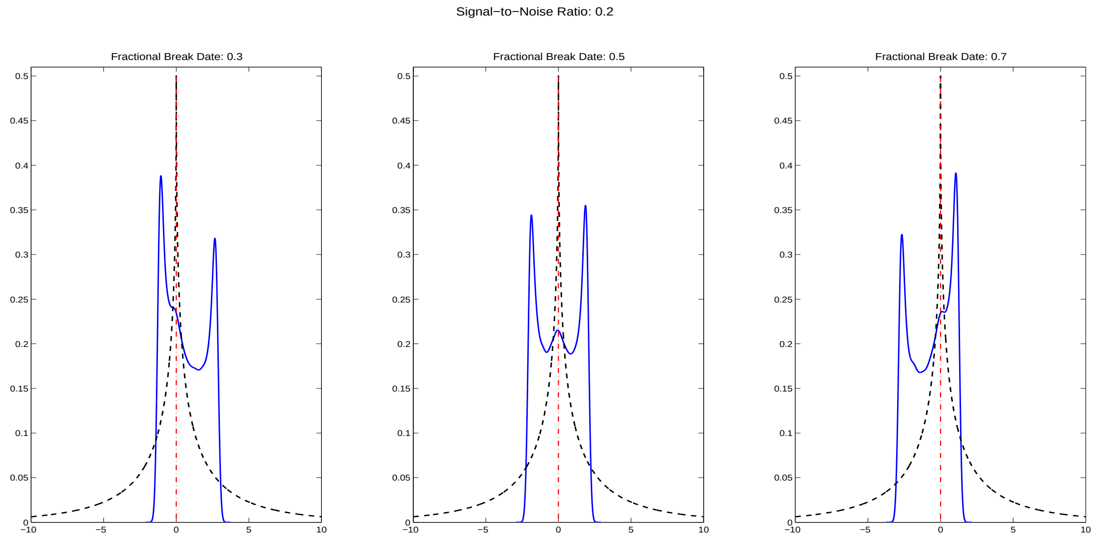

Figure 1: The limit probability density of $\rho\left(\widehat{T}_b - T_b^0\right)$ under a continuous record (solid line) and the density of the asymptotic distribution in Bai (1997) (broken line) when $\rho^2 = 0.2$ and the true fractional break point $\lambda_0 = 0.3$, 0.5 and 0.7 (the left, middle and right panel, respectively).

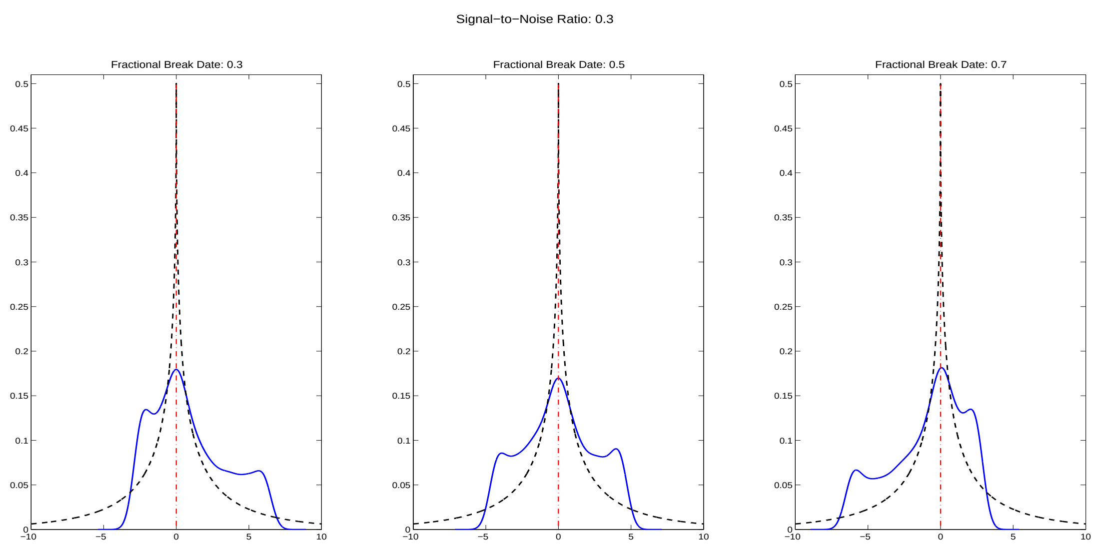

Figure 2: The limit probability density of $\rho\left(\widehat{T}_b - T_b^0\right)$ under a continuous record (solid line) and the density of the asymptotic distribution in Bai (1997) (broken line) when $\rho^2 = 0.3$ and the true fractional break point $\lambda_0 = 0.3$, 0.5 and 0.7 (the left, middle and right panel, respectively).

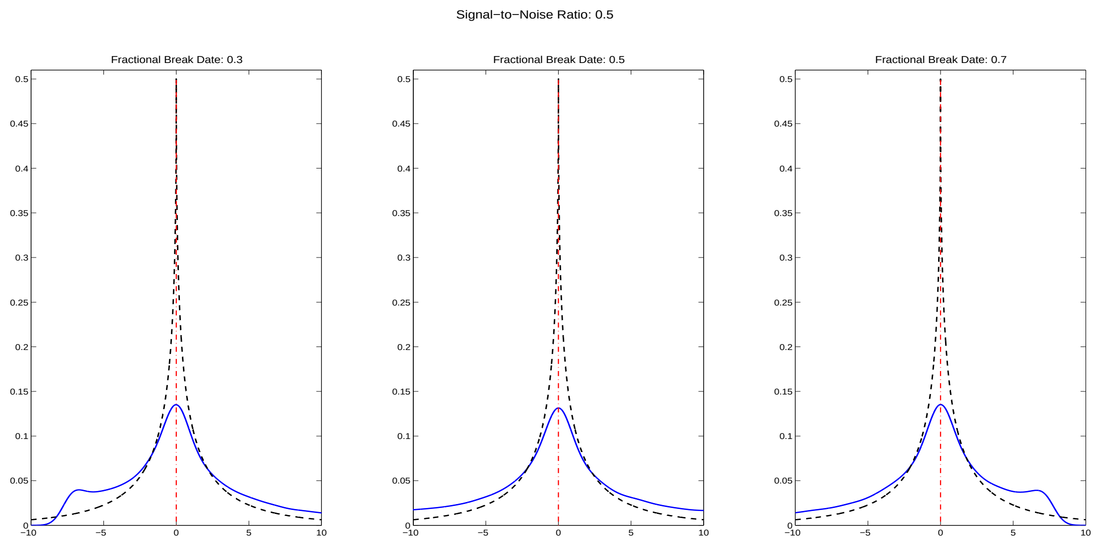

Figure 3: The limit probability density of $\rho\left(\widehat{T}_b - T_b^0\right)$ under a continuous record (solid line) and the density of the asymptotic distribution in Bai (1997) (broken line) when $\rho^2 = 0.5$ and true fractional break date $\lambda_0 = 0.3$, 0.5 and 0.7 (the left, middle and right panel, respectively).

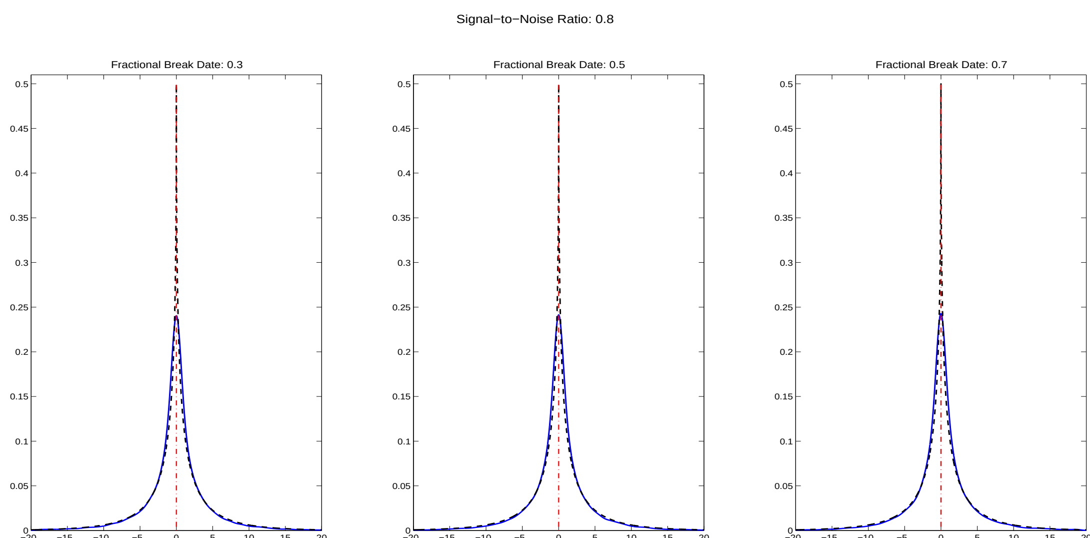

Figure 4: The limit probability density of $\rho\left(\widehat{T}_b - T_b^0\right)$ under a continuous record (solid line) and the density of the asymptotic distribution in Bai (1997) (broken line) when $\rho^2 = 0.8$ and the true fractional break date $\lambda_0 = 0.3$, 0.5 and 0.7 (the left, middle and right panel, respectively).

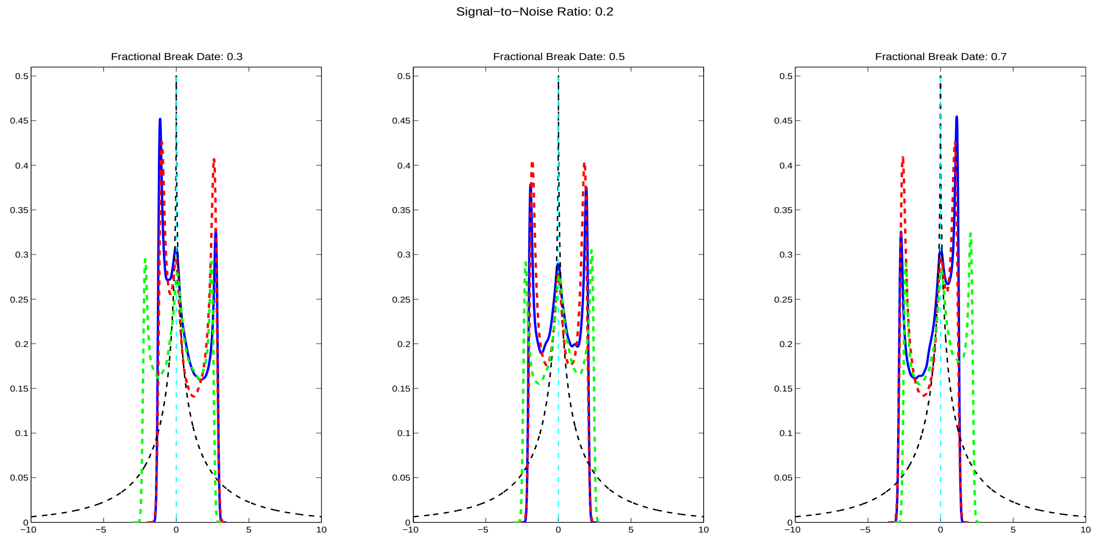

Figure 5: The probability density of $\rho\left(\widehat{T}_b - T_b^0\right)$ for model (5.1) with break magnitude $\delta^0 = 0.2$ and true break fraction $\lambda_0 = 0.3$, 0.5 and 0.7 (the left, middle and right panel, respectively). The signal-to-noise ratio is $\delta^0/\sigma_e = \delta^0$ since $\sigma_e^2 = 1$. 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.

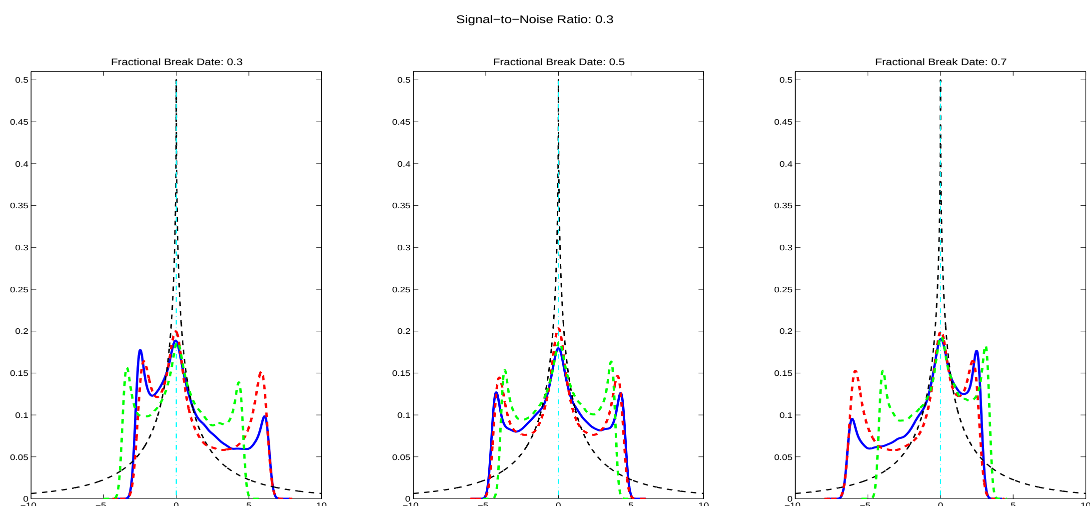

Figure 6: The probability density of $\rho\left(\widehat{T}_b - T_b^0\right)$ for model (5.1) with break magnitude $\delta^0 = 0.3$ and true break fraction $\lambda_0 = 0.3$, 0.5 and 0.7 (the left, middle and right panel, respectively). The signal-to-noise ratio is $\delta^0/\sigma_e = \delta^0$ since $\sigma_e^2 = 1$. 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.

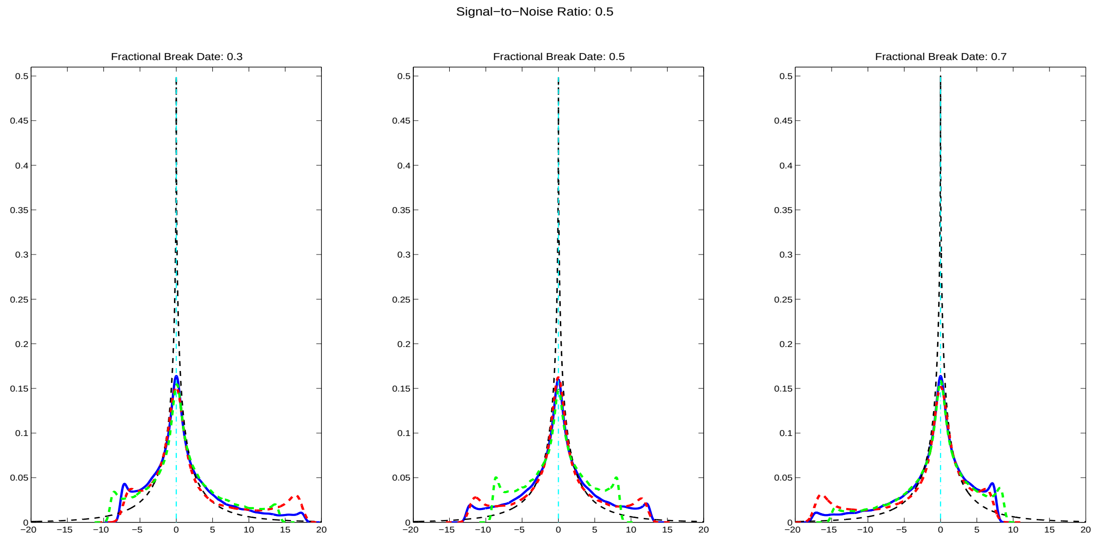

Figure 7: The probability density of $\rho\left(\widehat{T}_b - T_b^0\right)$ for model (5.1) with break magnitude $\delta^0 = 0.5$ and true break fraction $\lambda_0 = 0.3$, 0.5 and 0.7 (the left, middle and right panel, respectively). The signal-to-noise ratio is $\delta^0/\sigma_e = \delta^0$ since $\sigma_e^2 = 1$. 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.

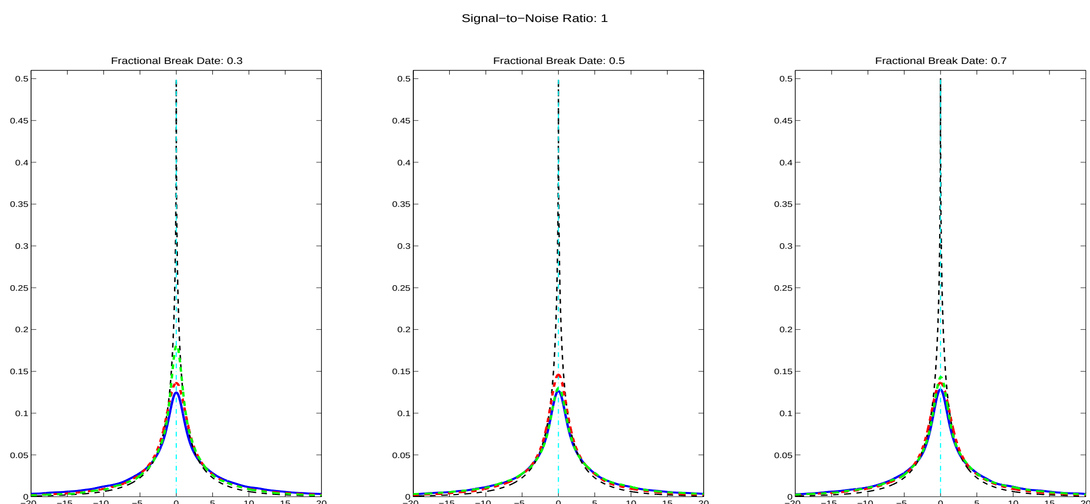

Figure 8: The probability density of $\rho\left(\widehat{T}_b - T_b^0\right)$ for model (5.1) with break magnitude $\delta^0 = 1$ and true break fraction $\lambda_0 = 0.3$, 0.5 and 0.7 (the left, middle and right panel, respectively). The signal-to-noise ratio is $\delta^0/\sigma_e = \delta^0$ since $\sigma_e^2 = 1$. 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.

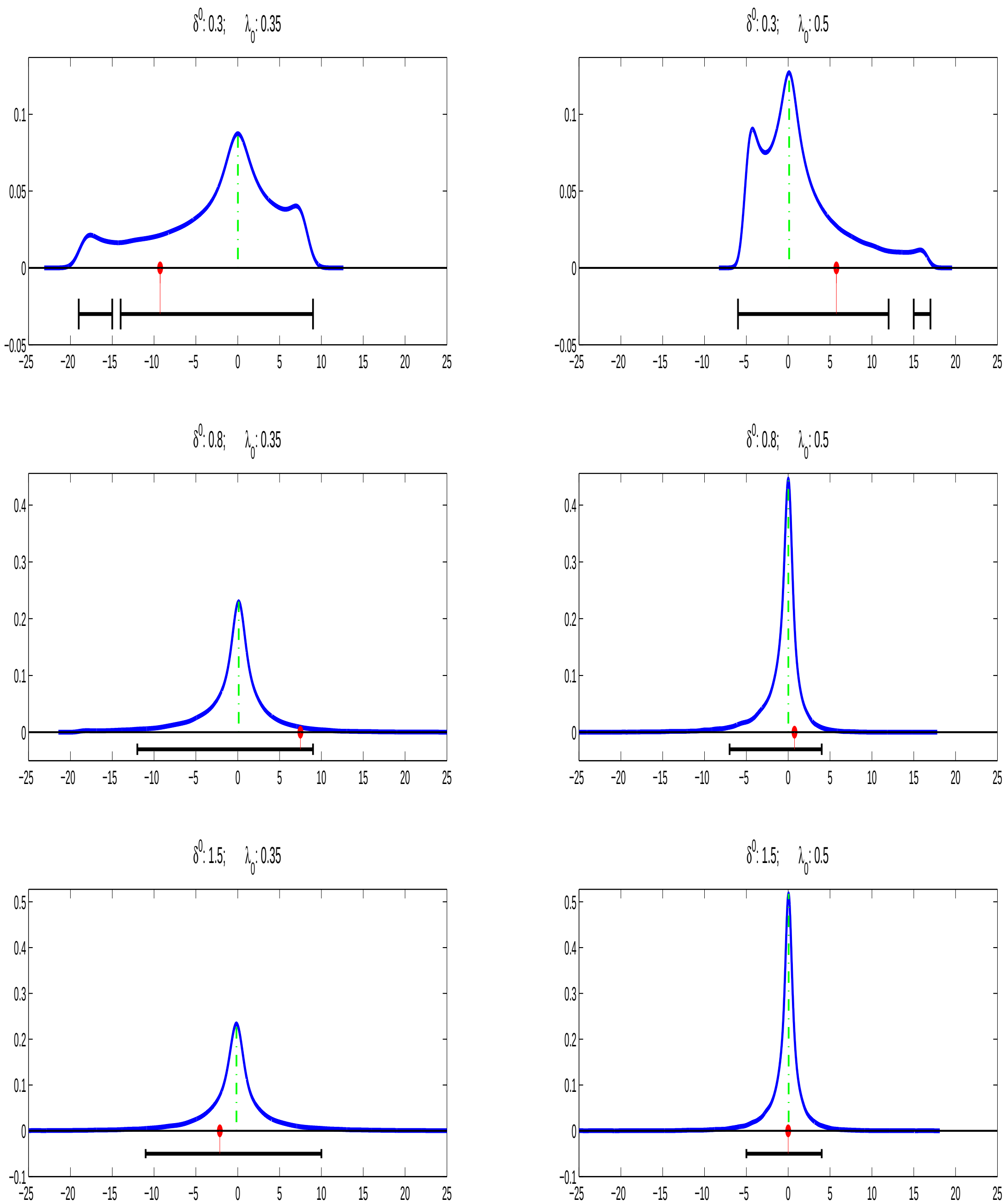

Figure 9: Highest Density Regions (HDRs) of the feasible probability density of $\rho\left(\widehat{T}_b - T_b^0\right)$ as described in Section 6. The significance level is $\alpha = 0.05$, the true break point is $\lambda_0 = 0.3$ and 0.5 (the left and right panels, respectively) and the break magnitude is $\delta^0 = 0.3$, 0.8 and 1.5 (the top, middle and bottom panels, respectively). The union of the black lines below the horizontal axis is the 95% HDR confidence region.

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

	$\delta^0 = 0.3$		$\delta^0 = 0.8$		$\delta^0 = 1.5$	
	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 9. 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

		$\delta^0 = 0.3$		$\delta^0 = 0.6$		$\delta^0 = 1$		$\delta^0 = 1.5$		$\delta^0 = 2$	
		Cov.	Lgth.	Cov.	Lgth.	Cov.	Lgth.	Cov.	Lgth.	Cov.	Lgth.
$\lambda_0 = 0.5$	HDR	0.956	75.63	0.940	65.39	0.949	35.96	0.969	12.53	0.960	5.93
	Bai (1997)	0.814	66.67	0.890	41.73	0.931	20.28	0.936	9.22	0.960	5.62
	$\widehat{U}_T$.eq	0.948	82.64	0.948	59.16	0.948	29.32	0.953	16.25	0.953	11.58
	ILR	0.955	83.22	0.954	55.97	0.969	21.65	0.974	8.49	0.983	4.56
	sup-W	0.202		0.455		0.912		0.999		1.000	
$\lambda_0 = 0.35$	HDR	0.958	74.01	0.948	61.50	0.951	33.84	0.954	12.14	0.965	5.95
	Bai (1997)	0.839	66.12	0.850	41.85	0.901	19.40	0.938	9.18	0.963	5.58
	$\widehat{U}_T$.eq	0.953	83.32	0.950	61.17	0.950	30.09	0.950	16.15	0.949	11.45
	ILR	0.949	83.15	0.960	58.69	0.966	22.94	0.975	8.25	0.985	4.06
	sup-W	0.192		0.651		0.983		1.000		1.000	
$\lambda_0 = 0.2$	HDR	0.901	73.55	0.934	57.34	0.968	31.15	0.975	12.59	0.967	6.16
	Bai (1997)	0.837	64.44	0.890	41.73	0.931	20.28	0.946	9.42	0.958	5.63
	$\widehat{U}_T$.eq	0.950	85.48	0.950	69.84	0.950	38.52	0.950	16.59	0.950	11.23
	ILR	0.953	85.71	0.958	66.48	0.967	29.42	0.976	9.214	0.981	4.81
	sup-W	0.118		0.405		0.878		0.998		1.000	

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

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

		$\delta^0 = 0.3$		$\delta^0 = 0.6$		$\delta^0 = 1$		$\delta^0 = 1.5$		$\delta^0 = 2$	
		Cov.	Lgth.	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	0.900	33.73	0.934	26.11
	Bai (1997)	0.854	70.60	0.843	58.27	0.857	40.70	0.894	23.33	0.923	14.24
	$\widehat{U}_T$.neq	0.961	88.95	0.961	80.33	0.961	61.15	0.961	39.69	0.964	32.16
	ILR	0.989	92.53	0.985	84.06	0.977	58.05	0.974	26.19	0.958	12.31
	sup-W	0.121		0.349		0.747		0.988		1.000	
$\lambda_0 = 0.35$	HDR	0.976	89.81	0.961	83.26	0.935	64.87	0.900	38.19	0.934	26.11
	Bai (1997)	0.823	69.86	0.822	55.87	0.844	38.91	0.898	23.56	0.932	14.24
	$\widehat{U}_T$.neq	0.963	89.84	0.963	82.26	0.961	65.87	0.961	43.63	0.964	32.16
	ILR	0.990	93.48	0.985	88.693	0.982	68.23	0.979	32.77	0.977	15.45
	sup-W	0.121		0.368		0.789				1.000	
$\lambda_0 = 0.2$	HDR	0.978	90.39	0.975	85.89	0.934	70.05	0.954	44.17	0.957	29.63
	Bai (1997)	0.782	70.24	0.805	56.37	0.831	37.66	0.897	23.19	0.928	14.80
	$\widehat{U}_T$.neq	0.968	91.11	0.968	87.62	0.972	78.17	0.968	60.80	0.967	46.24
	ILR	0.980	93.32	0.981	91.60	0.978	81.60	0.978	49.04	0.981	22.60
	sup-W	0.098		0.262		0.628		0.938		0.995	

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

Table 4: Small-sample coverage rate and length of the confidence set for model M3

		$\delta^0 = 0.3$		$\delta^0 = 0.6$		$\delta^0 = 1$		$\delta^0 = 1.5$		$\delta^0 = 2$	
		Cov.	Lgth.	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	0.945	11.29	0.971	5.85
	Bai (1997)	0.833	66.34	0.834	41.32	0.895	18.63	0.942	8.982	0.969	5.49
	$\widehat{U}_T$.eq	0.958	87.16	0.968	71.47	0.958	45.82	0.957	30.73	0.957	28.01
	ILR	0.932	79.38	0.944	53.48	0.966	21.98	0.986	8.59	0.993	4.87
	sup-W	0.314		0.749		0.990		1.000		1.000	
$\lambda_0 = 0.35$	HDR	0.910	70.98	0.902	53.88	0.917	28.07	0.948	11.18	0.973	5.99
	Bai (1997)	0.849	65.13	0.840	40.43	0.900	18.69	0.949	9.01	0.974	5.49
	$\widehat{U}_T$.eq	0.960	87.46	0.961	72.79	0.962	46.44	0.961	31.39	0.961	28.03
	ILR	0.942	80.94	0.946	55.20	0.965	23.55	0.983	8.88	0.993	4.93
	sup-W	0.308		0.705		0.990		1.000		1.000	
$\lambda_0 = 0.2$	HDR	0.905	72.26	0.913	50.61	0.933	25.07	0.947	11.10	0.973	6.35
	Bai (1997)	0.829	65.56	0.899	41.42	0.932	19.62	0.951	9.20	0.966	5.55
	$\widehat{U}_T$.eq	0.962	88.77	0.968	78.61	0.963	57.87	0.968	37.15	0.965	29.88
	ILR	0.938	83.24	0.951	63.66	0.972	28.94	0.985	10.18	0.994	5.16
	sup-W	0.272		0.595		0.921		0.997		0.999	

The model is $y_t = \beta^0 + \delta^0 \mathbf{1}_{\{t > \lfloor T\lambda_0 \rfloor\}} + e_t$, $e_t = 0.3 e_{t-1} + u_t$, $u_t \sim i.i.d. \mathscr{N}(0, 0.49)$, $T = 100$. The notes of Table 2 apply.

Table 5: Small-sample coverage rate and length of the confidence set for model M4

		$\delta^0 = 0.3$		$\delta^0 = 0.6$		$\delta^0 = 1$		$\delta^0 = 1.5$		$\delta^0 = 2$	
		Cov.	Lgth.	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	0.901	11.21	0.912	7.15
	Bai (1997)	0.805	65.94	0.821	44.07	0.850	20.71	0.878	9.88	0.887	5.96
	$\widehat{U}_T$.eq	0.950	85.23	0.951	67.40	0.951	39.87	0.950	23.58	0.955	17.46
	ILR	0.961	84.37	0.966	59.94	0.977	26.09	0.986	11.78	0.986	7.14
	sup-W	0.209		0.655		0.976		1.000		1.000	
$\lambda_0 = 0.35$	HDR	0.915	75.53	0.911	58.88	0.905	29.77	0.901	11.44	0.912	7.27
	Bai (1997)	0.821	64.69	0.826	42.93	0.849	20.77	0.880	9.92	0.888	5.99
	$\widehat{U}_T$.eq	0.948	85.48	0.948	68.95	0.948	41.40	0.948	24.01	0.954	17.57
	ILR	0.959	84.67	0.964	61.55	0.973	27.70	0.983	11.79	0.987	7.13
	sup-W	0.186		0.612		0.963		1.000		1.000	
$\lambda_0 = 0.2$	HDR	0.911	74.46	0.931	56.22	0.935	29.22	0.927	12.39	0.929	7.85
	Bai (1997)	0.820	64.06	0.870	42.86	0.896	22.11	0.898	10.40	0.887	6.16
	$\widehat{U}_T$.eq	0.952	86.80	0.956	75.20	0.952	51.99	0.956	29.92	95.20	19.92
	ILR	0.961	86.03	0.964	68.69	0.978	36.34	0.980	13.89	0.985	7.51
	sup-W	0.134		0.430		0.855		0.989		0.999	

The model is $y_t = \pi^0 + Z_t\beta^0 + Z_t\delta^0 \mathbf{1}_{\{t>\lfloor T\lambda_0 \rfloor\}} + e_t$, $X_t = 0.5X_{t-1} + u_t$, $u_t \sim i.i.d.\, \mathscr{N}(0, 0.75)$, $e_t \sim$ i.i.d.$\mathscr{N}(0, 1)$, $T = 100$. The notes of Table 2 apply.

Table 6: Small-sample coverage rate and length of the confidence set for model M5

		$\delta^0 = 0.3$		$\delta^0 = 0.6$		$\delta^0 = 1$		$\delta^0 = 1.5$		$\delta^0 = 2$	
		Cov.	Lgth.	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	0.956	47.06	0.969	35.03
	Bai (1997)	0.690	56.73	0.716	41.63	0.783	27.53	0.847	18.32	0.885	12.70
	$\widehat{U}_T$.eq	0.962	87.76	0.962	78.32	0.962	63.80	0.962	50.14	0.962	40.82
	ILR	0.790	71.07	0.805	59.66	0.824	40.78	0.868	21.59	0.909	11.63
	sup-W	0.316		0.517		0.918		0.986		0.997	
$\lambda_0 = 0.35$	HDR	0.928	76.41	0.925	68.21	56.17	0.933	43.18	0.946	31.73	0.964
	Bai (1997)	0.691	55.18	0.720	40.25	26.90	0.757	17.96	0.826	12.62	0.883
	$\widehat{U}_T$.eq	0.953	87.76	0.953	78.55	64.81	0.953	51.51	0.953	41.98	0.953
	ILR	0.795	71.34	0.804	60.48	0.832	30.42	0.870	20.77	0.903	10.78
	sup-W	0.313		0.667		0.895		0.977		0.996	
$\lambda_0 = 0.2$	HDR	0.915	75.86	0.919	66.79	0.926	52.50	0.945	38.63	0.957	27.46
	Bai (1997)	0.707	55.03	0.770	39.77	0.828	26.82	0.862	18.05	0.901	12.68
	$\widehat{U}_T$.eq	0.951	88.48	0.952	82.09	0.954	71.84	0.955	60.78	0.950	50.72
	ILR	0.795	72.01	0.809	62.75	0.829	45.18	0.870	24.86	0.913	12.62
	sup-W	0.257		0.517		0.757		0.891		0.950	

The model is $y_t = \pi^0 + Z_t\beta^0 + Z_t\delta^0 \mathbf{1}_{\{t>\lfloor T\lambda_0 \rfloor\}} + e_t$, $e_t = v_t\,|Z_t|$, $v_t \sim i.i.d.\, \mathscr{N}(0, 1)$, $Z_t = 0.5Z_{t-1} + u_t$, $u_t \sim i.i.d.\, \mathscr{N}(0, 1)$ $T = 100$. The notes of Table 2 apply.

Table 7: Small-sample coverage rate and length of the confidence set for model M6

		$\delta^0 = 0.3$		$\delta^0 = 0.6$		$\delta^0 = 1$		$\delta^0 = 1.5$		$\delta^0 = 2$	
		Cov.	Lgth.	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	0.947	25.10	0.957	12.50
	Bai (1997)	0.834	70.13	0.824	52.16	0.861	28.69	0.921	14.18	0.948	8.45
	$\widehat{U}_T$.eq	0.959	88.62	0.959	78.87	0.959	58.60	0.960	38.91	0.952	30.15
	ILR	0.969	86.75	0.959	67.91	0.967	34.13	0.985	15.97	0.995	9.17
	sup-W	0.245		0.573		0.911		0.997		1.000	
$\lambda_0 = 0.35$	HDR	0.926	74.78	0.914	64.86	0.924	45.69	0.945	23.57	0.956	12.25
	Bai (1997)	0.851	69.35	0.847	51.17	0.878	28.59	0.920	14.26	0.944	8.47
	$\widehat{U}_T$.eq	0.964	88.82	0.960	79.74	0.964	60.26	0.964	39.89	0.964	30.64
	ILR	0.972	88.69	0.975	73.95	0.981	39.08	0.986	16.06	0.992	9.08
	sup-W	0.244		0.559		0.904		0.994		1.000	
$\lambda_0 = 0.2$	HDR	0.909	78.12	0.921	61.87	0.933	40.66	0.948	20.95	0.961	11.70
	Bai (1997)	0.824	65.23	0.867	51.35	0.915	29.83	0.937	14.92	0.955	8.70
	$\widehat{U}_T$.eq	0.961	89.71	0.960	83.68	0.961	69.25	0.960	49.11	0.960	35.78
	ILR	0.966	91.48	0.971	82.78	0.984	51.93	0.988	21.36	0.995	10.87
	sup-W	0.232		0.467		0.804		0.962		0.995	

The model is $y_t = \beta^0 + \delta^0 \mathbf{1}_{\{t > \lfloor T\lambda_0 \rfloor\}} + e_t$, $e_t = 0.3 e_{t-1} + u_t$, $u_t \sim i.i.d.\, \mathrm{t}_\nu$, $\nu = 5$, $T = 100$. The notes of Table 2 apply.

Table 8: Small-sample coverage rate and length of the confidence set for model M7

		$\delta^0 = 0.3$		$\delta^0 = 0.6$		$\delta^0 = 1$		$\delta^0 = 1.5$		$\delta^0 = 2$	
		Cov.	Lgth.	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	0.946	12.60	0.965	6.34
	Bai (1997)	0.778	60.94	0.815	38.14	0.885	17.29	0.928	8.53	0.949	5.34
	$\widehat{U}_T$.eq	0.949	84.56	0.950	67.64	0.953	42.95	0.950	29.95	0.950	30.25
	ILR	0.943	83.69	0.946	63.24	0.956	32.85	0.967	16.20	0.982	10.49
	sup-W	0.275		0.753		0.991		1.000		1.000	
$\lambda_0 = 0.35$	HDR	0.919	74.16	0.916	58.53	0.931	32.10	0.948	12.95	0.965	6.48
	Bai (1997)	0.799	60.25	0.814	37.94	0.872	17.49	0.919	8.59	0.952	5.35
	$\widehat{U}_T$.eq	0.951	85.01	0.948	69.14	0.957	48.40	0.953	31.07	0.949	30.31
	ILR	0.946	84.12	0.944	63.99	0.960	33.45	0.973	14.91	0.977	8.71
	sup-W	0.258		0.700		0.986		1.000		1.000	
$\lambda_0 = 0.2$	HDR	0.912	73.43	0.929	56.18	0.949	31.23	0.956	13.65	0.965	6.96
	Bai (1997)	0.795	59.43	0.864	38.17	0.910	18.52	0.934	8.67	0.954	5.34
	$\widehat{U}_T$.eq	0.950	86.94	0.951	76.52	0.946	55.72	0.955	39.59	0.947	38.80
	ILR	0.945	83.94	0.953	63.55	0.963	32.41	0.973	24.42	0.982	15.01
	sup-W	0.195		0.546		0.920		0.998		1.000	

The model is $y_t = \delta^0 \left(1 - \pi^0\right) \mathbf{1}_{\{t > \lfloor T\lambda_0 \rfloor\}} + \pi^0 y_{t-1} + e_t$, $e_t \sim i.i.d.\, \mathscr{N}\,(0,\, 0.49)$, $\pi^0 = 0.3$, $T = 100$. The notes of Table 2 apply.

Table 9: Small-sample coverage rate and length of the confidence sets for model M8

		$\delta^0 = 1$		$\delta^0 = 1.5$		$\delta^0 = 2$		$\delta^0 = 2.5$		$\delta^0 = 3$	
		Cov.	Lgth.	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	0.986	5.99	0.995	4.55
	Bai (1997)	0.793	12.87	0.877	7.11	0.929	4.78	0.951	3.66	0.973	2.957
	$\widehat{U}_T$.eq	0.951	91.64	0.955	93.94	0.959	93.71	0.960	91.63	0.961	90.34
	ILR	0.951	46.31	0.967	34.19	0.977	26.48	0.990	16.15	0.991	16.49
	sup-W	0.996		1.000		1.000		1.000		1.000	
$\lambda_0 = 0.35$	HDR	0.925	33.02	0.933	16.67	0.971	9.40	0.986	4.39	0.994	4.33
	Bai (1997)	0.804	13.00	0.876	7.11	0.923	4.94	0.955	3.65	0.974	2.93
	$\widehat{U}_T$.eq	0.952	91.22	0.945	92.61	0.957	92.48	0.961	93.58	0.964	93.08
	ILR	0.949	47.54	0.967	34.18	0.982	25.84	0.987	5.98	0.984	16.76
	sup-W	0.992		1.000		1.000		1.000		1.000	
$\lambda_0 = 0.2$	HDR	0.937	34.66	0.953	19.24	0.954	11.42	0.984	7.36	0.994	5.36
	Bai (1997)	0.832	13.64	0.885	7.19	0.931	4.92	0.950	3.61	0.971	2.91
	$\widehat{U}_T$.eq	0.944	89.64	0.951	89.58	0.956	88.22	0.958	86.98	0.961	85.95
	ILR	0.946	49.13	0.970	33.54	0.980	24.48	0.984	16.82	0.989	12.51
	sup-W	0.935		0.995		1.000		1.000		1.000	

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

Table 10: Small-sample coverage rate and length of the confidence sets for model M9

		$\delta^0 = 0.3$		$\delta^0 = 0.6$		$\delta^0 = 1$		$\delta^0 = 1.5$		$\delta^0 = 2$	
		Cov.	Lgth.	Cov.	Lgth.	Cov.	Lgth.	Cov.	Lgth.	Cov.	Lgth.
$\lambda_0 = 0.5$	HDR	0.903	61.09	0.927	31.14	0930	18.33	0.940	13.32	0.930	9.10
	Bai (1997)	0.791	37.86	0.831	17.73	0.855	10.43	0.898	8.12	0.868	5.30
	$\widehat{U}_T$.eq	0.947	65.23	0.947	39.76	0.947	28.82	0.934	27.90	0.947	20.36
	ILR	0.909	72.62	0.946	45.06	0.962	23.97	0.973	13.60	0.978	9.34
	sup-W	0.746		0.941		0.976		0.960		0.990	
$\lambda_0 = 0.35$	HDR	0.904	60.58	0.918	30.96	0.904	18.16	0.923	13.31	0.928	0.34
	Bai (1997)	0.791	37.70	0.829	18.04	0.852	10.61	0.864	7.81	0.870	5.34
	$\widehat{U}_T$.eq	0.942	66.27	0.942	40.63	0.942	29.39	0.941	24.04	0.942	20.67
	ILR	0.922	72.20	0.947	45.27	0.959	24.93	0.970	12.96	0.973	8.55
	sup-W	0.734		0.931		0.971		0.972		0.988	
$\lambda_0 = 0.2$	HDR	0.920	61.37	0.946	31.00	0.942	20.44	0.941	13.38	0.944	9.04
	Bai (1997)	0.791	39.23	0.841	19.28	0.876	11.99	0.898	8.12	0.886	6.16
	$\widehat{U}_T$.eq	0.934	71.42	0.931	47.53	0.934	34.12	0.934	27.90	0934	24.06
	ILR	0.920	72.68	0.935	49.61	0.959	27.90	0.969	15.75	0.972	10.01
	sup-W	0.634		0.884		0.944		0.960		0.976	

The model is $y_t = \pi^0 + Z_t\beta^0 + Z_t\delta^0\mathbf{1}_{\{t > \lfloor T\lambda_0 \rfloor\}} + e_t$, $Z_t \sim i.i.d. \mathscr{N}(1, 1.44)$, $\{e_t\}$ follows a FIGARCH(1,0.6,1) process and $T = 100$. The notes of Table 2 apply.

Table 11: Small-sample coverage rate and length of the confidence set for model M10

		$\delta^0 = 0.3$		$\delta^0 = 0.6$		$\delta^0 = 1$		$\delta^0 = 1.5$		$\delta^0 = 2$	
		Cov.	Lgth.	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	6.53	0.916	4.34
	Bai (1997)	0.809	45.33	0.844	17.11	0.864	8.27	0.878	5.08	0.883	3.61
	$\widehat{U}_T$.eq	0.959	72.69	0.959	39.81	0.959	24.25	0.959	17.96	0.959	14.79
	ILR	0.929	83.23	0.951	69.67	0.971	44.40	0.978	20.76	0.987	10.44
	sup-W	0.600		0.988		1.000		1.000		1.000	
$\lambda_0 = 0.35$	HDR	0.934	73.08	0.937	35.37	0.923	13.68	0.920	6.82	0.920	4.55
	Bai (1997)	0.821	45.70	0.838	17.78	0.867	8.53	0.886	5.22	0.889	3.71
	$\widehat{U}_T$.eq	0.964	76.14	0.964	44.61	0.965	27.33	0.965	19.74	0.964	15.84
	ILR	0.934	81.32	0.959	62.98	0.977	34.38	0.982	16.73	0.984	9.12
	sup-W	0.529		0.970		0.999		1.000		1.000	
$\lambda_0 = 0.2$	HDR	0.941	71.46	0.959	59.03	0.950	15.39	0.926	7.78	0.919	5.03
	Bai (1997)	0.818	47.82	0.872	20.44	0.878	9.60	0.876	5.64	0.873	3.92
	$\widehat{U}_T$.eq	0.971	82.40	0.971	59.03	0.971	39.02	0.971	27.07	0.972	20.42
	ILR	0.928	83.26	0.952	70.03	0.964	42.65	0.979	20.15	0.982	10.30
	sup-W	0.346		0.839		0.981		0.997		0.999	

The model is $y_t = \pi^0 + Z_t\beta^0 + Z_t\delta^0\mathbf{1}_{\{t>\lfloor T\lambda_0\rfloor\}} + e_t$, $e_t \sim i.i.d.\,\mathscr{N}(0, 1)$, $Z_t \sim \text{ARFIMA}(0.3, 0.6, 0)$, $T = 100$. The notes of Table 2 apply.

Supplemental Material to

Continuous Record Asymptotics for Structural Change Models

Alessandro Casini Pierre Perron

Department of Economics
Boston University

30th November 2017
First Vesrion: 28th October 2015

Abstract

This supplemental material is structured as follows. Section S.A contains the Mathematical Appendix which includes proofs of most of the results in the paper. In Section S.B we supplement our discussion on the probability density of the continuous record asymptotic distribution with additional results.

S.A Mathematical Proofs

S.A.1 Additional Notations

For a matrix A, the orthogonal projection matrices P_A, M_A are defined as $P_A = A(A'A)^{-1}A'$ 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 $(A'A)_{i,j}$ and the $i\times j$ upper-left (resp., lower-right) sub-block of $A'A$ as $[A'A]_{\{i\times j,\cdot\}}$ (resp., $[A'A]_{\{\cdot,i\times j\}}$). For a random variable ξ and a number $r\geq 1$, we write $\|\xi\|_r = (\mathbb{E}\|\xi\|^r)^{1/r}$. 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 further use the same notations as explained in Section 2.

S.A.2 Preliminary Lemmas

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

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

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

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

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.A.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 $\overline{\Sigma}_{kh}\triangleq(T^\varpi h)^{-1}\int_{kh}^{kh+T^\varpi h}\Sigma_s ds$, and let $\widehat{\Sigma}_{kh}\triangleq(T^\varpi h)^{-1}\sum_{i=1}^{\lfloor T^\varpi\rfloor}\Delta_h X_{k+i}\Delta_h\widetilde{X}_{k+i}'$. Then, $\sup_{1\leq k\leq B_T}\left\|\widehat{\Sigma}_{kh}-\overline{\Sigma}_{kh}\right\|=o_p(1)$. Furthermore, for each k and some $K>0$ with $N-K>kh>K$, $\sup_{T^\epsilon\leq T^\varpi\leq T^{1-\epsilon}}\left\|\widehat{\Sigma}_{kh}-\overline{\Sigma}_{kh}\right\|=o_p(1)$.*

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

$$\left\| \sup_{T^{\epsilon} \leq T^{\varpi} \leq T^{1-\epsilon}} \left| \widehat{\Sigma}_{kh} - \overline{\Sigma}_{kh} \right| \right\|_{l} = \left\| \sup_{1 \leq T^{\varpi - \epsilon} \leq T^{1-2\epsilon}} \left| \widehat{\Sigma}_{kh} - \overline{\Sigma}_{kh} \right| \right\|_{l} \leq K_{l} T^{(1-2\epsilon)/l} T^{-\epsilon/2}$$

Choose $l > (2 - 4\epsilon)/\epsilon$ to verify the claim. $\square$

S.A.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 2.1-2.2 with the following stronger assumption.

Assumption S.A.1. *Let Assumption 2.1-2.2 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 introduce the following notation which will be useful in some of the proofs below.

S.A.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 $X = (X^i)_{1 \leq i \leq d} \in \mathbb{R}^d$ 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[5]

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

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 $h \downarrow 0$:

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

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.A.3-S.A.4 follow from their Theorem 3.3.1-(b), while Lemma S.A.5 follows from their Theorem 5.4.2.

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

Lemma S.A.3. *Under Assumption S.A.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* $(Z_2'e)_{i,1} = \sum_{k=T_b+1}^{T} z_{kh}^{(i)} e_{kh}$;

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

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

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

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

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

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

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

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

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

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

(i) with $(Z_\Delta' Z_\Delta)_{i,j} = \sum_{k=T_b^0+1}^{T_b} z_{kh}^{(i)} z_{kh}^{(j)}$ *we have* $\begin{cases} |(Z_\Delta' Z_\Delta)_{i,j} - \int_{(T_b+1)h}^{T_b^0 h} \Sigma_{ZZ,s}^{(i,j)} ds| \xrightarrow{P} 0, & \text{if } T_b < T_b^0 \\ |(Z_\Delta' Z_\Delta)_{i,j} - \int_{T_b^0 h}^{(T_b+1)h} \Sigma_{ZZ,s}^{(i,j)} ds| \xrightarrow{P} 0, & \text{if } T_b > T_b^0 \end{cases}$;

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

(ii) with $(Z_\Delta' X_\Delta)_{i,r} = \sum_{k=T_b^0+1}^{T_b} z_{kh}^{(i)} x_{kh}^{(r)}$ *we have* $\begin{cases} |(Z_\Delta' X_\Delta)_{i,r} - \int_{(T_b+1)h}^{T_b^0 h} \Sigma_{ZX,s}^{(i,r)} ds| \xrightarrow{P} 0, & \text{if } T_b < T_b^0 \\ |(Z_\Delta' X_\Delta)_{i,r} - \int_{T_b^0 h}^{(T_b+1)h} \Sigma_{ZX,s}^{(i,r)} ds| \xrightarrow{P} 0, & \text{if } T_b > T_b^0 \end{cases}$.

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 $\left(\Omega,\ \mathscr{F},\ \{\mathscr{F}_t\}_{t\geq 0},\ P\right)$. 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 $\left(\widetilde{\Omega},\ \widetilde{\mathscr{F}},\ \widetilde{P}\right)$ of the original space $\left(\Omega,\ \mathscr{F},\ \{\mathscr{F}_t\}_{t\geq 0},\ P\right)$. Any variable or process defined on either Ω or Ω^* is extended in the *usual way* 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 $\left(\widetilde{\Omega},\ \widetilde{\mathscr{F}},\ \{\mathscr{F}_t\}_{t\geq 0},\ \widetilde{P}\right)$. Finally, as for the filtration, we can consider another filtration $\left\{\widetilde{\mathscr{F}_t}\right\}_{t\geq 0}$ 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 $\left(\widetilde{\Omega},\ \widetilde{\mathscr{F}},\ \left\{\widetilde{\mathscr{F}_t}\right\}_{t\geq 0},\ \widetilde{P}\right)$ of the original filtered space $\left(\Omega,\ \mathscr{F},\ \{\mathscr{F}_t\}_{t\geq 0},\ P\right)$ is constructed in this paper. Assume that the auxiliary probability space $\left(\Omega^*,\ \mathscr{F}^*,\ \{\mathscr{F}_t^*\}_{t\geq 0},\ P^*\right)$ supports a p^2-dimensional standard Wiener process $W_s^\dagger$ which is adapted to $\left\{\widetilde{\mathscr{F}_t}\right\}$. 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 $\widehat{\Sigma}_{Z,s}^{(ij,kl)} = \Sigma_{Z,s}^{(ik)} \Sigma_{Z,s}^{(jl)}$. Due to the symmetry of $\Sigma_{Z,s}$, the matrix with entries $\left(\sigma_{Z,s}^{*,(ij,kl)} + \sigma_{Z,s}^{*,(ji,kl)}\right)/\sqrt{2}$ is a square-root of the matrix with entries $\widehat{\Sigma}_{Z,s}^{(ij,kl)} + \widehat{\Sigma}_{Z,s}^{(il,jk)}$. Then the process $\mathscr{U}_t$ with components $\mathscr{U}_t^{(r,j)} = 2^{-1/2} \sum_{k,l=1}^{p} \int_0^t \left(\sigma_{Z,s}^{(rj,kl)} + \sigma_{Z,s}^{(jr,kl)}\right) dW_s^{\dagger(kl)}$

is, conditionally on $\mathscr{F}$, a continuous Gaussian process with independent increments and (conditional) covariance $\widetilde{\mathbb{E}}\left(\mathscr{U}^{(r,j)}(v)\,\mathscr{U}^{(k,l)}(v)\,|\,\mathscr{F}\right)=\int_{T_b^0h+v}^{T_b^0h}\left(\Sigma_{Z,s}^{(rk)}\Sigma_{Z,s}^{(jl)}+\Sigma_{Z,s}^{(rl)}\Sigma_{Z,s}^{(jk)}\right)ds$, where $v\leq 0$. The CLT of interest is as follows.

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

S.A.4 Proofs of Sections 3 and 4

S.A.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 when we shall establish a $T^{1-\kappa}$-rate of convergence of $\widehat{\lambda}_b$ toward λ_0, we will consider the case where $N_b-N_b^0=T^{-(1-\kappa)}$ for some $\kappa\in(0,\,1/2)$. Hence, it is convenient to 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 $\widehat{\Sigma}_X\left(T_b,\,T_b^0\right)$ 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 reason conditionally on the processes $\mu_{\cdot,t}$ and Σ_t^0 (defined in Assumption 2.2) 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.A.4.2 Proof of Proposition 3.1

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

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

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

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

where

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

Denote

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

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

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

We arbitrarily define $r\left(T_b\right) = \left(\delta^0\right)'\delta^0$ 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 it must satisfy $g_e\left(\widehat{T}_b\right) \geq \left|\widehat{T}_b - T_b^0\right| r\left(\widehat{T}_b\right)$. Therefore,

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

where recall $p \leq T_b \leq T - p$ 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.A.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.,

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

The first term of $g_e(T_b)$ is

$$2\left(\delta^0\right)'(Z_0'MZ_2)(Z_2'MZ_2)^{-1/2}(Z_2'MZ_2)^{-1/2}Z_2Me. \tag{S.16}$$

Lemma S.A.5 implies that for any $1 \leq j \leq p$, $(Z_2e)_{j,1}/\sqrt{h} = O_p(1)$ and for any $1 \leq i \leq q+p$, $(Xe)_{i,1}/\sqrt{h} = O_p(1)$. These hold because they both involve a positive fraction of the data. Furthermore, from Lemma S.A.3, we also have that $Z_2'MZ_2$ and $Z_0'MZ_2$ are $O_p(1)$. Therefore, the supremum of $(Z_0'MZ_2)(Z_2'MZ_2)^{-1/2}$ over all T_b is $\sup_{T_b}(Z_0'MZ_2)(Z_2'MZ_2)^{-1}(Z_2'MZ_0) \leq Z_0'MZ_0 = O_p(1)$ by Lemma S.A.3. By Assumption (2.1)-(iii) $(Z_2'MZ_2)^{-1/2}Z_2'Me$ is $O_p(1)O_p\left(\sqrt{h}\right)$ uniformly, which implies that (S.16) is $O_p\left(\sqrt{h}\right)$ uniformly over $p \leq T_b \leq T-p$. As for the second term of (S.10), $Z_0'Me = O_p\left(\sqrt{h}\right)$. The first term in (S.11) is uniformly $o_p(1)$ and the same holds for the last term. Therefore, combining these results, $\sup_{T_b}|g_e(T_b)| = O_p\left(\sqrt{h}\right)$ uniformly when $\left|\widehat{\lambda}_b - \lambda_0\right| > K$. Therefore for some $B > 0$, these arguments combined with Lemma S.A.6 below result in $P\left(r_B^{-1}\sup_{p\leq T_b\leq T-p}|g_e(T_b)| \geq K\right) \leq \varepsilon$, 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 arbitrarily chosen. □

Lemma S.A.6. *For $B > 0$, let $r_B = \inf_{|T_b - T_b^0| > TB} Tr(T_b)$. There exists a $\kappa > 0$ such that for every $\varepsilon > 0$, there exists a $B < \infty$ such that $P(r_B \geq \kappa) \leq 1-\varepsilon$, 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.A.1, $r(T_b) \geq (\delta^0)' R'(X_\Delta' X_\Delta/(T_b^0 - T_b))(X_2'X_2)^{-1}(X_0'X_0)R\delta^0$. When multiplied by T, we have

$$Tr(T_b) \geq T\left(\delta^0\right)' R' \frac{X_\Delta' X_\Delta}{T_b^0 - T_b}(X_2'X_2)^{-1}(X_0'X_0)R\delta^0 = \left(\delta^0\right)' R' \frac{X_\Delta' X_\Delta}{N_b^0 - N_b}(X_2'X_2)^{-1}(X_0'X_0)R\delta^0.$$

Note that $0 < K < B < h(T_b^0 - T_b) < N$. Then, $Tr(T_b) \geq (\delta^0)' R'(X_\Delta' X_\Delta/N)(X_2'X_2)^{-1}(X_0'X_0)R\delta^0$, and by standard estimates for Itô semimartingales, $X_\Delta' X_\Delta = O_p(1)$ (i.e., use the Burkhölder-Davis-Gundy inequality and recall that $\left|\widehat{N}_b - N_b^0\right| > BN$). Hence, we conclude that $Tr(T_b) \geq (\delta^0)' R' O_p(1/N) O_p(1) R\delta^0 \geq \kappa > 0$, where κ is some positive constant. The last inequality follows whenever $X_\Delta' X_\Delta$ is positive definite since $R'X_\Delta' X_\Delta(X_2'X_2)^{-1}(X_0'X_0)R$ can be rewritten as $R'\left[(X_0'X_0)^{-1} + (X_\Delta' X_\Delta)^{-1}\right]R$. According to Lemma S.A.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 $r_B = \inf_{|T_b - T_b^0| > TB} Tr(T_b)$ satisfies $P(r_B \geq \kappa) \leq 1 - \varepsilon$. The case with $T_b > T_b^0$ is similar and is omitted. □

S.A.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 $\mathbf{B}_T \triangleq \{T_b : T\eta \leq T_b \leq T(1-\eta)\}$, 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 $T \to \infty$. 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 $\mathbf{B}_{K,T} \triangleq \{T_b : |N_b - N_b^0| \leq KT^{-1}\}$, for some $K > 0$. For any $K > 0$, define the intersection of $\mathbf{B}_T$

and the complement of $\mathbf{B}_{K,T}$ by $\mathbf{D}_{K,T} \triangleq \left\{T_b : N\eta \leq N_b \leq N(1-\eta), |N_b - N_b^0| > KT^{-1}\right\}$. Notice that

$$\begin{aligned}\left\{\left|\widehat{\lambda}_b - \lambda_0\right| > KT^{-1}\right\} &= \left\{\left|\widehat{\lambda}_b - \lambda_0\right| > KT^{-1} \cap \widehat{\lambda}_b \in (\eta,\, 1-\eta)\right\} \cup \left\{\left|\widehat{\lambda}_b - \lambda_0\right| > KT^{-1} \cap \widehat{\lambda}_b \notin (\eta,\, 1-\eta)\right\} \\ &\subseteq \left\{\left|\widehat{\lambda}_b - \lambda_0\right| > K\left(T^{-1}\right) \cap \widehat{\lambda}_b \in (\eta,\, 1-\eta)\right\} \cup \left\{\widehat{\lambda}_b \notin (\eta,\, 1-\eta)\right\},\end{aligned}$$

and so

$$P\left(\left|\widehat{\lambda}_b - \lambda_0\right| > KT^{-1}\right) \leq P\left(\widehat{\lambda}_b \notin (\eta,\, 1-\eta)\right) + P\left(\left|\widehat{T}_b - T_b^0\right| > K \cap \widehat{\lambda}_b \in (\eta,\, 1-\eta)\right),$$

and for large T,

$$\begin{aligned}P\left(\left|\widehat{\lambda}_b - \lambda_0\right| > KT^{-1}\right) &\leq \varepsilon + P\left(\left|\widehat{\lambda}_b - \lambda_0\right| > KT^{-1} \cap \widehat{\lambda}_b \in (\eta,\, 1-\eta)\right) \\ &\leq \varepsilon + P\left(\sup_{T_b \in \mathbf{D}_{K,T}} Q_T(T_b) \geq Q_T\left(T_b^0\right)\right).\end{aligned}$$

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 $P\left(\sup_{T_b \in \mathbf{D}_{K,T}} Q_T(T_b) \geq Q_T(T_b^0)\right) < \varepsilon$. 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 $g_e(T_b) / |T_b - T_b^0| \geq r(T_b)$, where $g_e(T_b)$ and $r(T_b)$ were 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.A.1,

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

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 becomes 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 $\mathbb{E}\left[x_{kh}^{(i)} x_{kh}^{(j)} | \mathscr{F}_{(k-1)h}\right] = \Sigma_{X,(k-1)h}^{(i,j)} h$, 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 $\left(h\left|T_b - T_b^0\right|\right)^{-1} X_\Delta' X_\Delta = \left(h\left|T_b - T_b^0\right|\right)^{-1} \sum_{k=T_b+1}^{T_b^0} x_{kh} x_{kh}' = A > 0$ 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.11). When $T_b \in \mathbf{D}_{K,T}$, Z_2 involves at least a positive fraction $N\eta$ of the data. From Lemma S.A.3, as $h \downarrow 0$, it follows that

$$h^{-1}\left(T_b^0 - T_b\right)^{-1} e'MZ_2\left(Z_2'MZ_2\right)^{-1} Z_2Me = \left(T_b^0 - T_b\right)^{-1} h^{-1} O_p\left(h^{1/2}\right) O_p(1)\, O_p\left(h^{1/2}\right) = \frac{O_p(1)}{T_b^0 - T_b},$$

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 $g_e(T_b)$. Using $Z_2 = Z_0 \pm Z_\Delta$ we can deduce that

$$\begin{aligned}\left(\delta^0\right)'\left(Z_0'MZ_2\right)\left(Z_2'MZ_2\right)^{-1} Z_2Me &= \left(\delta^0\right)'\left(\left(Z_2' \pm Z_\Delta'\right) MZ_2\right)\left(Z_2'MZ_2\right)^{-1} Z_2Me \\ &= \left(\delta^0\right)' Z_0'Me \pm \left(\delta^0\right)' Z_\Delta'Me \pm \left(\delta^0\right)'\left(Z_\Delta'MZ_2\right)\left(Z_2'MZ_2\right)^{-1} Z_2Me,\end{aligned} \tag{S.18}$$

from which it follows that

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

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

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

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.A.5 that the second term of (S.19) is $O_p\left(h^{1/2}\right)$. Next, note that $Z_\Delta'Me = Z_\Delta'e - Z_\Delta'X\left(X'X\right)^{-1}X'e$. We can write

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

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

$$P\left(\sup_{T_b\in\mathbf{D}_{K,T}}\left\|\left(\delta^0\right)'\frac{2}{h}\frac{1}{T_b^0-T_b}\sum_{k=T_b+1}^{T_b^0}z_{kh}e_{kh}\right\|>\frac{B}{4N}\right) \tag{S.22}$$

$$\leq P\left(\sup_{T_b\leq T_b^0-KN^{-1}}\left\|\frac{1}{h}\frac{1}{T_b^0-T_b}\sum_{k=T_b+1}^{T_b^0}z_{kh}e_{kh}\right\|>\frac{B}{8N\left\|\delta^0\right\|}\right)<\varepsilon. \tag{S.23}$$

Note that $\left|T_b-T_b^0\right|$ is bounded away from zero in $\mathbf{D}_{K,T}$. Observe that $\left(z_{kh}/\sqrt{h}\right)\left(e_{kh}/\sqrt{h}\right)$ 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)],

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

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.A.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 $\left|\widehat{\lambda}_b-\lambda_0\right|$ can be made arbitrary small. Proposition 3.2 gives the associated rate of convergence: $T\left(\widehat{\lambda}_b-\lambda_0\right)=O_p(1)$. 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 $\left|T_b-T_b^0\right|\leq K$, 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, δ_1^0 and δ_2^0. Toward this end, we first find a lower bound on the convergence rate for $\widehat{\lambda}_b$ that guarantees its estimation problem 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.A.7. *Fix $\gamma\in(0,1/2)$ and some constant $A > 0$. For all large $T>\overline{T}$, if $\left|\widehat{N}_b-N_b^0\right|\leq AO_p\left(h^{1-\gamma}\right)$, then $X'\left(Z_0-\widehat{Z}_0\right)=O_p\left(h^{1-\gamma}\right)$ and $Z_0'\left(Z_0-\widehat{Z}_0\right)=O_p\left(h^{1-\gamma}\right)$.*

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 $h\downarrow 0$. By assumption, there exists some constant $C>0$ such that $P\left(h^{\gamma}\left|\widehat{T}_b-T_b^0\right|>C\right)<\varepsilon$. We have to show that although we only know $\left|\widehat{T}_b-T_b^0\right|\leq Ch^{-\gamma}$, 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 $\left|\widehat{N}_b-N_b^0\right|\to 0$ at rate at least $h^{1-\gamma}$ 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 $h\downarrow 0$, 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 $X'\left(Z_0-\widehat{Z}_0\right)/h^{1-\gamma}$ 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 $h \downarrow 0$,

$$X'_\Delta Z_\Delta / h^{1-\gamma} \xrightarrow{P} \Sigma_{XX,N_b^0}, \tag{S.24}$$

which implies that $h^{-1/2} X' \left(Z_0 - \widehat{Z}_0 \right) = O_p \left(h^{1/2-\gamma} \right)$. This shows that the order of the error in replacing Z_0 by $Z_2 = \widehat{Z}_0$ goes to zero at a enough fast rate. That is, by definition we can write $Y = X\beta^0 + \widehat{Z}_0 \delta^0 + \left(Z_0 - \widehat{Z}_0 \right) \delta^0 + e$, from which it follows that $X'\widehat{Z}_0 = X'Z_0 + o_p(1)$, $X' \left(Z_0 - \widehat{Z}_0 \right) \delta^0 = o_p(1)$ and $Z'_0 \left(Z_0 - \widehat{Z}_0 \right) \delta^0 = o_p(1)$. 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 proven similarly. This concludes the proof of the Lemma. □

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

Proof of Proposition 3.3. By standard arguments,

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

from which it follows that

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

and a similar reasoning applies to $\widehat{Z}'_0 \left(Z_0 - \widehat{Z}_0 \right) \delta^0$. 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.A.3-S.A.4. The rest of the arguments (including mixed normality) follows from Barndorff-Nielsen and Shephard (2004) and are omitted. □

S.A.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 $\widehat{T}_b = \arg\max_{T_b} Q_T(T_b)$, $Q_T\left(\widehat{T}_b\right) = \widehat{\delta}'_{T_b} \left(Z'_2 M Z_2 \right) \widehat{\delta}_{T_b}$, and the decomposition

$$Q_T(T_b) - Q_T\left(T_b^0\right) = \widehat{\delta}'_{T_b} \left(Z'_2 M Z_2 \right) \widehat{\delta}_{T_b} - \widehat{\delta}'_{T_b^0} \left(Z'_0 M Z_0 \right) \widehat{\delta}_{T_b^0} \tag{S.25}$$

$$= \delta'_h \left\{ \left(Z'_0 M Z_2 \right) \left(Z'_2 M Z_2 \right)^{-1} \left(Z'_2 M Z_0 \right) - Z'_0 M Z_0 \right\} \delta_h \tag{S.26}$$

$$+ g_e(T_b), \tag{S.27}$$

where

$$g_e(T_b) = 2\delta'_h \left(Z'_0 M Z_2 \right) \left(Z'_2 M Z_2 \right)^{-1} Z_2 M e - 2\delta'_h \left(Z'_0 M e \right) \tag{S.28}$$

$$+ e' M Z_2 \left(Z'_2 M Z_2 \right)^{-1} Z_2 M e - e' M Z_0 \left(Z'_0 M Z_0 \right)^{-1} Z'_0 M e. \tag{S.29}$$

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

$$
\begin{aligned}
P\left(\left|\widehat{\lambda}_b - \lambda_0\right| > K\right) &= P\left(\left|\widehat{T}_b - T_b^0\right| > TK\right) \\
&\leq P\left(\sup_{\left|T_b - T_b^0\right| > TK} h^{-1/2}\left|g_e\left(T_b\right)\right| \geq \inf_{\left|T_b - T_b^0\right| > TK} h^{-1/2}\left|T_b - T_b^0\right| r\left(T_b\right)\right) \\
&\leq P\left(\sup_{p \leq T_b \leq T-p} h^{-1/2}\left|g_e\left(T_b\right)\right| \geq TK \inf_{\left|T_b - T_b^0\right| > TK} h^{-1/2} r\left(T_b\right)\right) \\
&= P\left(r_T^{-1} \sup_{p \leq T_b \leq T-p} h^{-1/2}\left|g_e\left(T_b\right)\right| \geq K\right),
\end{aligned} \tag{S.30}
$$

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.A.8. Thus, it is sufficient to verify that

$$
\sup_{p \leq T_b \leq T-p} h^{-1/2}\left|g_e\left(T_b\right)\right| = o_p(1). \tag{S.31}
$$

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

$$
\begin{aligned}
&2\delta_h'\left(Z_0'MZ_2\right)\left(Z_2'MZ_2\right)^{-1/2}\left(Z_2'MZ_2\right)^{-1/2} Z_2Me \\
&\qquad \leq 2h^{1/4}\left(\delta^0\right)'\left(Z_0'MZ_2\right)\left(Z_2'MZ_2\right)^{-1/2}\left(Z_2'MZ_2\right)^{-1/2} Z_2Me.
\end{aligned} \tag{S.32}
$$

For any $1 \leq j \leq p$, $\left(Z_2\widetilde{e}\right)_{j,1}/\sqrt{h} = O_p(1)$ by Theorem S.A.5, and similarly, for any $1 \leq i \leq q+p$, $\left(X\widetilde{e}\right)_i/\sqrt{h} = O_p(1)$. Furthermore, from Lemma S.A.3 we also have that $Z_2'MZ_2$ and $Z_0'MZ_2$ are $O_p(1)$. Therefore, the supremum of $\left(Z_0'MZ_2\right)\left(Z_2'MZ_2\right)^{-1/2}$ over all T_b is such that

$$
\sup_{T_b}\left(Z_0'MZ_2\right)\left(Z_2'MZ_2\right)^{-1}\left(Z_2'MZ_0\right) \leq Z_0'MZ_0 = O_p(1),
$$

by Lemma S.A.3. By Assumption 2.1-(iii) $\left(Z_2'MZ_2\right)^{-1/2} Z_2M\widetilde{e}$ is $O_p(1)\, O_p\left(\sqrt{h}\right)$ uniformly, which implies that (S.32) is $O_p\left(\sqrt{h}\right)$ uniformly over $p \leq T_b \leq T-p$. In view of Assumption 4.1 [recall (4.1)], it is crucial to study the behavior of $\left(X'e\right)_{j,1}$ for $1 \leq j \leq p+q$. Note first that $\left|\widehat{\lambda}_b - \lambda_0\right| > K$ or $N > \left|\widehat{N}_b - N_b^0\right| > KN$. Then, by Itô formula proceeding as in the proof of Lemma S.A.2, we have a standard result for the local volatility of a continuous Itô semimartingale; namely that for some $A > 0$ (recall the condition $T^{1-\kappa}\epsilon \to B > 0$),

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

From Assumption 2.1-(iv) since $\Sigma_{Xe,t} = 0$ for all $t \geq 0$, we have

$$
\begin{aligned}
X'e &= \sum_{k=1}^{T_b^0 - \lfloor T^\kappa \rfloor} x_{kh}\widetilde{e}_{kh} + h^{-1/4}\sum_{k=T_b^0 - \lfloor T^\kappa \rfloor + 1}^{T_b^0 + \lfloor T^\kappa \rfloor} x_{kh}\widetilde{e}_{kh} + \sum_{k=T_b^0 + \lfloor T^\kappa \rfloor + 1}^{T} x_{kh}\widetilde{e}_{kh} \\
&= O_p\left(h^{1/2}\right) + h^{-1/4} O_p\left(h^{1-\kappa+1/2}\right) + O_p\left(h^{1/2}\right) = O_p\left(h^{1/2}\right).
\end{aligned} \tag{S.33}
$$

The same bound applies to $Z_2'e$ and $Z_0'e$. Thus, equation (S.32) is such that

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

As for the second term of (S.28),

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

using (S.33). Again using (S.33), the first term in (S.29) is, uniforly in T_b,

$$h^{-1/2}e'MZ_2\left(Z_2'MZ_2\right)^{-1}Z_2Me=h^{-1/2}BO_p\left(h^{1/2}\right)O_p\left(1\right)O_p\left(h^{1/2}\right)=O_p\left(h^{1/2}\right). \tag{S.34}$$

Similarly, the last term in (S.29) is $O_p\left(h^{1/2}\right)$. Therefore, combining these results we have $h^{-1/2}\sup_{T_b}\left|g_e\left(T_b\right)\right|$ $=BO_p\left(h^{1/4}\right)$, from which it follows that the right-hand side of (S.30) is weakly smaller than ε.

Lemma S.A.8. *For $B>0$, let $r_{B,h}=\inf_{\left|T_b-T_b^0\right|>TB}Th^{-1/2}r\left(T_b\right)$. 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.A.1,

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

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

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

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

By Lemma S.A.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 $h \downarrow 0$

$$\frac{1}{h\left(T_b^0 - T_b\right)} \sum_{k=T_b+1}^{T_b^0} x_{kh} x'_{kh} \xrightarrow{P} \Sigma_{XX,N_b^0}, \tag{S.35}$$

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.A.2 this approximation is uniform, establishing that

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

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

$$P\left(\sup_{T_b \in \mathbf{D}_{K,T}} h^{-3/2} \frac{g_e\left(T_b, \delta_h\right)}{|T_b - T_b^0|} \geq B\right) < \varepsilon, \tag{S.36}$$

for some $B > 0$. Consider the terms of $g_e\left(T_b\right)$ in (S.29). Using $Z_2 = Z_0 \pm Z_\Delta$, we can deduce for the first term,

$$\begin{aligned}
&\delta'_h \left(Z'_0 M Z_2\right) \left(Z'_2 M Z_2\right)^{-1} Z_2 M e \\
&\quad = \delta'_h \left(\left(Z'_2 \pm Z_\Delta\right) M Z_2\right) \left(Z'_2 M Z_2\right)^{-1} Z_2 M e \\
&\quad = \delta'_h Z'_0 M e \pm \delta'_h Z'_\Delta M e \pm \delta'_h \left(Z'_\Delta M Z_2\right) \left(Z'_2 M Z_2\right)^{-1} Z_2 M e.
\end{aligned} \tag{S.37}$$

First, we can apply Lemma S.A.3 [(vi)-(viii)], together with Assumption 2.1-(iii), to the terms that do not involve $|N_b - N_b^0|$. Let us focus on the third term,

$$K^{-1} h^{-(1-\kappa)} \left(Z'_\Delta M Z_2\right) = \frac{Z'_\Delta Z_2}{Kh^{1-\kappa}} - \frac{Z'_\Delta X_\Delta}{Kh^{1-\kappa}} \left(X'X\right)^{-1} X' Z_2. \tag{S.38}$$

Consider $Z'_\Delta Z_\Delta$ (the argument for $Z'_\Delta X_\Delta$ is analogous). By Lemma S.A.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 $h \downarrow 0$,

$$Z'_\Delta Z_\Delta / Kh^{1-\kappa} = B O_p(1), \tag{S.39}$$

for some $B > 0$, uniformly in T_b. The second term in (S.38) is thus also $O_p(1)$ uniformly using Lemma S.A.3. Then, using (S.33) and (S.38) into the third term of (S.37), we have

$$\begin{aligned}
&\frac{1}{K} h^{-(1-\kappa)-1/2} \left(\delta_h\right)' \left(Z'_\Delta M Z_2\right) \left(Z'_2 M Z_2\right)^{-1} Z_2 M e \\
&\quad \leq \frac{1}{K} h^{-1/4} \left(\delta^0\right)' \left(\frac{Z'_\Delta M Z_2}{h^{1-\kappa}}\right) \left(Z'_2 M Z_2\right)^{-1} Z_2 M e \\
&\quad \leq h^{-1/4} \frac{Z'_\Delta M Z_2}{Kh^{1-\kappa}} O_p(1)\, O_p\left(h^{1/2}\right) \leq O_p\left(h^{1/4}\right),
\end{aligned} \tag{S.40}$$

where $(Z_2'MZ_2)^{-1} = O_p(1)$. So the right-and side of (S.40) is less than $\varepsilon/4$ in probability. Therefore, for the second term of (S.37),

$$\begin{aligned}
K^{-1}h^{-(1-\kappa)-1/2}\delta_h' Z_\Delta' Me &= \frac{h^{-1/2}}{Kh^{1-\kappa}}\delta_h' \sum_{k=T_b+1}^{T_b^0} z_{kh}e_{kh} - \frac{h^{-1/2}}{h^{1-\kappa}}\delta_h' \left(\sum_{k=T_b+1}^{T_b^0} z_{kh}x_{kh}'\right)(X'X)^{-1}(X'e) \\
&\leq \frac{h^{-1/2}}{Kh^{1-\kappa}}\delta_h' \sum_{k=T_b+1}^{T_b^0} z_{kh}e_{kh} - B\frac{1}{K}\frac{h^{-1/4}}{h^{1-\kappa}}\left(\delta^0\right)\left(\sum_{k=T_b+1}^{T_b^0} z_{kh}x_{kh}'\right)(X'X)^{-1}(X'e) \\
&\leq \frac{h^{-1/2}}{Kh^{1-\kappa}}\delta_h' \sum_{k=T_b+1}^{T_b^0} z_{kh}e_{kh} - h^{-1/4}O_p(1)\,O_p\left(h^{1/2}\right). \qquad \text{(S.41)}
\end{aligned}$$

Thus, using (S.37), (S.28) is such that

$$\begin{aligned}
&2\delta_h' Z_0' Me \pm 2\delta_h' Z_\Delta' Me \pm 2\delta_h'\left(Z_\Delta' MZ_2\right)\left(Z_2'MZ_2\right)^{-1} Z_2 Me - 2\delta_h'\left(Z_0'Me\right) \\
&\quad = 2\delta_h' Z_\Delta' Me \pm 2\delta_h'\left(Z_\Delta' MZ_2\right)\left(Z_2'MZ_2\right)^{-1} Z_2 Me \\
&\quad \leq \frac{h^{-1/2}}{Kh^{1-\kappa}}\left(\delta^0\right)' \sum_{k=T_b+1}^{T_b^0} z_{kh}\widetilde{e}_{kh} - h^{-1/4}O_p(1)\,O_p\left(h^{1/2}\right) + O_p\left(h^{-1/4}\right),
\end{aligned}$$

in view of (S.40) and (S.41). Next, consider equation (S.29). 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 $K^{-1}h^{-(3/2-\kappa)}$,

$$\begin{aligned}
K^{-1}h^{-(3/2-\kappa)}e'MZ_2\left(Z_2'MZ_2\right)^{-1}Z_2Me &= K^{-1}h^{-(3/2-\kappa)}e'MZ_0\left(Z_2'MZ_2\right)^{-1}Z_2Me \qquad \text{(S.42)} \\
&\quad \pm K^{-1}h^{-(3/2-\kappa)}e'MZ_\Delta\left(Z_2'MZ_2\right)^{-1}Z_2Me.
\end{aligned}$$

By the same argument as in (S.33), $Z_2'Me = O_p\left(h^{1/2}\right)$. Then, using the Burkhölder-Davis-Gundy inequality, estimates for the local volatility of continuous Itô semimartingales yield

$$\widetilde{e}'MZ_\Delta = \widetilde{e}'Z_\Delta - \widetilde{e}'X\left(X'X\right)^{-1}X'Z_\Delta = O_p\left(Kh^{1/2+1-\kappa}\right) - O_p\left(h^{1/2}\right)O_p(1)\,O_p\left(Kh^{1-\kappa}\right).$$

Thus, the second term in (S.42) is such that

$$\begin{aligned}
&K^{-1}h^{-(3/2-\kappa)}\widetilde{e}'MZ_\Delta\left(Z_2'MZ_2\right)^{-1}Z_2Me \qquad \text{(S.43)} \\
&\quad = B\left(K^{-1}h^{-(3/2-\kappa)}\right)O_p\left(Kh^{1-\kappa+1/2}\right)O_p(1)\,O_p\left(h^{1/2}\right) = BO_p\left(h^{1/2}\right).
\end{aligned}$$

Next, let us consider (S.29). 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.29) is asymptotically negligible. First, note that using (S.33),

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

By the fact that $Z_0 = Z_2 \pm Z_\Delta$ applied repeatedly in (S.42), and noting that the cross-product terms involving Z_Δ are $o_p(1)$ by the same reasoning as in (S.43), we obtain that the difference between the first and second term of (S.29) is negligible. The more intricate step is the one arising from

$$e'MZ_0\left(Z_0'MZ_2 \pm Z_\Delta'MZ_2\right)^{-1}Z_0'Me - e'MZ_0\left(Z_0'MZ_0\right)^{-1}Z_0'Me$$

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

On $\mathbf{D}_{K,T}$, $|N_b - N_b^0| = O_p\left(Kh^{1-\kappa}\right)$, 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.33)]

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

Therefore, (S.29) is stochastically small uniformly in $T_b \in \mathbf{D}_{K,T}$ when T is large. Altogether, we have

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

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,

$$\mathbb{E}\left(\left\|\widehat{\Sigma}_{Ze}\left(T_2, T_b^0\right) - \overline{\Sigma}_{Ze}\left(T_2, T_b^0\right)\right\| \mid \mathscr{F}_{T_bh}\right) \leq Bh^{1/2}, \tag{S.44}$$

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

$$\begin{aligned}
P\left(\sup_{T_b < T_b^0 - KT^\kappa} 2\frac{h^{-1/2}}{Kh^{1-\kappa}}\delta_h'\left\|\sum_{k=T_b+1}^{T_b^0} z_{kh}e_{kh}\right\| \geq C\right) &= P\left(\sup_{T_b < T_b^0 - KT^\kappa}\frac{h^{-1/2}}{Kh^{1-\kappa}}\delta_h'\left\|R_{1,h} + R_{2,h}(T_b)\right\| \geq 2^{-1}C\right) \\
&\leq P\left(\frac{1}{Kh^{1-\kappa}}\left\|R_{1,h}\right\| > 4^{-1}C\left\|\delta^0\right\|^{-1}h^{1/2}\right) \\
&\quad + P\left(\sup_{T_b < T_b^0 - KT^\kappa}\frac{K^{-1}}{h^{1-\kappa}}\left\|R_{2,h}(T_b)\right\| > 4^{-1}C\left\|\delta^0\right\|^{-1}h^{1/4}\right)
\end{aligned} \tag{S.45}$$

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

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

$$\leq C_r (B+1) B^{-1} \left\| \delta^0 \right\|^r h^{-r/4} T^{\kappa} h^{r/2} \leq C_r \left\| \delta^0 \right\|^r h^{r/2-\kappa-r/4} \to 0,$$

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

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

by choosing K large enough where we have used (S.44). Altogether, the right-hand side of (S.45) is less than ε, which concludes the proof. □

Proof of part (iii) of Proposition 4.1. Observe that Lemma S.A.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 have to show

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

and that the limiting distribution of $X'e/h^{1/2}$ is Gaussian. The first claim can be proven in a manner analogous to that in the proof of Proposition 3.3. For the second claim, we have the following decomposition from (S.33),

$$X'e = \sum_{k=1}^{T_b^0-\lfloor T^\kappa \rfloor} x_{kh} \widetilde{e}_{kh} + h^{-1/4} \sum_{T_b^0-\lfloor T^\kappa \rfloor+1}^{T_b^0+\lfloor T^\kappa \rfloor} x_{kh} \widetilde{e}_{kh} + \sum_{k=T_b^0+\lfloor T^\kappa \rfloor+1}^{T} x_{kh} \widetilde{e}_{kh} \triangleq R_{1,h} + R_{2,h} + R_{3,h}.$$

By Theorem S.A.5, $h^{-1/2} R_{1,h} \xrightarrow{\mathcal{L}\text{-s}} \mathscr{M}\mathscr{N}(0, V_1)$, where $V_1 \triangleq \lim_{T\to\infty} T \sum_{k=1}^{T_b^0-\lfloor T^\kappa \rfloor} \mathbb{E}\left(x_{kh} x_{kh}' \widetilde{e}_{kh}^2\right)$. Similarly, $h^{-1/2} R_{3,h} \xrightarrow{\mathcal{L}\text{-s}} \mathscr{M}\mathscr{N}(0, V_3)$, where $V_3 \triangleq \lim_{T\to\infty} T \sum_{k=T_b^0+\lfloor T^\kappa \rfloor+1}^{T} \mathbb{E}\left(x_{kh} x_{kh}' \widetilde{e}_{kh}^2\right)$. If $\kappa \in (0, 1/4)$, $h^{-(1-\kappa)} \sum_{T_b^0-\lfloor T^\kappa \rfloor+1}^{T_b^0+\lfloor T^\kappa \rfloor} x_{kh} \widetilde{e}_{kh} \xrightarrow{P} \Sigma_{Xe,N_b^0}$ by Theorem 9.3.2 in Jacod and Protter (2012) and so $h^{-1/2} R_{2,h} = h^{-3/4} \sum_{T_b^0-\lfloor T^\kappa \rfloor}^{T_b^0+\lfloor T^\kappa \rfloor} x_{kh} \widetilde{e}_{kh} \xrightarrow{P} 0$. 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 2.1-(iv) $\Sigma_{Xe,t} = 0$ 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.A.4.6 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.A.9. *The following identity holds*

$$(\delta_h)' \left\{ Z_0' M Z_0 - \left(Z_0' M Z_2\right) \left(Z_2' M Z_2\right)^{-1} \left(Z_2' M Z_0\right) \right\} \delta_h$$

$$= (\delta_h)' \left\{ Z_\Delta' M Z_\Delta - (Z_\Delta' M Z_2) (Z_2' M Z_2)^{-1} (Z_2' M Z_\Delta) \right\} \delta_h.$$

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

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

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

$$\begin{aligned}
&Q_T(T_b) - Q_T(T_0) \\
&\quad = -\delta_h' \left\{ Z_\Delta' M Z_\Delta - (Z_\Delta' M Z_2) (Z_2' M Z_2)^{-1} (Z_2' M Z_\Delta) \right\} \delta_h + g_e(T_b, \delta_h),
\end{aligned} \tag{S.46}$$

where

$$g_e(T_b, \delta_h) = 2\delta_h' (Z_0' M Z_2) (Z_2' M Z_2)^{-1} Z_2 M e - 2\delta_h' (Z_0' M e) \tag{S.47}$$

$$+ e' M Z_2 (Z_2' M Z_2)^{-1} Z_2 M e - e' M Z_0 (Z_0' M Z_0)^{-1} Z_0' M e. \tag{S.48}$$

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.A.2, we have uniformly in u as $h \downarrow 0$

$$\frac{1}{h^{1-\kappa}} \sum_{k=T_b^0+uT^\kappa}^{T_b^0} x_{kh} x_{kh}' \xrightarrow{P} \Sigma_{XX,N_b^0}. \tag{S.49}$$

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.A.4.1 [recall (S.6)], by the Burkhölder-Davis-Gundy inequality, we have that standard estimates for the local volatility yield,

$$\left\| \mathbb{E}\left(\widehat{\Sigma}_{ZX}\left(T_b, T_b^0\right) - \Sigma_{ZX,(T_b^0-1)h} | \mathscr{F}_{(T_b^0-1)h} \right) \right\| \leq B h^{1/2}. \tag{S.50}$$

Equation (S.49)-(S.50) can be used to yield, uniformly in T_b,

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

and

$$Z_\Delta' M Z_2 = Z_\Delta' Z_\Delta - Z_\Delta' X (X'X)^{-1} X' Z_2 = O_p(\psi_h) - O_p(\psi_h) O_p(1) O_p(1). \tag{S.52}$$

Now, expand the first term of (S.46),

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

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

$$\psi_h^{-1} \delta_h' Z_\Delta' M Z_\Delta \delta_h = \psi_h^{-1} \delta_h' Z_\Delta' Z_\Delta \delta_h - \psi_h^{-1} \delta_h' Z_\Delta' X (X'X)^{-1} X' Z_\Delta \delta_h. \tag{S.54}$$

By (S.51), the second term above is such that

$$\left\|\delta^{0}\right\|^{2} h^{1/2}\frac{Z_{\Delta}^{\prime}X}{\psi_{h}}\left(X^{\prime}X\right)^{-1}X^{\prime}Z_{\Delta}=\left\|\delta^{0}\right\|^{2} h^{1/2}O_{p}\left(1\right)X^{\prime}Z_{\Delta}, \tag{S.55}$$

uniformly in $T_{b}\left(u\right)$. Therefore,

$$\psi_{h}^{-1}\delta_{h}^{\prime}Z_{\Delta}^{\prime}MZ_{\Delta}\delta_{h}=\psi_{h}^{-1}\delta_{h}^{\prime}Z_{\Delta}^{\prime}Z_{\Delta}\delta_{h}-\left\|\delta^{0}\right\|^{2} h^{1/2}O_{p}\left(1\right)O_{p}\left(\psi_{h}\right). \tag{S.56}$$

The last equality shows that the second term of $\delta^{\prime}Z_{\Delta}^{\prime}MZ_{\Delta}\delta$ 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.46) when multiplied by ψ_{h}^{-1} is, uniformly in $T_{b}\left(u\right)$,

$$\psi_{h}^{-1}\delta_{h}\left(Z_{\Delta}^{\prime}MZ_{2}\right)\left(Z_{2}^{\prime}MZ_{2}\right)^{-1}\left(Z_{2}^{\prime}MZ_{\Delta}\right)\delta_{h}^{\prime}=\left\|\delta^{0}\right\|^{2} h^{1/2}O_{p}\left(1\right)O_{p}\left(1\right)O_{p}\left(\psi_{h}\right),$$

where we have used the fact that $Z_{\Delta}^{\prime}MZ_{2}/\psi_{h}=O_{p}\left(1\right)$ [cf. (S.52)]. Hence, the second term of (S.46), 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.48) 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,

$$\psi_{h}^{-1}\left[e^{\prime}MZ_{2}\left(Z_{2}^{\prime}MZ_{2}\right)^{-1}Z_{2}Me-e^{\prime}MZ_{0}\left(Z_{0}^{\prime}MZ_{0}\right)^{-1}Z_{0}^{\prime}Me\right]=o_{p}\left(c_{T}h\right).$$

Next, consider the first two terms of $g_{e}\left(T_{b},\,\delta_{h}\right)$. Using $Z_{0}^{\prime}MZ_{2}=Z_{2}^{\prime}MZ_{2}\pm Z_{\Delta}^{\prime}MZ_{2}$, it is easy to show that

$$\begin{aligned}2h^{1/4}\left(\delta^{0}\right)^{\prime}\left(Z_{0}^{\prime}MZ_{2}\right)\left(Z_{2}^{\prime}MZ_{2}\right)^{-1}Z_{2}Me-2h^{1/4}\left(\delta^{0}\right)^{\prime}\left(Z_{0}^{\prime}Me\right)&\\=2h^{1/4}\left(\delta^{0}\right)^{\prime}Z^{\prime}{}_{\Delta}Me\pm2h^{1/4}\left(\delta^{0}\right)^{\prime}Z^{\prime}{}_{\Delta}MZ_{2}\left(Z_{2}^{\prime}MZ_{2}\right)^{-1}Z_{2}^{\prime}Me.&\end{aligned} \tag{S.57}$$

Note that, uniformly in $T_{b}\left(u\right)$,

$$\begin{aligned}\psi_{h}^{-1}h^{1/4}\left(\delta^{0}\right)^{\prime}Z^{\prime}{}_{\Delta}MZ_{2}&=h^{1/4}\left(\delta^{0}\right)^{\prime}Z^{\prime}{}_{\Delta}Z_{\Delta}+\left(\delta^{0}\right)^{\prime}h^{1/4}\frac{Z_{\Delta}^{\prime}X}{\psi_{h}}\left(X^{\prime}X\right)^{-1}X^{\prime}Z_{2}\\&=h^{1/4}\left(\delta^{0}\right)^{\prime}\frac{Z^{\prime}{}_{\Delta}Z_{\Delta}}{\psi_{h}}+\left(\delta^{0}\right)^{\prime}h^{1/4}O_{p}\left(1\right)=h^{1/4}\left\|\delta^{0}\right\|O_{p}\left(1\right)+\left\|\delta^{0}\right\|h^{1/4}O_{p}\left(1\right),\end{aligned}$$

where we have used (S.49) and the fact that $\left(X^{\prime}X\right)^{-1}$ and $X^{\prime}Z_{2}$ are each $O_{p}\left(1\right)$. Recall the decomposition in (S.33):

$$X^{\prime}e=O_{p}\left(h^{1-\kappa+1/4}\right)+O_{p}\left(h^{1/2}\right). \tag{S.58}$$

Thus, the last term in (S.57) multiplied by ψ_{h}^{-1} is

$$\psi_{h}^{-1}2h^{1/4}\left(\delta^{0}\right)^{\prime}Z^{\prime}{}_{\Delta}MZ_{2}\left(Z_{2}^{\prime}MZ_{2}\right)^{-1}Z_{2}^{\prime}Me=h^{1/4}\left\|\delta^{0}\right\|O_{p}\left(1\right)O_{p}\left(1\right)\left[O_{p}\left(h^{1-\kappa+1/4}\right)+O_{p}\left(h^{1/2}\right)\right]$$

$$= \left\| \delta^0 \right\| h^{1/4} O_p(1) O_p\left(h^{1/2}\right) = \left\| \delta^0 \right\| O_p\left(h^{3/4}\right).$$

The first term of (S.57) 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, uniformly in T_b,

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

where we have used (S.49) and (S.58). Combining the last results, we have uniformly in T_b,

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

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

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

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.A.4.7 Proof of Proposition 5.1

Proof. Replace ξ_1, ξ_2, ρ and ϑ in (4.5) by their corresponding estimates ξ_1, ξ_2, ρ and ϑ, respectively. Multiply both sides of (4.5) by h^{-1} and apply a change in variable $v = s/h$. Consider the case $s < 0$. On the "fast time scale" $W^*_{\cdot}$ is replaced by $\widehat{W}_{1,h}(s) = W^*_{1,h}(sh)$ $(s < 0)$ where $W^*_{1,h}(s)$ is a sample-size dependent Wiener process. It follows that

$$-h^{-1} \frac{|s|}{2} + h^{-1} W^*_{1,h}(hs) = -\frac{|v|}{2} + W^*_1(v).$$

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

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

which is equal to the right-hand side of (4.5). Recall that $\vartheta = \left\|\delta^0\right\|^2 \overline{\sigma}^{-2} \left(\left(\delta^0\right)' \langle Z,\, Z\rangle_1 \delta^0\right)^2 / \left(\delta^0\right)' \Omega_{\mathscr{W},1} \left(\delta^0\right)$. Therefore, equation (4.5) holds when we use the proposed plug-in estimates. □

S.A.5 Proofs of Section A.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 (2.6) 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 on the order $O\left(h^{-1/2}\right)$, it has unbounded variation and does not belong to the usual class of semimartingales.[6] The predictable process $\left\{Y_{(k-1)h}\right\}_{k=1}^{T}$ 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:

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

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 $\left\{\widetilde{Y}_{(k-1)h}\right\}_{k=1}^{T}$ 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 have a corresponding assumption to Assumption S.A.1.

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

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

$$X = \begin{bmatrix} h^{1/2} & Y_0 h & \Delta_h D_1' & \Delta_h Z_1' \\ h^{1/2} & Y_1 h & \Delta_h D_2' & \Delta_h Z_2' \\ \vdots & \vdots & \vdots & \vdots \\ h^{1/2} & Y_{Th} h & \Delta_h D_T' & \Delta_h Z_T' \end{bmatrix}_{T\times(q+p+2)}. \tag{S.59}$$

Thus, $X'X$ is a $(q+p+2)\times(q+p+2)$ matrix given by

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

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.A.2, $X'X$ is $O_p\left(1\right)$ 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

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

left sub-block. Furthermore, we have

$$X'e = \begin{bmatrix} \sum_{k=1}^{T} h^{1/2} e_{kh} \\ \sum_{k=1}^{T} \left(Y_{(k-1)h} h\right) e_{kh} \\ \sum_{k=1}^{T} \Delta_h D_k e_{kh} \\ \sum_{k=1}^{T} \Delta_h Z_k e_{kh} \end{bmatrix}. \tag{S.60}$$

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 reason conditionally 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.A.10. *For $1 \le i \le 2$, $3 \le j \le p+2$ and $\gamma > 0$, the following estimates are asymptotically negligible: $\sum_{k=\lfloor s/h \rfloor}^{\lfloor t/h \rfloor} z_{kh}^{(i)} z_{kh}^{(j)} \overset{\text{u.c.p.}}{\Rightarrow} 0$, for all $N > t > s + \gamma > s > 0$.*

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

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

where the second inequality follows from the Burkhölder-Davis-Gundy inequality with parameter $r = 2$. This shows that the array $\left\{\left|z_{kh}^{(i)} z_{kh}^{(j)}\right|^2\right\}$ 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 $z_{kh}^{(2)} z_{kh}^{(j)} = \left(Y_{(k-1)h} h\right)\left(\Delta_h M_{Z,k}^{(j)}\right)$, 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.A.5.1 Proof of Proposition A.1

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

$$Q_T(T_b) - Q_T(T_0) = -\left|T_b - T_b^0\right| d(T_b) + g_e(T_b), \qquad \text{for all } T_b, \tag{S.61}$$

where

$$d(T_b) \triangleq \frac{(\delta^0)' \left\{(Z_0' M Z_0) - (Z_0' M Z_2)(Z_2' M Z_2)^{-1}(Z_2' M Z_0)\right\} \delta^0}{\left|T_b - T_b^0\right|}, \tag{S.62}$$

and we arbitrarily define $d(T_b) = (\delta^0)' \delta^0$ 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.A.11 below. Then

$$\begin{aligned} P\left(\left|\widehat{\lambda}_b - \lambda_0\right| > K\right) &= P\left(\left|\widehat{T}_b - T_b^0\right| > TK\right) \\ &\le P\left(\sup_{|T_b - T_b^0| > TK} |g_e(T_b)| \ge \inf_{|T_b - T_b^0| > TK} \left|T_b - T_b^0\right| d(T_b)\right) \\ &\le P\left(\sup_{p+2 \le T_b \le T-p-2} |g_e(T_b)| \ge TK \inf_{|T_b - T_b^0| > TK} d(T_b)\right) \end{aligned}$$

$$= P\left(d_T^{-1} \sup_{p+2\leq T_b\leq T-p-2} |g_e(T_b)| \geq K\right). \tag{S.63}$$

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

$$2\left(\delta^0\right)'\left(Z_0'MZ_2\right)\left(Z_2'MZ_2\right)^{-1/2}\left(Z_2'MZ_2\right)^{-1/2} Z_2Me. \tag{S.64}$$

For the stochastic regressors, Theorem S.A.5 implies that for any $3\leq j\leq p+2$, $(Z_2e)_{j,1}/\sqrt{h}=O_p(1)$ 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 $Z_2'MZ_2$ and $Z_0'MZ_2$ are $O_p(1)$. Then,

$$\sup_{T_b}\left(Z_0'MZ_2\right)\left(Z_2'MZ_2\right)^{-1}\left(Z_2'MZ_0\right)\leq Z_0'MZ_0=O_p(1),$$

by Lemma S.A.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.60)]. By Assumption 2.1-(iii) and the inequality

$$\sup_{T_b}\left\|\left(Z_2'MZ_2\right)^{-1/2}Z_2Me\right\|\leq\sup_{T_b}\left\|\left(Z_2'MZ_2\right)^{-1/2}\right\|\sup_{T_b}\|Z_2Me\|,$$

we have that $\left(Z_2'MZ_2\right)^{-1/2}Z_2Me$ is $O_p\left(h^{1/2}\right)$ 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.64) is $O_p\left(h^{1/2}\right)$. As for the second term of (S.10), $Z_0'Me=O_p\left(h^{1/2}\right)$. The first term in (S.11) is uniformly negligible and so is the last. Therefore, combining these results we can show that $\sup_{T_b}|g_e(T_b)|=O_p\left(\sqrt{h}\right)$. Using Lemma S.A.11 below, we have $P\left(d_T^{-1}\sup_{p+2\leq T_b\leq T-p-2}|g_e(T_b)|\geq K\right)\leq\varepsilon$, which shows that $\widehat{\lambda}_b\xrightarrow{P}\lambda_0$. □

Lemma S.A.11. *Let $d_B=\inf_{|T_b-T_b^0|>TB} Td(T_b)$. 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.A.6 (but replacing R by $\overline{R}$)

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

Under Assumption 2.1-(iii) and in view of (S.59), it can be seen that $X_\Delta'X_\Delta$ is positive definite: for the $p\times p$ lower-right sub-block apply Lemma S.A.3 as in the proof of Lemma S.A.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

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

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

Proof of part (ii) of Proposition A.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 $P\left(\sup_{T_b\in\mathbf{D}_{K,T}}Q_T(T_b)\geq Q_T\left(T_b^0\right)\right)<\varepsilon$, which is equivalently to

$$P\left(\sup_{T_b\in\mathbf{D}_{K,T}} h^{-1}g_e(T_b)\geq\inf_{T_b\in\mathbf{D}_{K,T}} h^{-1}\left|T_b-T_b^0\right|d(T_b)\right)<\varepsilon. \tag{S.65}$$

By Lemma S.A.1,

$$\inf_{T_b\in\mathbf{D}_{K,T}} d(T_b)\geq\inf_{T_b\in\mathbf{D}_{K,T}}\left(\delta^0\right)'\overline{R}'\frac{X'_\Delta X_\Delta}{T_b^0-T_b}\left(X'_2X_2\right)^{-1}\left(X'_0X_0\right)\overline{R}\delta^0.$$

For the $(q+p)\times(q+p)$ lower right sub-block of $X'_\Delta X_\Delta$ the arguments of Proposition 3.2 apply: $\left(h\left(T_b^0-T_b\right)\right)^{-1}\left[X'_\Delta X_\Delta\right]_{\{\cdot,(q+p)\times(q+p)\}}$ is bounded away from zero for all $T_b\in\mathbf{D}_{K,T}$ by choosing K large enough (recall $\left|T_b^0-T_b\right|>K$), 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 $\sum_{k=T_b+1}^{T_b^0}\left(h^{1/2}\right)^2$. Thus $\left(1/h\left(T_b^0-T_b\right)\right)\sum_{k=T_b+1}^{T_b^0}\left(h^{1/2}\right)^2>0$. The same argument applies to (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 $\sup_{T_b\in\mathbf{D}_{K,T}}\left(h\left|T_b-T_b^0\right|\right)^{-1}g_e(T_b)$ 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.A.3 to those elements which involve the stochastic regressors only, whereas the other terms are treated directly using the definition of $X'e$ in (S.60). Consider the first term of $g_e(T_b)$. Using the same steps which led to (S.19), we have

$$\begin{aligned}&\left|2\left(\delta^0\right)'\left(Z'_0MZ_2\right)\left(Z'_2MZ_2\right)^{-1}Z_2Me-2\left(\delta^0\right)'\left(Z'_0Me\right)\right|\\&=\left|\left(\delta^0\right)'Z'_\Delta Me\right|+\left|\left(\delta^0\right)'\left(Z'_\Delta MZ_2\right)\left(Z'_2MZ_2\right)^{-1}\left(Z_2Me\right)\right|.\end{aligned}\tag{S.66}$$

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

$$\left(h\left(T_b^0-T_b\right)\right)^{-1}\left(\delta^0\right)'\left(Z'_\Delta MZ_2\right)=\frac{\left(\delta^0\right)'\left(Z'_\Delta Z_\Delta\right)}{h\left(T_b^0-T_b\right)}-\left(\delta^0\right)'\left(\frac{Z'_\Delta X_\Delta}{h\left(T_b^0-T_b\right)}\left(X'X\right)^{-1}X'Z_2\right).$$

Applying the same manipulations as those used above for the $p\times p$ lower right sub-block of $Z'_\Delta Z_\Delta$, we have $\left(h\left(T_b^0-T_b\right)\right)^{-1}\left[Z'_\Delta Z_\Delta\right]_{\{\cdot,p\times p\}}=O_p(1)$, 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 $Z'_\Delta X_\Delta/\left(h\left(T_b^0-T_b\right)\right)$. 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.A.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 $C=\sup_k\left|\widetilde{Y}^2_{(k-1)h}\right|$ is a fixed constant given the localization in Assumption S.A.2. Thus, when multiplied by $\left(h\left(T_b^0-T_b\right)\right)^{-1}$, 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 $3 \leq j \leq p+2$,

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

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.A.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 $\left(Z'_\Delta Z_\Delta\right)_{1,j}$, $3 \leq j \leq p+2$. This shows that the term $\left(Z'_\Delta X_\Delta / \left(h\left(T_b^0-T_b\right)\right)\right)\left(X'X\right)^{-1} X'Z_2$ is bounded and so is $Z'_\Delta X_\Delta / \left(h\left(T_b^0-T_b\right)\right)$ using the same reasoning. Thus, $\left(h\left(T_b^0-T_b\right)\right)^{-1}\left(\delta^0\right)'\left(Z'_\Delta M Z_2\right)$ is $O_p(1)$. By the same arguments as before, we can use Theorem S.A.5 to show that the second term of (S.66) is $O_p\left(h^{1/2}\right)$ when multiplied by $\left(h\left(T_b^0-T_b\right)\right)^{-1}$ since the last term involves a positive fraction of the data. Now, expand the $(p+2)$-dimensional vector $Z'_\Delta Me$ as

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

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

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

where we recall that by Assumption 2.1-(iv) $\Sigma_{Ze,N_b^0}=0$. Note that the convergence is uniform over T_b by Lemma S.A.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 $e_{kh}/\sqrt{h} \sim \text{i.n.d.}\mathscr{N}\left(0,\, \sigma_{e,k-1}^2\right)$, 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 $C=\sup_k\left|\widetilde{Y}_{(k-1)h}\right|$ is finite by Assumption S.A.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 includes a positive fraction of data and thus

$$\sum_{k=1}^{T} x_{kh}^{(1)} e_{kh} = \sum_{k=1}^{T} h^{1/2} e_{kh} = h^{1/2} O_p(1), \tag{S.67}$$

using the basic 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.A.2. Combining the above derivations, we have

$$\frac{1}{h\left(T_b^0-T_b\right)} g_e\left(T_b\right) = \frac{1}{h\left(T_b^0-T_b\right)} \left(\delta^0\right)' 2 Z'_\Delta e + o_p(1). \tag{S.68}$$

In order to prove

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

we can use (S.68). To this end, we shall find a $K>0$, such that

$$\begin{aligned}P\left(\sup_{T_b\leq T_b^0-\frac{K}{N}}\left|\mu_\delta^0\frac{2}{h}\left(T_b^0-T_b\right)^{-1}\sum_{k=T_b+1}^{T_b^0}z_{kh}^{(1)}e_{kh}\right|>\frac{B}{3N}\right)\\ \leq P\left(\sup_{T_b\leq T_b^0-\frac{K}{N}}\left(T_b^0-T_b\right)^{-1}\left|\sum_{k=T_b+1}^{T_b^0}\frac{e_{kh}}{\sqrt{h}}\right|>\frac{B}{6\left|\mu_\delta^0\right|N}\right)<\frac{\varepsilon}{3}.\end{aligned}\tag{S.69}$$

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

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

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 $j=3,\ldots,p$, $\left\{z_{kh}^{(j)}e_{kh}/h\right\}$ 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.68), we have

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

which verifies (S.65) and thus proves our claim. □

S.A.5.2 Proof of Theorem A.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 A.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.A.12. *Under Assumption S.A.2, uniformly in T_b,*

$$\left(Q_T\left(T_b\right)-Q_T\left(T_b^0\right)\right)/\psi_h=-\delta_h\left(Z'_\Delta Z_\Delta/\psi_h\right)\delta_h\pm2\delta'_h\left(Z'_\Delta\widetilde{e}/\psi_h\right)+O_p\left(h^{3/4\wedge1-\kappa/2}\right).$$

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

$$Q_T\left(T_b\right)-Q_T\left(T_b^0\right)=-\delta'_h\left\{Z'_\Delta MZ_\Delta+\left(Z'_\Delta MZ_2\right)\left(Z'_2MZ_2\right)^{-1}\left(Z'_2MZ_\Delta\right)\right\}\delta_h+g_e\left(T_b,\,\delta_h\right).\tag{S.70}$$

We can expand the first term of (S.70) as

$$\delta'_hZ'_\Delta MZ_\Delta\delta_h=\delta'_hZ'_\Delta Z_\Delta\delta_h-\delta'_hA\delta_h,\tag{S.71}$$

where $A = Z'_{\Delta} X (X'X)^{-1} X' Z_{\Delta}$. 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.A.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. By symmetry so 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 statistics such as A separately. By Lemma S.A.3, $(X'X)^{-1} = O_p(1)$. 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 $[Z'_{\Delta} X_{\Delta}]_{\{\cdot, p \times p\}} = O_p\left(1/\left(h\left(T_b^0 - T_b\right)\right)\right)$. The same bound applies to the corresponding blocks of $Z'_{\Delta} Z_{\Delta}$ and $X'_{\Delta} Z_{\Delta}$. Now let us 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.A.2), $\widetilde{Y}_{(k-1)h}$ is bounded. Then, since the number of summands grows at a rate T^{κ}, we have $\left(Z'_{\Delta} X\right)_{2,2} = O_p\left(Kh^{1-\kappa}\right)$. The same proof can be used for $\left(Z'_{\Delta} X\right)_{1,1}$, which gives $\left(Z'_{\Delta} X\right)_{1,1} = O_p\left(Kh^{1-\kappa}\right)$. Thus, in view of (S.72), we conclude that (S.71) when divided by ψ_h is such that

$$\begin{aligned}
\delta'_h Z'_{\Delta} M Z_{\Delta} \delta_h / \psi_h &= \delta'_h Z'_{\Delta} Z_{\Delta} \delta_h / \psi_h - \delta'_h Z'_{\Delta} X \left(X'X\right)^{-1} X' Z_{\Delta} \delta_h / \psi_h \\
&= \psi_h^{-1} \left(\delta^0\right)' Z'_{\Delta} Z_{\Delta} \delta^0 - \psi_h^{-1} h^{1/2} O_p\left(h^{2(1-\kappa)}\right). \qquad \text{(S.72)}
\end{aligned}$$

For the second term of (S.70), we have

$$\begin{aligned}
&\psi_h^{-1} h^{1/2} \left(\delta^0\right)' \left\{ \left(Z'_{\Delta} M Z_2\right) \left(Z'_2 M Z_2\right)^{-1} \left(Z'_2 M Z_{\Delta}\right) \right\} \delta^0 \qquad \text{(S.73)} \\
&\quad = \psi_h^{-1} h^{1/2} \left\|\delta_0\right\|^2 O_p\left(\psi_h\right) O_p(1) O_p\left(\psi_h\right) \leq K \psi_h^{-1} h^{1/2} O_p\left(h^{2(1-\kappa)}\right)
\end{aligned}$$

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

$$\begin{aligned}
g_e\left(T_b,\, \delta_h\right) &= 2\delta'_h \left(Z'_0 M Z_2\right) \left(Z'_2 M Z_2\right)^{-1} Z_2 M e - 2\delta'_h \left(Z'_0 M e\right) \qquad \text{(S.74)} \\
&\quad + e' M Z_2 \left(Z'_2 M Z_2\right)^{-1} Z_2 M e - e' M Z_0 \left(Z'_0 M Z_0\right)^{-1} Z'_0 M e.
\end{aligned}$$

Recall (S.60), 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 $\widetilde{e}_{kh} \sim \text{i.n.d.}\, \mathscr{N}\left(0,\, \sigma_{e,k-1}^2 h\right)$. The first and third terms are $O_p\left(h^{1/2}\right)$ in view of (S.67). 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,

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

or $h^{-1/4}\sum_{k=T_b^0-\lfloor T^\kappa\rfloor}^{T_b^0+\lfloor T^\kappa\rfloor}x_{kh}^{(2)}\widetilde{e}_{kh}=h^{3/4-\kappa/2}O_p(1)$. This implies that $(X'e)_{2,1}$ is $O_p\left(h^{1/2\wedge 3/4-\kappa/2}\right)$. The same observation holds for $(X'e)_{1,1}$. Therefore, one follows the same steps as in the concluding part of the proof of Lemma 4.1 [cf. equation (S.55) and the derivations thereafter]. That is, for the first two terms of $g_e(T_b,\,\delta_h)$, using $Z_0'MZ_2=Z_2'MZ_2\pm Z_\Delta'MZ_2$, we have

$$\begin{aligned}2h^{1/4}\left(\delta^0\right)'\left(Z_0'MZ_2\right)\left(Z_2'MZ_2\right)^{-1}Z_2Me-2h^{1/4}\left(\delta^0\right)'\left(Z_0'Me\right)\\ =2h^{1/4}\left(\delta^0\right)'Z'_\Delta Me\pm 2h^{1/4}\left(\delta^0\right)'Z'_\Delta MZ_2\left(Z_2'MZ_2\right)^{-1}Z_2'Me.\end{aligned}\tag{S.75}$$

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

$$\psi_h^{-1}2h^{1/4}\left(\delta^0\right)'Z'_\Delta MZ_2\left(Z_2'MZ_2\right)^{-1}Z_2'Me=\left\|\delta^0\right\|O_p(1)\,O_p\left(h^{1\wedge 5/4-\kappa/2}\right),$$

where we have used the fact that $Z'_\Delta MZ_2/\psi_h=O_p(1)$. For the first term of (S.75),

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

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,

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

This leads to

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

for sufficiently small c_T. This together with (S.72) and (S.73) yields,

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

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

Proof of part (iii) of Theorem A.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 $\theta_h,\sigma_{h,t}$ does not alter the result of Lemma S.A.12. In addition, we have now,

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

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 $\left[N_b^0-|v|\,,\,N_b^0+|v|\right]$ 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 (A.10)-(A.11). Consider $T_b\leq T_b^0$ (i.e., $v\leq 0$). By Lemma S.A.12, there exists a $\overline{T}$ such that for all $T>\overline{T}$,

$h^{-1/2}\left(Q_T\left(T_b\right) - Q_T\left(T_b^0\right)\right)$ is

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

and note that this relationship corresponds to (A.12). 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 $\mathbb{D}\left(\mathcal{D}^*, \mathbb{R}\right)$,

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

for $\left(T_b\left(v\right)+1\right)h \leq t_v < \left(T_b\left(v\right)+2\right)h$. Recall that the lower limit of the summation is $T_b\left(v\right)+1 = T_b^0 + \lfloor v/h \rfloor$ $\left(v \leq 0\right)$ 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.A.10). Given the previous lemma we can decompose $\overline{Q}_T\left(\theta, v\right)$ as follows,

$$\overline{Q}_T\left(\theta, v\right) = \left(\delta_p^0\right)' R_{1,h}\left(v\right)\delta_p^0 + \left(\delta_Z^0\right)' R_{2,h}\left(v\right)\delta_Z^0 + 2\left(\delta^0\right)'\left(\frac{1}{\sqrt{h}}\sum_{k=T_b+1}^{T_b^0} z_{kh}\widetilde{e}_{kh}\right), \tag{S.76}$$

where

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

and δ^0 has been partitioned accordingly; that is, $\delta_p^0 = \left(\mu_\delta^0, \alpha_\delta^0\right)'$ 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 ordinary results for convergence of Riemann sums,

$$\left(\delta_p^0\right)' R_{1,h}\left(v\right)\delta_p^0 \overset{\text{u.c.p.}}{\Rightarrow} \left(\delta_p^0\right)'\begin{bmatrix} N_b^0 - N_b & \int_{N_b^0+v}^{N_b^0}\widetilde{Y}_s ds \\ \int_{N_b^0+v}^{N_b^0}\widetilde{Y}_s ds & \int_{N_b^0+v}^{N_b^0}\widetilde{Y}_s^2 ds\end{bmatrix}\delta_p^0. \tag{S.77}$$

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

$$R_{2,h}\left(v\right) \overset{\text{u.c.p.}}{\Rightarrow} \left\langle Z_\Delta, Z_\Delta\right\rangle\left(v\right). \tag{S.78}$$

We now turn to examine the asymptotic behavior of the second term in (S.76) on $\mathcal{D}^*$. We follow 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.A.2] we then have

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

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:

$$\frac{1}{\sqrt{h}}\sum_{k=T_b(v)+1}^{T_b^0}\mu_\delta^0 z_{kh}^{(1)}\widetilde{e}_{kh}\overset{\text{u.c.p.}}{\Rightarrow}\int_{N_b^0+v}^{N_b^0}\mu_\delta^0 dW_{e,s}. \tag{S.79}$$

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

$$2\left(\delta_Z^0\right)'\left(\frac{1}{\sqrt{h}}\sum_{k=T_b(v)+1}^{T_b^0}\widetilde{z}_{kh}\widetilde{e}_{kh}\right), \tag{S.80}$$

where we have partitioned z_{kh} as $z_{kh}=\left[h^{1/2}\quad Y_{(k-1)h}h\quad \widetilde{z}'_{kh}\right]'$. 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.80) 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 $(\alpha_\delta)^2\left(\sum_{k=T_b(v)+1}^{T_b^0}\left(z_{kh}^{(2)}\right)^2-\left(\int_{N_b^0+v}^{N_b^0}\widetilde{Y}_s^2ds\right)\right)$ as $\overline{Q}_{6,T}(\theta,\,v)+\overline{Q}_{7,T}(\theta,\,v)$, where $\overline{Q}_{6,T}(\theta,\,v)\triangleq(\alpha_\delta)^2\left(\sum_k\zeta_{2,h,k}^*\right)$ and $\overline{Q}_{7,T}(\theta,\,v)\triangleq(\alpha_\delta)^2\left(\sum_k{\zeta^{**}}_{2,h,k}\right)$, with

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

and

$$\zeta_{2,h,k}^{**}=2\widetilde{Y}_{(k-1)h}\left(\widetilde{Y}_{(k-1)h}h-\int_{(k-1)h}^{kh}\widetilde{Y}_sds-\mathbb{E}\left[\left(\widetilde{Y}_{(k-1)h}h-\int_{(k-1)h}^{kh}\widetilde{Y}_sds\right)\mid\mathscr{F}_{(k-1)h}\right]\right).$$

Then, we have the following decomposition for $\overline{Q}_T^{\mathrm{c}}(\theta^*) \triangleq \overline{Q}_T(\theta^*) + (\delta^0)' \Lambda(v) \delta^0$ (if $v \leq 0$ and defined analogously for $v > 0$): $\overline{Q}_T^{\mathrm{c}}(\theta^*) = \sum_{r=1}^{9} \overline{Q}_{r,T}(\theta, v)$, where $\overline{Q}_{r,T}(\theta, v)$, $r = 1, \ldots, 4$ are defined in (A.13) and $\overline{Q}_{5,T}(\theta, v) \triangleq (\mu_\delta)^2 \left(\sum_k \zeta_{1,h,k}\right)$, $\overline{Q}_{8,T}(\theta, v) \triangleq (\mu_\delta)^2 \left(h^{-1/2} \sum_k \xi_{1,h,k}\right)$, $\overline{Q}_{9,T}(\theta, v) \triangleq (\alpha_\delta)^2 \left(h^{-1/2} \sum_k \xi_{2,h,k}\right)$ where $\zeta_{1,h,k} \triangleq \left(z_{kh}^{(1)}\right)^2 - h$, $\xi_{1,h,k} \triangleq h^{1/2} \widetilde{e}_{kh}$ and $\xi_{2,h,k} \triangleq \left(\widetilde{Y}_{(k-1)h} h^{1/2}\right) \widetilde{e}_{kh}$. Moreover, recall that $\sum_k$ replaces $\sum_{T_{b(v)+1}}^{T_b^0}$ for $N_b(v) \in \mathcal{D}^*(C)$. Let us consider $\overline{Q}_{6,T}(\theta, v)$ first. For $s \in [(k-1)h, kh]$, 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, by using a maximal inequality for any $r > 1$,

$$\left[\mathbb{E} \left(\sup_{(\theta, v)} \left| (\alpha_\delta)^2 \sum_k L_{2,h,k} \right| \right)^r \right]^{1/r} \leq K_r \left(\sup_{(\theta, v)} (\alpha_\delta)^{2r} \sum_k h^r \right)^{1/r} = K_r h^{\frac{r-1}{r}}. \tag{S.81}$$

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

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

where the second inequality follows from the Burkhölder-Davis-Gundy inequality. Thus, using a maximal inequality as in (S.81), we have for $r > 1$

$$\left[\mathbb{E} \left(\sup_{(\theta, v)} \left| (\alpha_\delta)^2 \sum_k L_{1,h,k} \right| \right)^r \right]^{1/r} = K_r h^{\frac{r-1}{r}}. \tag{S.82}$$

(S.81) and (S.82) imply that $\overline{Q}_{6,T}(\cdot, \cdot)$ is stochastically equicontinuous. Next, note that $\overline{Q}_{7,T}(\theta, v)$ is a sum of martingale differences times $h^{1/2}$ (recall the definition of $\Delta_h \widetilde{V}_k = h^{1/2} \Delta_h V_k(\pi, \delta_{Z,1}, \delta_{Z,2})$). Therefore by Assumption S.A.2, for any $0 \leq s < t \leq N$, $V_t - V_s = O_p(1)$ uniformly and therefore,

$$\sup_{(\theta, v)} \left| \overline{Q}_{7,T}(\theta, v) \right| \leq K O_p\left(h^{1/2} \right). \tag{S.83}$$

Given (S.77) and (S.81)-(S.83), we deduce that $\sup_{(\theta, v)} \left\{ \left| \overline{Q}_{6,T}(\theta, v) \right| + \left| \overline{Q}_{7,T}(\theta, v) \right| \right\} = o_p(1)$. As for the term involving $R_{1,h}(v)$, it is easy to see that $\sup_{(\theta, v)} \left| \overline{Q}_{5,T}(\theta, v) \right| \to 0$. Next, we can use some of the results proved in the proof of Theorem 4.1. In particular, the asymptotic stochastic equicontinuity of the sequence of processes $\left\{ 2 (\delta_Z)' \mathscr{W}_h(v) \right\}$ follows from the same property for $\left\{ \overline{Q}_{3,T}(\theta, v) \right\}$ and $\left\{ \overline{Q}_{4,T}(\theta, v) \right\}$ proved in that proof. The stochastic equicontinuity of $(\delta_Z)' (R_{2,h}(\theta, v) - \langle Z_\Delta, Z_\Delta \rangle (v)) \delta_Z$ also follows from the same proof. Recall $\overline{Q}_{1,T}(\theta, v) + \overline{Q}_{2,T}(\theta, v)$ as defined in (A.13). Thus, stochastic equicontinuity follows from (A.15) 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 $\left\{ \left(\bar{\theta}, \bar{v} \right) \right\}$ (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,

$$\left| \overline{Q}_{9,T}(\theta, v) - \overline{Q}_{9,T}\left(\bar{\theta}, \bar{v} \right) \right| = \left| \alpha_\delta \sum_{k=T_b(v)+1}^{T_b^0} \widetilde{Y}_{(k-1)h} \widetilde{e}_{kh} - \bar{\alpha}_\delta \sum_{k=T_b^0}^{T_b(\bar{v})} \widetilde{Y}_{(k-1)h} \widetilde{e}_{kh} \right|$$

$$\leq |\alpha_\delta| \left| \sum_{k=T_b(v)+1}^{T_b^0} \widetilde{Y}_{(k-1)h} \widetilde{e}_{kh} \right| + |\bar{\alpha}_\delta| \left| \sum_{k=T_b^0}^{T_b(\bar{v})} \widetilde{Y}_{(k-1)h} \widetilde{e}_{kh} \right|. \tag{S.84}$$

For the second term, by the Burkhölder-Davis-Gundy inequality for any $r \geq 1$,

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

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 $B > 0$

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

The argument for the first term in (S.84) 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 $\left\{ \overline{Q}_{9,T} (\cdot, \cdot) \right\}$ 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 $\left\{ \overline{Q}_T^{\mathrm{c}} (\theta^*), T \geq 1 \right\}$ in (S.76) is stochastically equicontinuous. Furthermore, by (S.77) and (S.78) 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 $\left\{ \overline{Q}_T (\cdot, \cdot), T \geq 1 \right\}$ 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 (A.6). By definition, $\mathcal{D}^* (C)$ is compact and $Th \left(\widehat{\lambda}_b - \lambda_0 \right) = O_p (1)$, 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. $\square$

S.A.5.3 Proof of Proposition A.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 $\left\{ \widetilde{Y}_t^* \right\}$ converges to the non-stochastic process $\left\{ \widetilde{Y}_t^0 \right\}$ defined in (A.4) as $h \downarrow 0$. Assumption S.A.2 is maintained throughout and the constant $K > 0$ may vary from line to line.

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

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 $t < N_b^0$,

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

We can use Cauchy-Schwarz's inequality,

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

By Gronwall's inequality,

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

where $C<\infty$ is a bound on the sum of the supremum terms in the last equation above. The bound follows from Assumption S.A.2. Then, $\sup_{0\leq t\leq N}\left|\widetilde{Y}_t-\widetilde{Y}_t^0\right|\leq K\sqrt{2}B\gamma_h\exp\left(K^2N^2\right)\rightarrow 0$, as $h\downarrow 0$ (and so $\gamma_h\downarrow 0$). The assertion then follows from $\lfloor t/h\rfloor h\rightarrow t$ as $h\downarrow 0$. For $t\geq N_b^0$, one follows the same steps. □

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

Proof. Note that

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

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

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

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

Proof. Note that

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

By Markov's inequality and the dominated convergence theorem, for every $\varepsilon>0$ and every $\eta>0$

$$
\begin{aligned}
&P\Big(\left|\int_0^{N_b^0}\alpha_1\left(\widetilde{Y}_s^*-\widetilde{Y}_s^0\right)\delta_{Z,1}^{(j)}dZ_s^{(j)}\right|>\eta\Big) \\
&\quad\leq\frac{\left(\sup_{0\leq s\leq N}\sum_{r=1}^p\left(\sigma_{Z,s}^{(j,r)}\right)^2\right)^{1/2}}{\eta}\left|\alpha_1\right|\left|\delta_{Z,1}^{(j)}\right|\left(\int_0^{N_b^0}\mathbb{E}\left[\left(\widetilde{Y}_s^*-\widetilde{Y}_s^0\right)^2\right]ds\right)^{1/2},
\end{aligned}
$$

which goes to zero as $h \downarrow 0$ in view of Lemma S.A.13 and Assumption S.A.2. □

Lemma S.A.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 (A.3)],

$$\widetilde{Y}_{kh} - \widetilde{Y}_{(k-1)h} = \int_{(k-1)h}^{kh} \left(\mu_1^0 + \alpha_1^0 \widetilde{Y}_s\right) ds + \Delta_h \widetilde{V}_k \left(\pi^0,\, \delta_{Z,1}^0,\, \delta_{Z,2}^0\right).$$

Then,

$$\begin{aligned}
&\sum_{k=1}^{T_b^0} \left(\mu_1 + \alpha_1 \widetilde{Y}_{(k-1)h}\right) \left(\widetilde{Y}_{kh} - \widetilde{Y}_{(k-1)h} - \left(\mu_1^0 + \alpha_1^0 \widetilde{Y}_{(k-1)h}\right) h\right) \\
&\quad = \sum_{k=1}^{T_b^0} \int_{(k-1)h}^{kh} \left(\mu_1 + \alpha_1 \widetilde{Y}_{(k-1)h}\right) \left(\mu_1^0 + \alpha_1^0 \widetilde{Y}_s - \left(\mu_1^0 + \alpha_1^0 \widetilde{Y}_{(k-1)h}\right)\right) \\
&\quad\quad + \sum_{k=1}^{T_b^0} \int_{(k-1)h}^{kh} \left(\mu_1 + \alpha_1 \widetilde{Y}_{(k-1)h}\right) \Delta_h \widetilde{V}_k \left(\pi^0,\, \delta_{Z,1}^0,\, \delta_{Z,2}^0\right) \\
&\quad = \int_0^{N_b^0} \left(\mu_1 + \alpha_1 \widetilde{Y}_{(k-1)h}^*\right) \left(\alpha_1^0 \left(\widetilde{Y}_s - \widetilde{Y}_{(k-1)h}^*\right)\right) ds + B\gamma_h \int_0^{N_b^0} \left(\mu_1 + \alpha_1 \widetilde{Y}_s^*\right) dV_s.
\end{aligned}$$

For the first term on the right-hand side,

$$\begin{aligned}
\sup_{\mu_1, \alpha_1} \left| \int_0^{N_b^0} \left(\mu_1 + \alpha_1 \widetilde{Y}_s^*\right) \left(\alpha_1^0 \left(\widetilde{Y}_s - \widetilde{Y}_s^*\right)\right) ds \right| &\leq \left|\alpha_1^0\right| \left| \int_0^{N_b^0} \sup_{\mu_1, \alpha_1} \left(\mu_1 + \alpha_1 \widetilde{Y}_s^*\right) \left(\widetilde{Y}_s - \widetilde{Y}_s^0 + \widetilde{Y}_s^0 - \widetilde{Y}_s^*\right) ds \right| \\
&\leq \left|\alpha_1^0\right| K \left(\int_0^{N_b^0} \sup_{0 \leq s \leq N_b^0} \left|\widetilde{Y}_s - \widetilde{Y}_s^0\right| + \sup_{0 \leq s \leq N_b^0} \left|\widetilde{Y}_s^0 - \widetilde{Y}_s^*\right| ds \right),
\end{aligned}$$

which is $o_p(1)$ as $h \downarrow 0$ from Lemma S.A.13 and Assumption S.A.2. Next, consider the vector of regressors Z, and note that for any $3 \leq j \leq p+2$,

$$B\gamma_h \sup_{\mu_1, \alpha_1} \left| \int_0^{N_b^0} \left(\mu_1 + \alpha_1 \widetilde{Y}_s^*\right) dZ_s^{(j)} \right| \leq B\gamma_h \sup_{\mu_1, \alpha_1} \left| \int_0^{N_b^0} \left(\mu_1 + \alpha_1 \widetilde{Y}_s^*\right) \sum_{r=1}^{p} \sigma_{Z,s}^{(j,r)} dW_Z^{(r)} \right|.$$

Let $R_{j,h} = R_{j,h}(\mu_1,\, \alpha_1) \triangleq \int_0^{N_b^0} B\gamma_h \left(\mu_1 + \alpha_1 \widetilde{Y}_s^*\right) \sum_{r=1}^{p} \sigma_{Z,s}^{(j,r)} dW_Z^{(r)}$ (we index R_j by h because $\widetilde{Y}_s^*$ depends on h). Then, we want to show that, for every $\varepsilon > 0$ and $K > 0$,

$$P\left(\sup_{\mu_1, \alpha_1} \left|R_{j,h}(\mu_1,\, \alpha_1)\right| > K\right) \leq \varepsilon. \tag{S.85}$$

In view of Chebyshev's inequality and the Itô's isometry,

$$P\left(\left|R_{j,h}\right| > K\right) \leq \left(\frac{B\gamma_h}{K}\right)^2 \mathbb{E}\left[\left|\int_0^{N_b^0} \left(R_{j,h} / \left(B\gamma_h\right)\right)\right|^2\right],$$

$$\leq \left[\sup_{0\leq s\leq N}\sum_{r=1}^{p}\left(\sigma_{Z,s}^{(j,r)}\right)^{2}\right]\left(\frac{B\gamma_{h}}{K}\right)^{2}\int_{0}^{N_{b}^{0}}\mathbb{E}\left[\left|\mu_{1}+\alpha_{1}\widetilde{Y}_{s}^{*}\right|^{2}ds\right],$$

so that by the boundness of the processes (cf. Assumption S.A.2) and the compactness of Θ_0, we have for some $A<\infty$,

$$P\left(\left|R_{j,h}\right|>K\right)\leq A\left[\sup_{0\leq s\leq T}\sum_{r=1}^{p}\left(\sigma_{Z,s}^{(j,r)}\right)^{2}\right]\left(\frac{B\gamma_{h}}{K}\right)^{2}\rightarrow 0, \tag{S.86}$$

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 $\mathbb{E}\left[\left|R_{j,h}\right|^{2m}\right]\leq K$ which follows using the same steps that led to (S.86) with the Burkhölder-Davis-Gundy inequality in place of the Itô's isometry. Let $g\left(\widetilde{Y}_{s}^{*},\,\widetilde{\theta}\right)\triangleq\mu_{1,1}+\alpha_{1,1}\widetilde{Y}_{s}^{*}$, $\widetilde{\theta}_{1}\triangleq(\mu_{1,1},\alpha_{1,1})'$ 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,

$$\begin{aligned}
&\mathbb{E}\left[\left|R_{j,h}\left(\widetilde{\theta}_{2}\right)-R_{j,h}\left(\widetilde{\theta}_{1}\right)\right|^{2m}\right]\\
&\leq\left(B\gamma_{h}\right)^{2m}K_{m}\left[\sup_{0\leq s\leq N}\sum_{r=1}^{p}\left(\sigma_{Z,s}^{(j,r)}\right)^{2}\right]^{m}\mathbb{E}\left[\left(\int_{0}^{N_{b}^{0}}\left(g\left(\widetilde{Y}_{s}^{*},\,\widetilde{\theta}_{2}\right)-g\left(\widetilde{Y}_{s}^{*},\,\widetilde{\theta}_{1}\right)\right)^{2}ds\right)^{m}\right]\\
&\leq\left(B\gamma_{h}\right)^{2m}K_{m}\left[\sup_{0\leq s\leq N}\sum_{r=1}^{p}\left(\sigma_{Z,s}^{(j,r)}\right)^{2}\right]^{m}\mathbb{E}\left[\left(\int_{0}^{N_{b}^{0}}\left(\left(\mu_{1,2}-\mu_{1,1}\right)+\left(\alpha_{1,2}-\alpha_{1,1}\right)\widetilde{Y}_{s}^{*}\right)^{2}ds\right)^{m}\right]\\
&\leq\left(B\gamma_{h}\right)^{2m}K_{m}\left[\sup_{0\leq s\leq N}\sum_{r=1}^{p}\left(\sigma_{Z,s}^{(j,r)}\right)^{2}\right]^{m}\mathbb{E}\left[\left(\int_{0}^{N_{b}^{0}}\left(\left(\mu_{1,2}-\mu_{1,1}\right)+\left(\alpha_{1,2}-\alpha_{1,1}\right)C\right)^{2}ds\right)^{m}\right]\\
&\leq\left(B\gamma_{h}\right)^{2m}K_{m}\mathbb{E}\left[\left(\int_{0}^{N_{b}^{0}}\left(2\left(\mu_{1,2}-\mu_{1,1}\right)^{2}+2C\left(\alpha_{1,2}-\alpha_{1,1}\right)^{2}\right)ds\right)^{m}\right]\\
&\leq\left(B\gamma_{h}\right)^{2m}K_{m}\mathbb{E}\left[\left(\int_{0}^{N_{b}^{0}}\left(2\left(\mu_{1,2}-\mu_{1,1}\right)^{2}+2\left(\alpha_{1,2}-\alpha_{1,1}\right)^{2}-2\left(\alpha_{1,2}-\alpha_{1,1}\right)^{2}+2C\left(\alpha_{1,2}-\alpha_{1,1}\right)^{2}\right)ds\right)^{m}\right]\\
&\leq 2^{m}\left(B\gamma_{h}\right)^{2m}K_{m}\left\Vert 2\left(\widetilde{\theta}_{2}-\widetilde{\theta}_{1}\right)\right\Vert^{2m}\left(\int_{0}^{N_{b}^{0}}ds\right)^{m}+2^{m}\left(B\gamma_{h}\right)^{2m}K\left(\widetilde{\theta}_{1},\,\widetilde{\theta}_{2},\,m,\,C\right)
\end{aligned} \tag{S.87}$$

where $C=\sup_{s\geq 0}\left|\widetilde{Y}_{s}^{*}\right|$, $K\left(\widetilde{\theta}_{1},\,\widetilde{\theta}_{2},\,m,\,C\right)$ is some constant that depends on its arguments and we have used 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.86) and (S.87) yield (S.85). Since the same result can be shown to remain valid for each term in the stochastic element $\Delta_h V_k(\pi,\,\delta_{Z,1},\,\delta_{Z,2})$, this establishes the claim. □

Proof of Proposition A.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,

$$\widehat{\mu}_{1}\widehat{N}_{b}=\widetilde{Y}_{\widehat{N}_{b}}-\widetilde{Y}_{0}-\widehat{\alpha}_{1}h\sum_{k=1}^{\widehat{T}_{b}}\widetilde{Y}_{(k-1)h} \tag{S.88}$$

$$\widehat{\alpha}_1 = \frac{\sum_{k=1}^{\widehat{T}_b} \left(\widetilde{Y}_{kh} - \widetilde{Y}_{(k-1)h}\right) \widetilde{Y}_{(k-1)h} - \left(\widehat{N}_b^{-1} \left(\widetilde{Y}_{\widehat{N}_b} - \widetilde{Y}_0\right)\right) h \sum_{k=1}^{\widehat{T}_b} \widetilde{Y}_{(k-1)h}}{h \sum_{k=1}^{\widehat{T}_b} \widetilde{Y}^2_{(k-1)h} - \widehat{N}_b^{-1} \left(h \sum_{k=1}^{\widehat{T}_b} \widetilde{Y}_{(k-1)h}\right)^2}. \tag{S.89}$$

Then, assuming $\widehat{T}_b < T_b^0$,

$$\widehat{\alpha}_1 = \frac{\sum_{k=1}^{\widehat{T}_b} \left(\mu_1^0 h + \alpha_1^0 \widetilde{Y}_{(k-1)h} h + \Delta_h \widetilde{V}_{h,k}\right) \widetilde{Y}_{(k-1)h}}{h \sum_{k=1}^{\widehat{T}_b} \widetilde{Y}^2_{(k-1)h} - \widehat{N}_b^{-1} \left(h \sum_{k=1}^{\widehat{T}_b} \widetilde{Y}_{(k-1)h}\right)^2}$$
$$- \frac{\left(\mu_1^0 + \alpha_1^0 \widehat{N}_b^{-1} \sum_{k=1}^{\widehat{T}_b} \widetilde{Y}_{(k-1)h} h + \widehat{N}_b^{-1} B \gamma_h \left(V_{\widehat{N}_b} - V_0\right)\right) h \sum_{k=1}^{\widehat{T}_b} \widetilde{Y}_{(k-1)h}}{h \sum_{k=1}^{\widehat{T}_b} \widetilde{Y}^2_{(k-1)h} - \widehat{N}_b^{-1} \left(h \sum_{k=1}^{\widehat{T}_b} \widetilde{Y}_{(k-1)h}\right)^2} + o_p(1),$$

and thus

$$\widehat{\alpha}_1 = \frac{\sum_{k=1}^{T_b^0} \left(\mu_1^0 h + \alpha_1^0 \widetilde{Y}_{(k-1)h} h + \Delta_h \widetilde{V}_k\right) \widetilde{Y}_{(k-1)h}}{h \sum_{k=1}^{\widehat{T}_b} \widetilde{Y}^2_{(k-1)h} - \widehat{N}_b^{-1} \left(h \sum_{k=1}^{\widehat{T}_b} \widetilde{Y}_{(k-1)h}\right)^2}$$
$$- \frac{\left(\mu_1^0 + \alpha_1^0 \widehat{N}_b^{-1} \sum_{k=1}^{T_b^0} \widetilde{Y}_{(k-1)h} h + \widehat{N}_b^{-1} B \gamma_h \left(V_{N_b^0} - V_0\right)\right) h \sum_{k=1}^{T_b^0} \widetilde{Y}_{(k-1)h}}{h \sum_{k=1}^{\widehat{T}_b} \widetilde{Y}^2_{(k-1)h} - \widehat{N}_b^{-1} \left(h \sum_{k=1}^{\widehat{T}_b} \widetilde{Y}_{(k-1)h}\right)^2}$$
$$- \frac{\sum_{k=\widehat{T}_b+1}^{T_b^0} \left(\mu_1^0 h + \alpha_1^0 \widetilde{Y}_{(k-1)h} h + \Delta_h \widetilde{V}_k\right) \widetilde{Y}_{(k-1)h}}{h \sum_{k=1}^{\widehat{T}_b} \widetilde{Y}^2_{(k-1)h} - \widehat{N}_b^{-1} \left(h \sum_{k=1}^{\widehat{T}_b} \widetilde{Y}_{(k-1)h}\right)^2}$$
$$+ \frac{\widehat{N}_b^{-1} \left(\sum_{k=\widehat{T}_b+1}^{T_b^0} \mu_1^0 h + \alpha_1^0 \sum_{k=\widehat{T}_b+1}^{T_b^0} \widetilde{Y}_{(k-1)h} h + B \gamma_h \left(V_{N_b^0} - V_{\widehat{N}_b}\right)\right) h \sum_{k=\widehat{T}_b+1}^{T_b^0} \widetilde{Y}_{(k-1)h}}{h \sum_{k=1}^{\widehat{T}_b} \widetilde{Y}^2_{(k-1)h} - \widehat{N}_b^{-1} \left(h \sum_{k=1}^{\widehat{T}_b} \widetilde{Y}_{(k-1)h}\right)^2}.$$

By part (ii) of Theorem A.1, $N_b^0 - \widehat{N}_b = O_p\left(h^{1-\kappa}\right)$, and thus it is easy to see that the third and fourth terms go to zero in probability at a slower rate than $h^{1-\kappa}$. As for the first and second terms, recalling that $\Delta_h \widetilde{V}_{h,k} = h^{1/2} \Delta V_{h,k}$ from (A.3), we have by ordinary convergence of approximations to Riemann sums, Lemma S.A.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.A.15, $\sum_{k=1}^{T_b^0} \widetilde{Y}_{(k-1)h} \Delta_h \widetilde{V}_k \xrightarrow{P} 0$. Thus, we deduce that

$$\widehat{\alpha}_1 = \alpha_1^0 + O_p\left(B \gamma_h\right). \tag{S.90}$$

Using (S.90) into (S.88),

$$\widehat{\mu}_1 \widehat{N}_b = \widetilde{Y}_{\widehat{N}_b} - \widetilde{Y}_0 - \alpha_1^0 h \sum_{k=1}^{\widehat{T}_b} \widetilde{Y}_{(k-1)h} - O_p\left(B \gamma_h\right),$$

$$= \widetilde{Y}_{\widehat{N}_b} - \widetilde{Y}_0 - \alpha_1^0 h \sum_{k=1}^{T_b^0} \widetilde{Y}_{(k-1)h} - \alpha_1^0 h \sum_{k=\widehat{T}_b+1}^{T_b^0} \widetilde{Y}_{(k-1)h} - o_p(1).$$

By part (ii) of Theorem A.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

$$\widehat{\mu}_1 \widehat{N}_b = \widetilde{Y}_{N_b^0} - \widetilde{Y}_0 - \alpha_1^0 \int_0^{N_b^0} \widetilde{Y}_s ds - \alpha_1^0 O_p\left(h^{1-\kappa}\right) - o_p(1).$$

Noting that

$$\widetilde{Y}_{N_b^0} - \widetilde{Y}_0 = \mu_1^0 N_b^0 + \alpha_1^0 \int_0^{N_b^0} \widetilde{Y}_s ds + O_p(B\gamma_h)\left(V_{N_b^0} - V_0\right),$$

we have $\widehat{\mu}_1 N_b^0 = \mu_1^0 N_b^0 + O_p(B\gamma_h)\left(V_{N_b^0} - V_0\right)$, which yields

$$\widehat{\mu}_1 = \mu_1^0 + O_p(B\gamma_h). \tag{S.91}$$

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. Further, 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.90) and (S.91),

$$\begin{aligned} \widehat{u}_{kh} &= h^{-1/2}\left(\mu_1^0 h + \alpha_1^0 \widetilde{Y}_{(k-1)h} h + B\gamma_h \Delta_h V_k - \mu_1^0 h - O_p\left(h^{3/2}\right) - \alpha_1^0 \widetilde{Y}_{(k-1)h} h - O_p\left(h^{3/2}\right)\right) \\ &= h^{-1/2} B\gamma \Delta_h V_k - O_p(h). \end{aligned} \tag{S.92}$$

Similarly, for $T_b^0 \leq \widehat{T}_b \leq k$,

$$\widehat{u}_{kh} = h^{-1/2} B\gamma_h \Delta_h V_k - O_p(h), \tag{S.93}$$

whereas for $\widehat{T}_b < k \leq T_b^0$,

$$\begin{aligned} \widehat{u}_{kh} &= h^{-1/2}\left(\mu_1^0 h + \alpha_1^0 \widetilde{Y}_{(k-1)h} h + B\gamma_h \Delta_h V_k - \mu_2^0 h - O_p\left(h^{3/2}\right) - \alpha_2^0 \widetilde{Y}_{(k-1)h} h - O_p\left(h^{3/2}\right)\right) \\ &= h^{-1/2}\left(-\mu_\delta^0 h - \alpha_\delta^0 \widetilde{Y}_{(k-1)h} h + B\gamma_h \Delta_h V_k - O_p\left(h^{3/2}\right)\right) \\ &= -\mu_\delta^0 h^{1/2} - \alpha_\delta^0 \widetilde{Y}_{(k-1)h} h^{1/2} + h^{-1/2} B\gamma_h \Delta_h V_k - O_p(h). \end{aligned} \tag{S.94}$$

Next, note that $\sum_{k=\widehat{T}_b+1}^{T_b^0} \mu_\delta^0 h^{1/2} \leq Kh^{1/2-\kappa}$ and $\sum_{k=\widehat{T}_b+1}^{T_b^0} \alpha_\delta^0 \widetilde{Y}_{(k-1)h} h^{1/2} \leq Kh^{1/2-\kappa}$ since by Theorem

A.1-(ii) there are T^{κ} terms in each sum. Moreover, recall that $e_{kh} = \Delta_h e_k^* \sim \mathscr{N}\left(0,\, \sigma_{e,k-1}^2 h\right)$ and thus[7] $\sum_{k=\widehat{T}_b+1}^{T_b^0} e_{kh} = \sqrt{h}\sum_{k=\widehat{T}_b+1}^{T_b^0} h^{-1/2} e_{kh} = h^{1/2-\kappa} o_p(1)$. Therefore, $\sum_{k=\widehat{T}_b+1}^{T_b^0} \widehat{u}_{kh} = K o_p\left(h^{1/2-\kappa}\right)$. Since $\kappa \in (0,\, 1/2)$, this shows that the residuals $\widehat{u}_{kh}$ from equation (S.94) are asymptotically negligible. That is, asymptotically the estimator of $\left((\beta_S^0)',\, \left(\delta_{Z,1}^0\right)',\, \left(\delta_{Z,2}^0\right)'\right)'$ minimizes (assuming $\widehat{T}_b \leq T_b^0$),

$$\sum_{k=1}^{\widehat{T}_b} \left(\widehat{u}_{kh} - \widetilde{x}_{kh}'\beta_S\right)^2 + \sum_{k=T_b^0+1}^{T} \left(\widehat{u}_{kh} - \widetilde{x}_{kh}'\beta_S - \widetilde{z}_{0,kh}'\delta_S\right)^2 + o_p(1),$$

where $X = \begin{bmatrix}\widetilde{X}^{(1)} & \widetilde{X}^{(2)} & \widetilde{X}\end{bmatrix}$, $\beta^0 = \begin{bmatrix}\mu_1^0 & \alpha_1^0 & (\beta_S^0)'\end{bmatrix}'$, and Z_0 and δ_S^0 are partitioned accordingly. The subscript S indicates that these are the parameters of the stochastic semimartingale regressors. But 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: $\widehat{U} = \widetilde{X}\widehat{\beta}_S + \widehat{Z}_0\widehat{\delta}_S +$ 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 $\widehat{U} = \left(\widehat{u}_{kh};\, k = 1, \ldots, \widehat{T}_b,\, T_b^0+1, \ldots, N\right)$. Therefore, using (S.92) and (S.93), we have

$$h^{-1/2}\begin{bmatrix}\widehat{\beta}_S - \beta^0 \\ \widehat{\delta}_S - \delta^0\end{bmatrix} = \begin{bmatrix}\widetilde{X}'\widetilde{X} & \widetilde{X}'\widehat{Z}_0 \\ \widehat{Z}_0'\widetilde{X} & \widehat{Z}_0'\widehat{Z}_0\end{bmatrix}^{-1} h^{-1/2}\begin{bmatrix}\widetilde{X}'e & \widetilde{X}'\left(Z_0 - \widehat{Z}_0\right)\delta^0 + \widetilde{X}'AO_p(h) \\ \widehat{Z}_0'e & \widehat{Z}_0'\left(Z_0 - \widehat{Z}_0\right)\delta^0 + \widehat{Z}_0'AO_p(h)\end{bmatrix},$$

for some matrix $A = O_p(1)$. It then follows by the same proof as in Proposition 3.3 that

$$\begin{bmatrix}\widetilde{X}'\widetilde{X} & \widetilde{X}'\widehat{Z}_0 \\ \widehat{Z}_0'\widetilde{X} & \widehat{Z}_0'\widehat{Z}_0\end{bmatrix}^{-1} \widetilde{X}'AO_p\left(h^{1/2}\right) = o_p(1), \tag{S.95}$$

and

$$\begin{bmatrix}\widetilde{X}'\widetilde{X} & \widetilde{X}'\widehat{Z}_0 \\ \widehat{Z}_0'\widetilde{X} & \widehat{Z}_0'\widehat{Z}_0\end{bmatrix}^{-1} \frac{1}{h^{1/2}}\widetilde{X}'\left(Z_0 - \widehat{Z}_0\right)\delta^0 = O_p(1)\, o_p(1) = o_p(1). \tag{S.96}$$

The same arguments can be used for $\widehat{Z}_0'\left(Z_0 - \widehat{Z}_0\right)\delta^0$ and $\widehat{Z}_0'AO_p(h)$. Therefore, in view of (S.90) and (S.91), we obtain $\widehat{\mu}_1 = \mu_1^0 + o_p(1)$ and $\widehat{\alpha}_1 = \alpha_1^0 + o_p(1)$, respectively, whereas (S.95) and (S.96) 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, $\widehat{N}_b - \widehat{N}_b^0 = O_p\left(h^{1-\kappa}\right)$ and one can follow the same steps that led to (S.90) and (S.91) and proceed as above. The final result is $\widehat{\theta} = \theta^0 + o_p(1)$, which is what we wanted to show. □

[7]The same bound holds for the corresponding sum involving the other terms in $\Delta_h V_k$.

S.A.5.4 Negligibility of the Drift Term

Recall Lemma S.A.10 and apply the same proof as in Section A.3.3. 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.B Additional Discussion about the Continuous Record Asymptotic Density Function

S.B.1 Further Discussion from Section 4.2

In this section, we continue our discussion about the properties of the continuous record asymptotic distribution from Section 4. It is useful to plot the probability densities for a fractional break date λ_0 close to the endpoints. Figure S-1 presents the densities of $\rho\left(\widehat{T}_b - T_b^0\right)$ given in equation (4.5) for $\rho^2 = 0.2$, 0.3, 0.5 (the left, middle and right panel, respectively) and a true break point $\lambda_0 = 0.2$. The figure also reports the density of the shrinkage large-N asymptotic distribution. We report corresponding plots for $\lambda_0 = 0.35$, 0.5, 0.75 in Figure S-2-S-4. The shape of the density of the shrinkage large-N asymptotic distribution is seen to remain unchanged as we raise the signal-to-noise ratio. It is always symmetric, uni-modal and centered at the true value λ_0. This contrasts with the density derived under a continuous record. From Figure S-1 it is easily seen that when the break size is small, the density from Theorem 4.2 is always asymmetric suggesting that the location of the break date indeed plays a key role in shaping the asymptotic distribution even if the regressors and errors have the same distribution across adjacent regimes. As we raise the signal-to-noise ratio (from left to right panel) the distribution becomes less asymmetric and accordingly less positively skewed but both features are still evident. An additional feature arises from this plot. There are only two modes when $\lambda_0 = 0.2$ (cf. Figure S-1, left and middle panels), the mode at the true value being no longer present. When the date of the break is not in the middle 80% of the sample, the density shows bi-modality rather than tri-modality as we discussed in Section 4. This constitutes the only exception to the otherwise similar comments that can be made when $\lambda_0 = 0.35$ and 0.5 (cf. Figure S-2-S-3). Figure S-4 displays the case for $\lambda_0 = 0.75$. Once again, the distribution is asymmetric. Since λ_0 is located in the second half of the sample, the density is negatively skewed.

When we consider nearly stationary regimes, that is, we allow for low heterogeneity across regimes according to the restrictions in (4.6), the results are not affected. However, observe that when the heterogeneity is higher, there are few notable distinctions as explained in the main text.

The features of the density under a continuous record arise from the properties of the limiting process. Consider the process $\mathscr{V}(s)$ as defined before Theorem 4.2. The limiting distribution is related to the extremum of $\mathscr{V}(s)$ over a fixed time interval with boundary points $-\rho N_b^0/\left\|\delta^0\right\|^{-2}\overline{\sigma}^2$ and $\rho\left(N-N_b^0\right)/\left\|\delta^0\right\|^{-2}\overline{\sigma}^2$. $\mathscr{V}(s)$ has a continuous sample path and it is the sum of a deterministic component or drift and a stochastic Gaussian component. The deterministic part is given by the second moments of the regressors and thus it is always negative because of the minus in front of it. The term $(|s|/2)\left(\delta^0\right)'\langle Z, Z\rangle_{(\cdot)}\delta^0$ is of $|s|$ whereas the stochastic term is of order $|s|^{1/2}$. This means that for small $|s|$, the highly-fluctuating Gaussian part is more influential. However, when the signal-to-noise ratio is large (ρ is high), the deterministic part dominates the stochastic one. Thus, the maximum of $\mathscr{V}(s)$ cannot be attained at large values of $|s|$. This explains why there is only one mode at the origin when the signal is high. In contrast, when the signal-to-noise ratio is low, the interval over which $\mathscr{V}(s)$ is maximized is short. Hence, the fluctuations in the stochastic part dominates that of the deterministic one as the former is of higher order on that interval. This has at least two consequences. First, there is another mode at each of the endpoints because by the property of the Gaussian part of $\mathscr{V}(s)$ it is much more likely to attain a maximum close to the boundary points than at any interior point. We refer to Karatzas and Shreve (1996) for an accessible treatment about the probabilistic aspects of this class of processes. Second, when ρ is low, so is $(|s|/2)\left(\delta^0\right)'\langle Z, Z\rangle_{(\cdot)}\delta^0$, and thus it is more likely that the maximum is achieved at either endpoint than at zero. This explains why when the signal is low the highest mode is not at the origin. When $\lambda_0 \neq 0.5$, the interval over which $\mathscr{V}(s)$ is maximized is asymmetric and as a consequence the density is also asymmetric. If ρ is not very large, when λ_0 is less (larger) than 0.5 there is a higher probability for $\mathscr{V}(s)$ to attain a maximum closer to the left (right) boundary point since the deterministic component takes a less negative value at $-\rho N_b^0/\left\|\delta^0\right\|^{-2}\overline{\sigma}^2$ and $\rho\left(N-N_b^0\right)/\left\|\delta^0\right\|^{-2}\overline{\sigma}^2$. When the size of the break is sufficiently high, the density is always symmetric and has unique mode at a value corresponding to $\widehat{\lambda}_b$ being close to λ_0 because the deterministic component $(|s|/2)\left(\delta^0\right)'\langle Z, Z\rangle_{(\cdot)}\left(\delta^0\right)$ is large enough that $\mathscr{V}(s)$ decreases as it moves away from the origin. Thus, with very high probability, the maximum is located at the origin.

When considering non-stationary regimes, the heterogeneity across regimes determines the stochastic order of the process $\mathscr{V}(s)$. If the post-break regime has higher volatility, there is a high probability that the limiting process attains a maximum on the interval $\left[0, \rho\left(N-N_b^0\right)/\left\|\delta^0\right\|^{-2}\overline{\sigma}^2\right]$ since it fluctuates more in that region. This explains why the density is clearly negatively skewed and the mode near the right boundary point is always higher than the mode near the left boundary point (Figure S-9-S-11, right panel).

Consider now the extreme cases $\lambda_0 = 0.1$, 0.45, 0.55 and $\lambda_0 = 0.9$. The characteristics discussed in Section 5 remain valid as can be seen from Figure S-5-S-8. The features of skewness, asymmetry, tri-modaility (only when ρ is low) and peakedness (when ρ is high) are all more pronounced for those relatively more extreme cases. For example, in Figure S-5 we plot the densities of $\widehat{T}_b$ for a true break fraction $\lambda_0 = 0.1$ (near the beginning of the sample). When ρ is low there are now only two modes because the mode associated with the middle point has disappeared. This bi-modaility vanishes as we increase ρ, and the density is positively skewed for all values of the signal-to-noise ratio. Similar comments apply to the other cases.

That the density is symmetric only if the break date is at half sample ($\lambda_0 = 0.5$) and that this property is sharp, can be seen from Figure S-6-S-7, left panel. When the true break date is not exactly at 0.5 but, e.g. as close as 0.45, the density is visibly asymmetric. Further, in such a case the density is positively skewed and the highest mode is towards beginning of the sample.

S.B.2 Further Discussion from Section 5

We continue with the analysis of cases allowing differences between the distribution of the errors and regressors in the pre- and post-break regimes (i.e., non-stationary regimes). We consider a scenario where the second regime is twice as volatile as the first. Here the signal-to-noise ratio is given by $\delta^0/\sigma_{e,1}$, where $\sigma_{e,1}^2$ is the variance of the error term in the first regime. We notice substantial similarities with the cases considered above but there is one notable exception. In Figure S-9-S-12, the shrinkage asymptotic density of Bai (1997) is asymmetric and unimodal for all pairs (ρ^2, λ_0) considered. The density is negatively skewed and the right tail much fatter then the left tail. Turning to the density from (4.6), we can make the following observations. Even if the signal-to-noise ratio is moderately high, the asymptotic distribution deviates from being symmetric when the break occurs at exactly middle sample ($\lambda_0 = 0.5$, Figure S-11, right panel). This is in contrast to the nearly stationary framework since the density was shown to be always symmetric no matter the value taken by ρ if $\lambda_0 = 0.5$. This suggests that when the statistical properties of the errors and regressors display significant differences across the two regimes, the probability densities derived under a continuous record is not symmetric even with $\lambda_0 = 0.5$. This means that the asymptotic distribution attributes different weights to the informational content of the two regimes since they possess highly heterogeneous statistical characteristics. Figure S-11 makes this point clear. It reports plots for the case with $\lambda_0 = 0.5$ and $\rho^2 = 0.3, 0.8, 1.5$ (from left to right panels). The density is no longer symmetric and the right tail is much fatter than the left one. This follows simply because there is more variability in the post-break region. In such cases, there is a tendency to overestimate the break point which leads to an upward bias if the the post-break regime displays larger variability. There are important differences with respect to Bai's (1997) density. First, although the density under a continuous record asymptotics is also asymmetric for all λ_0, the degree of asymmetry varies across different break dates. Second, there is multi-modality when the size of the break is small which is not shared with Bai's (1997) density since the latter is always unimodal. Finally, one should expect the density to be symmetric when the magnitude of the break is large as the distribution should collapse at λ_0 for large breaks. The continuous record asymptotics reproduces this property whereas the large-N asymptotic distribution of Bai (1997) remains asymmetric even for large break sizes (Figure S-12).

S.B.3 Figures

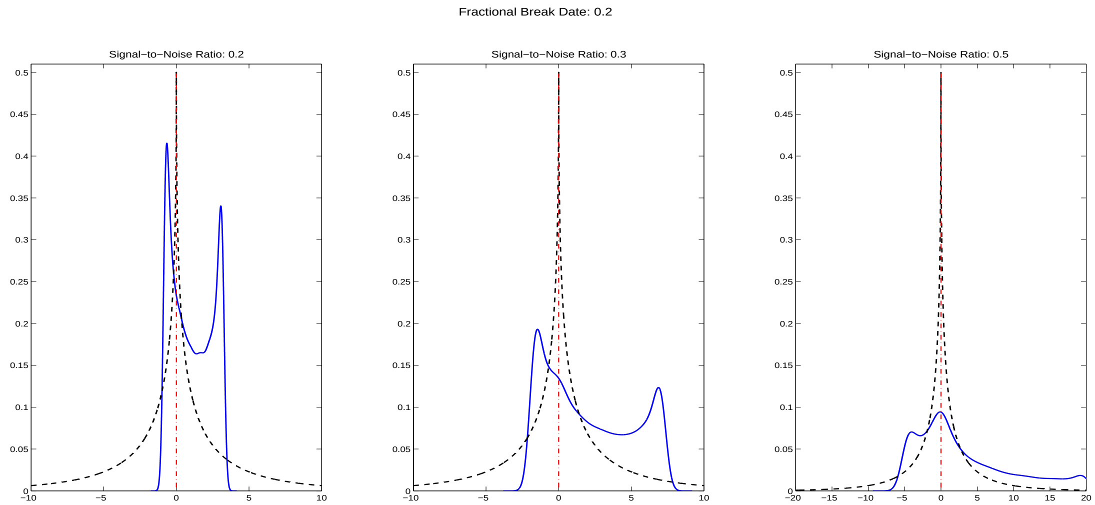

Figure S-1: The asymptotic probability density of $\rho\left(\widehat{T}_b - T_b^0\right)$ derived under a continuous record (blue solid line) and the density of Bai's (1997) asymptotic distribution (black broken line) for $\lambda_0 = 0.2$ and $\rho^2 = 0.2$, 0.3 and 0.5 (the left, middle and right panel, respectively).

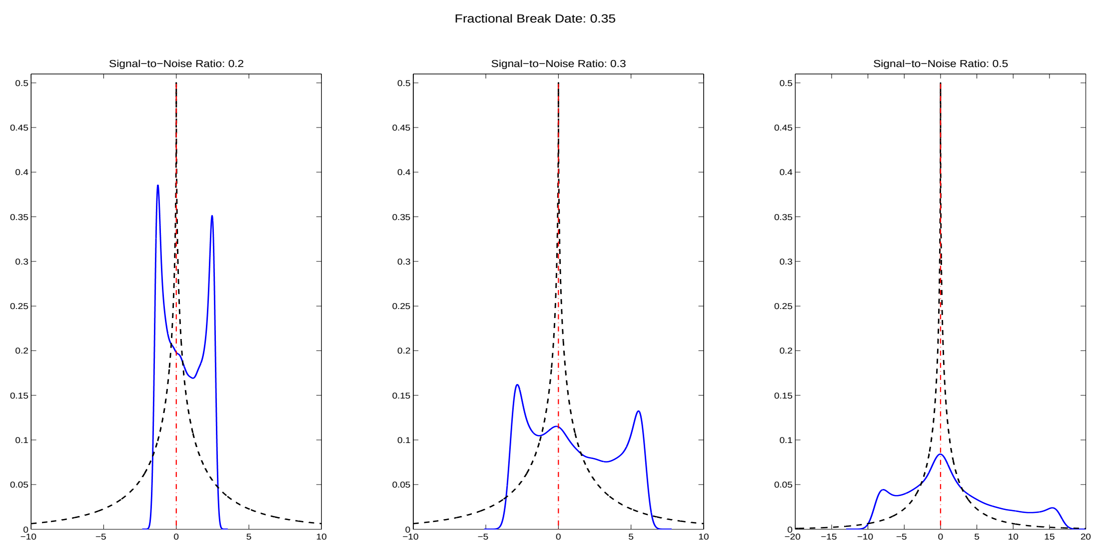

Figure S-2: The asymptotic probability density of $\rho\left(\widehat{T}_b - T_b^0\right)$ derived under a continuous record (blue solid line) and the density of Bai's (1997) asymptotic distribution (black broken line) for $\lambda_0 = 0.35$ and $\rho^2 = 0.2$, 0.3 and 0.5 (the left, middle and right panel, respectively).

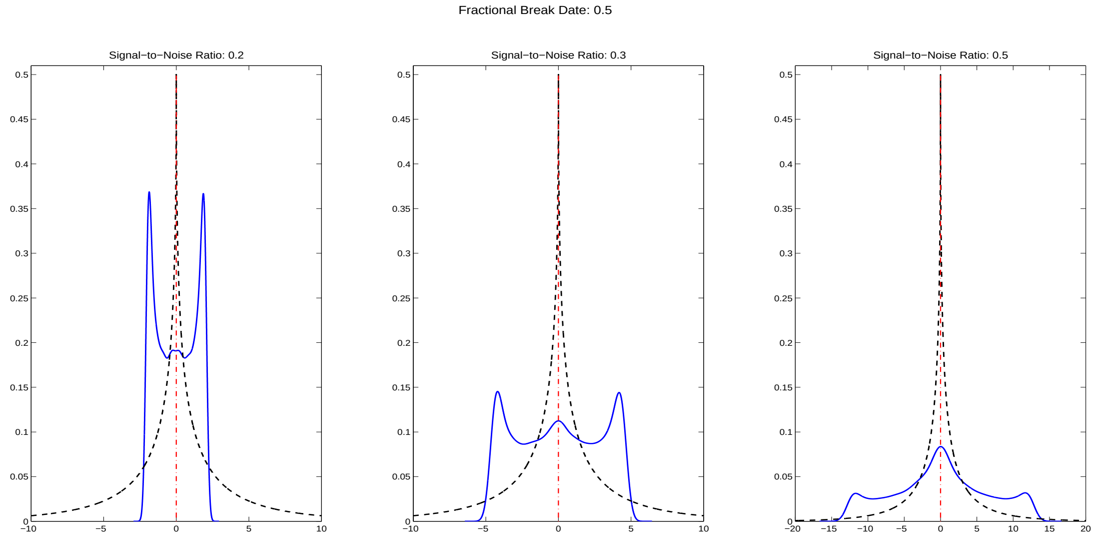

Figure S-3: The asymptotic probability density of $\rho\left(\widehat{T}_b - T_b^0\right)$ derived under a continuous record (blue solid line) and the density of Bai's (1997) asymptotic distribution (black broken line) for a true fractional break date $\lambda_0 = 0.5$ and $\rho^2 = 0.2$, 0.3 and 0.5 (the left, middle and right panel, respectively).

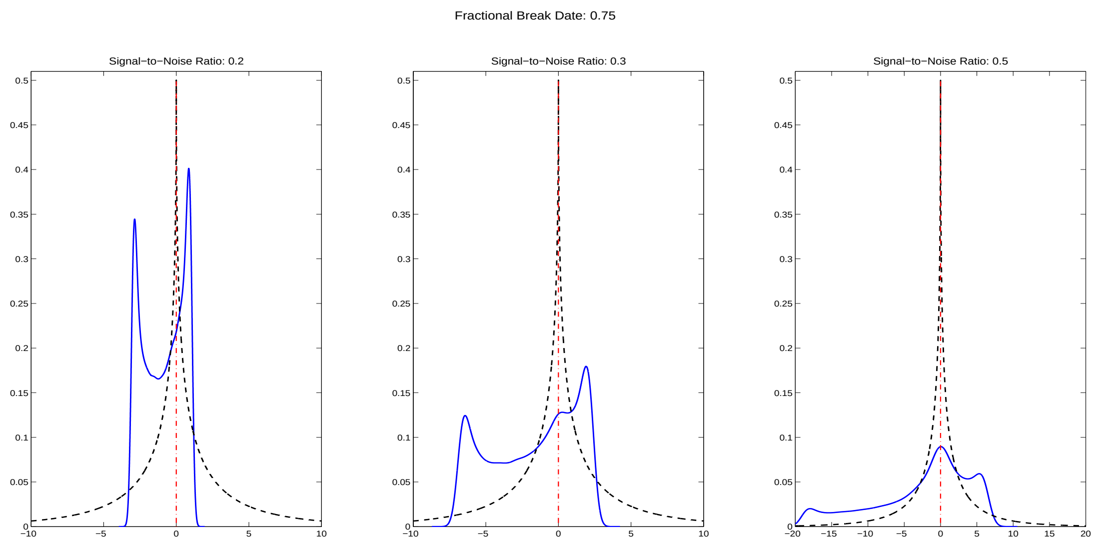

Figure S-4: The asymptotic probability density of $\rho\left(\widehat{T}_b - T_b^0\right)$ derived under a continuous record (blue solid line) and the density of Bai's (1997) asymptotic distribution (black broken line) for $\lambda_0 = 0.75$ and $\rho^2 = 0.2$, 0.3 and 0.5 (the left, middle and right panel, respectively).

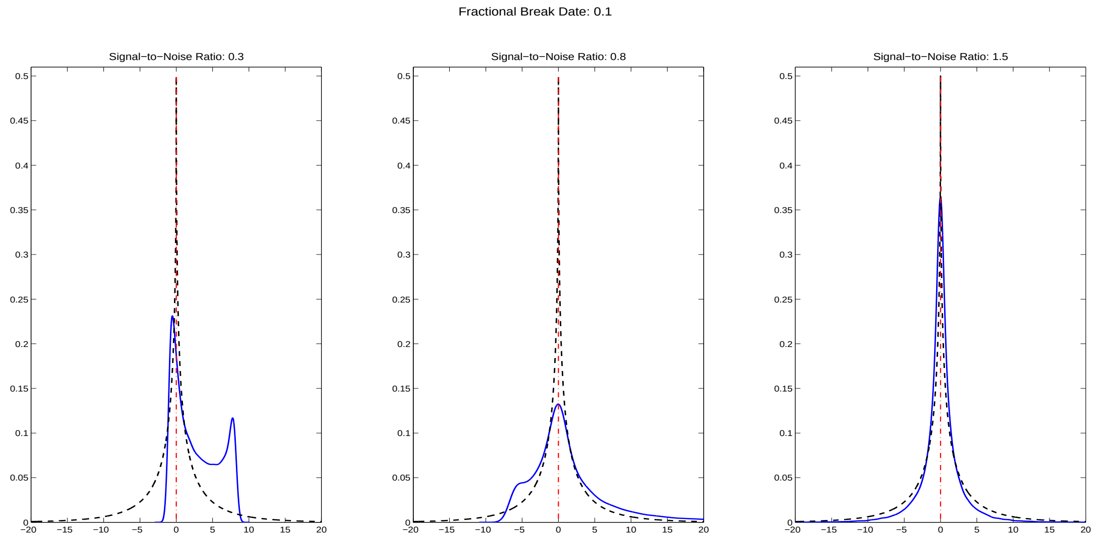

Figure S-5: The asymptotic probability density of $\rho\left(\widehat{T}_b - T_b^0\right)$ derived under a continuous record (blue solid line) and the density of of Bai's (1997) asymptotic distribution (black broken line) for $\lambda_0 = 0.1$ and $\rho^2 = 0.3$, 0.8 and 1.5 (the left, middle and right panel, respectively).

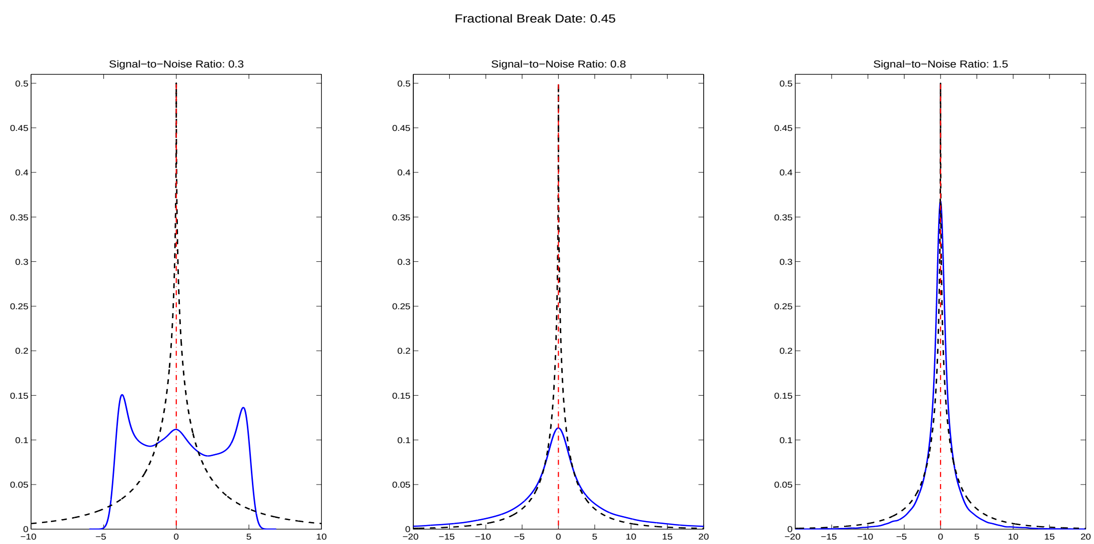

Figure S-6: The asymptotic probability density of $\rho\left(\widehat{T}_b - T_b^0\right)$ derived under a continuous record (blue solid line) and the density of Bai's (1997) asymptotic distribution (black broken line) for $\lambda_0 = 0.45$ and $\rho^2 = 0.3$, 0.8 and 1.5 (the left, middle and right panel, respectively).

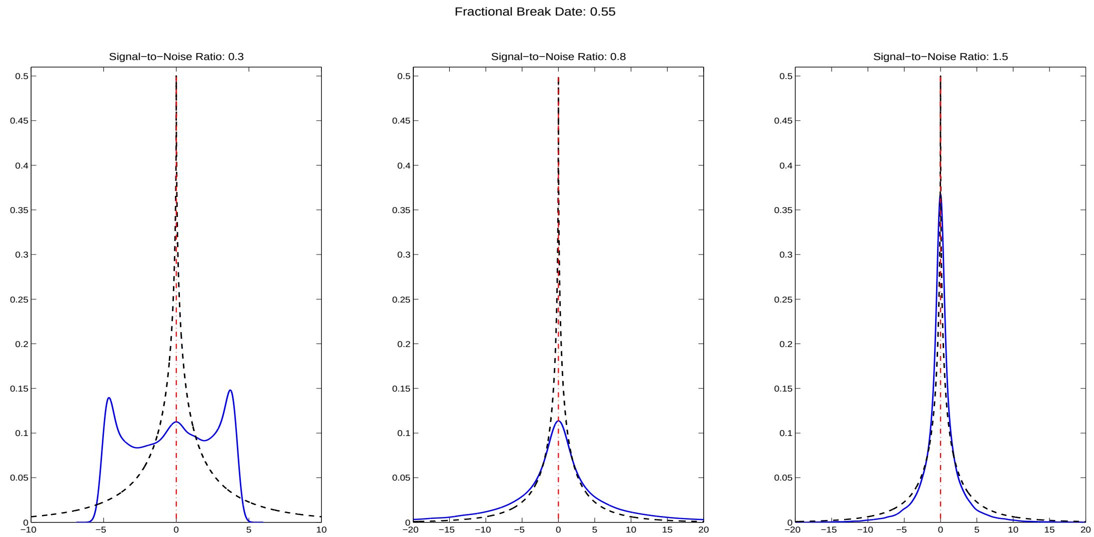

Figure S-7: The asymptotic probability density of $\rho\left(\widehat{T}_b - T_b^0\right)$ derived under a continuous record (blue solid line) and the density of Bai's (1997) asymptotic distribution (black broken line) for a true break point $\lambda_0 = 0.55$ and $\rho^2 = 0.3$, 0.8 and 1.5 (the left, middle and right panel, respectively).

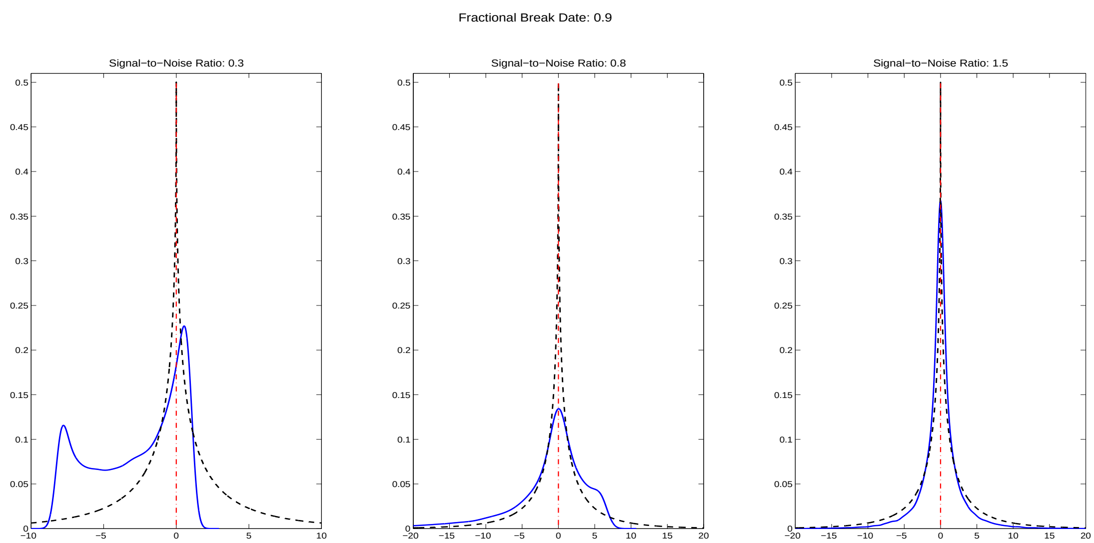

Figure S-8: The asymptotic probability density of $\rho\left(\widehat{T}_b - T_b^0\right)$ derived under a continuous record (blue solid line) and the density of Bai's (1997) asymptotic distribution (black broken line) for $\lambda_0 = 0.9$ and $\rho^2 = 0.3$, 0.8 and 1.5 (the left, middle and right panel, respectively).

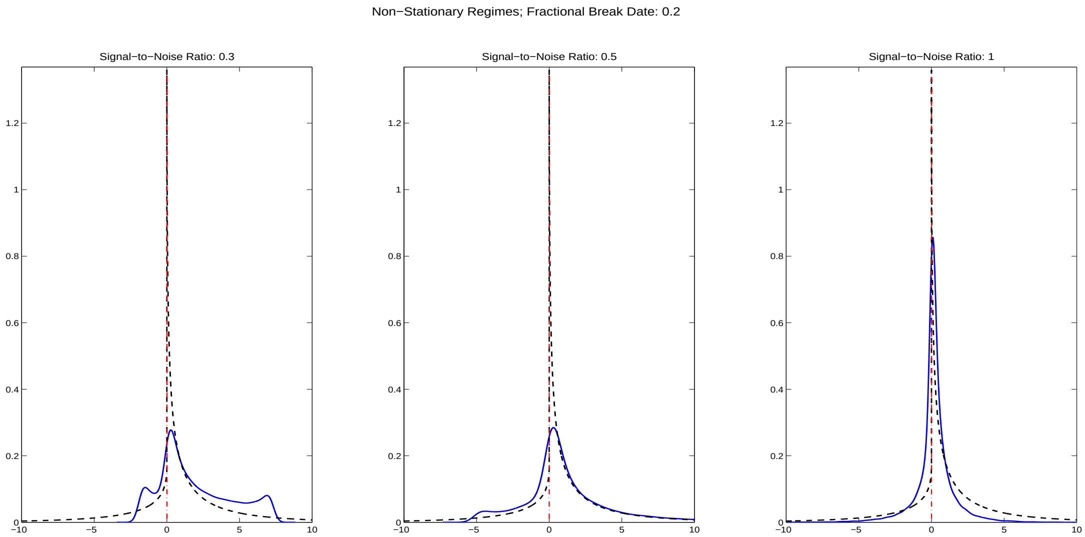

Figure S-9: The asymptotic probability density of $\rho\left(\widehat{T}_b - T_b^0\right)$ derived under a continuous record (blue solid line) and the density of Bai's (1997) asymptotic distribution (black broken line) under non-stationary regimes for $\lambda_0 = 0.2$ and $\rho^2 = 0.3$, 0.5 and 1 (the left, middle and right panel, respectively).

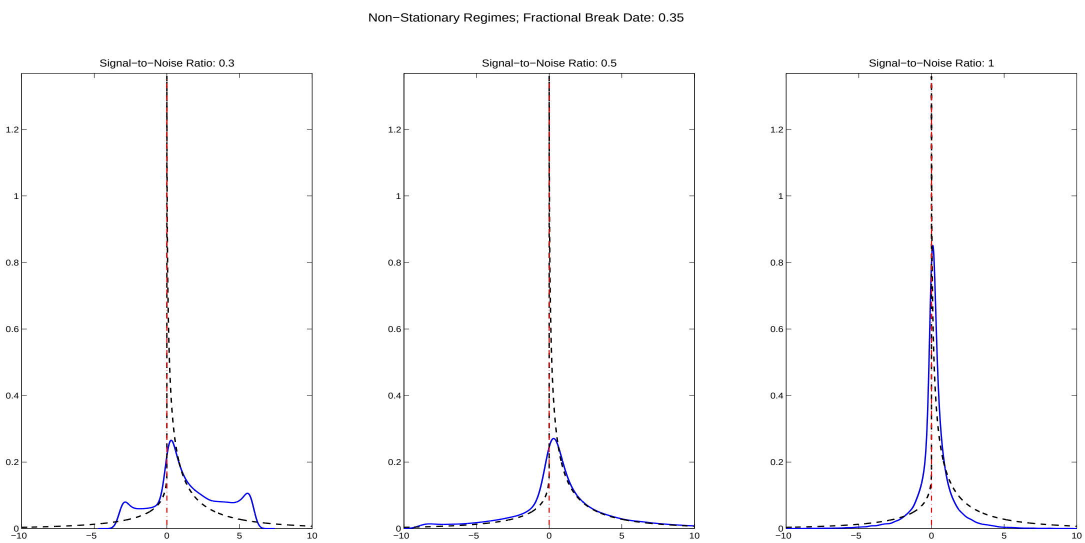

Figure S-10: The asymptotic probability density of $\rho\left(\widehat{T}_b - T_b^0\right)$ derived under a continuous record (blue solid line) and the density of Bai's (1997) asymptotic distribution (black broken line) under non-stationary regimes for $\lambda_0 = 0.2$, and $\rho^2 = 0.3$, 0.5 and 1 (the left, middle and right panel, respectively).

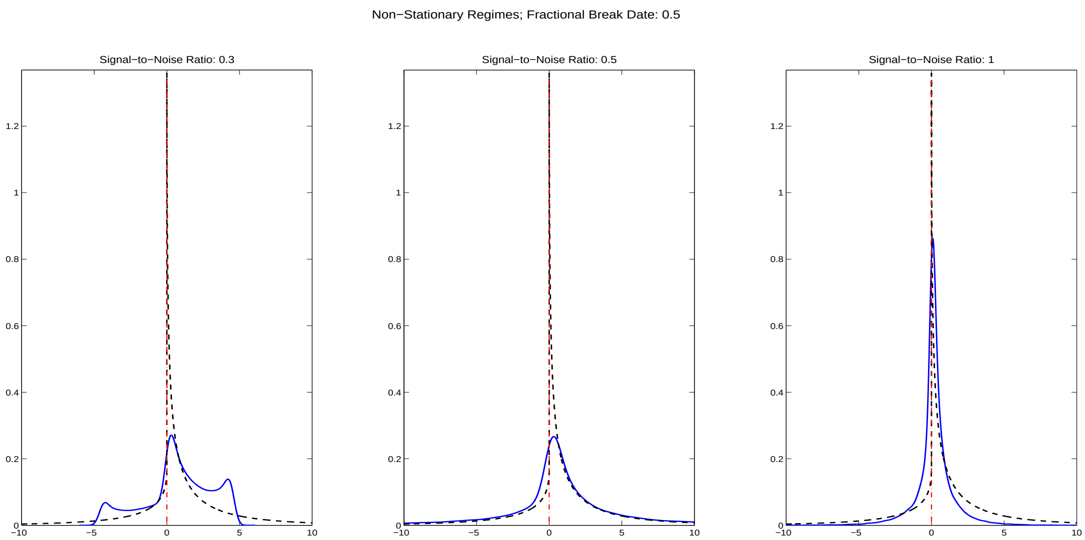

Figure S-11: The asymptotic probability density of $\rho\left(\widehat{T}_b - T_b^0\right)$ derived under a continuous record (solid line) and the density of Bai's (1997) asymptotic distribution (broken line) under non-stationary regimes for $\lambda_0 = 0.5$, and $\rho^2 = 0.3$, 0.5 and 1 (the left, middle and right panel, respectively).

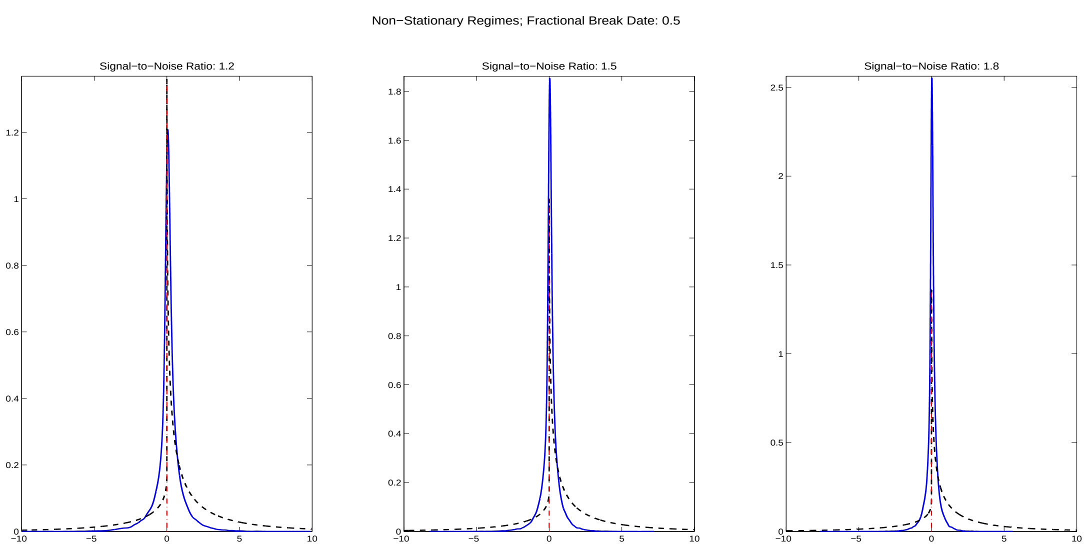

Figure S-12: The asymptotic probability density of $\rho\left(\widehat{T}_b - T_b^0\right)$ derived under a continuous record (solid line) and the density of Bai's (1997) asymptotic distribution (broken line) under non-stationary regimes for $\lambda_0 = 0.5$, and $\rho^2 = 1.2$, 1.5 and 2 (the left, middle and right panel, respectively).

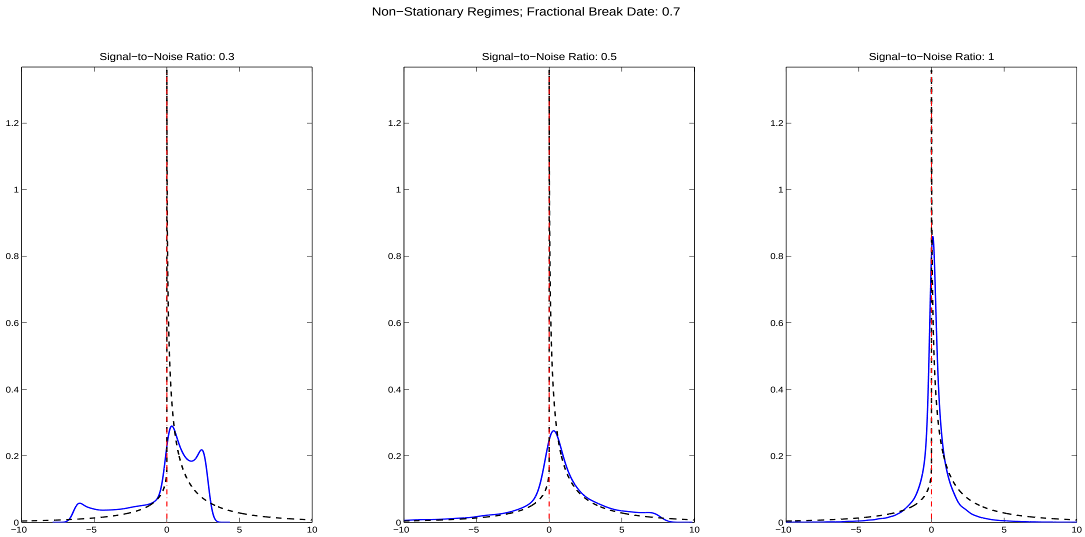

Figure S-13: The asymptotic probability density of $\rho\left(\widehat{T}_b - T_b^0\right)$ derived under a continuous record (solid line) and the density of Bai's (1997) asymptotic distribution (broken line) under non-stationary regimes for $\lambda_0 = 0.7$, and $\rho^2 = 0.3$, 0.5 and 1 (the left, middle and right panel, respectively).

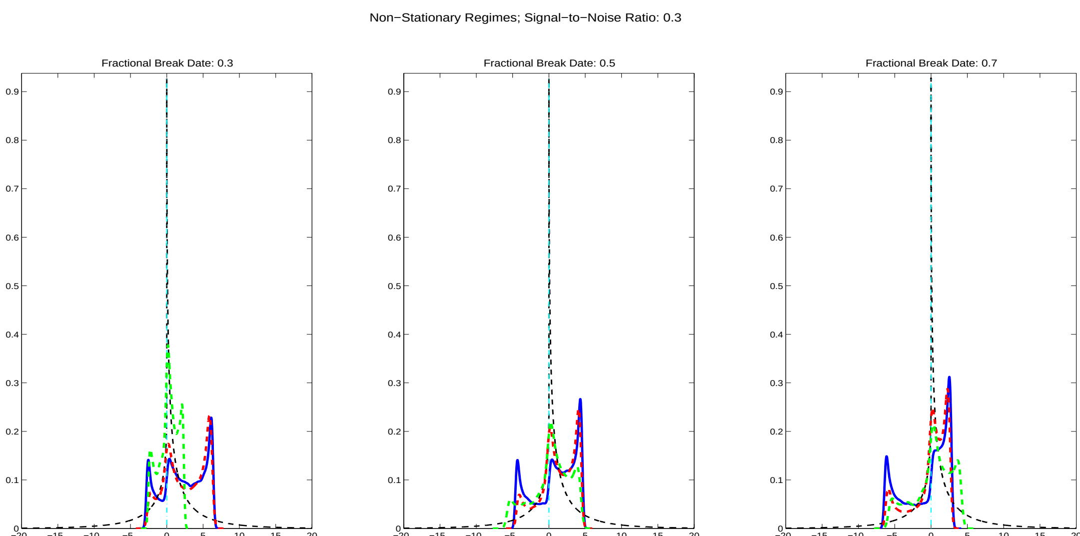

Figure S-14: The probability density of $\rho\left(\widehat{T}_b - T_b^0\right)$ for model (6.1) with break magnitude $\delta^0 = 0.3$ and $\lambda_0 =$ 0.3, 0.5 and 0.7 (the left, middle and right panel, respectively). The signal-to-noise ratio is $\delta^0/\sigma_{e,1} = \delta^0$ since $\sigma_{e,1}^2 = 1$ where $\sigma_{e,1}^2$ is the variance of the errors in the first regime. 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 of Bai (1997) and the red broken line break is the density of the finite-sample distribution.

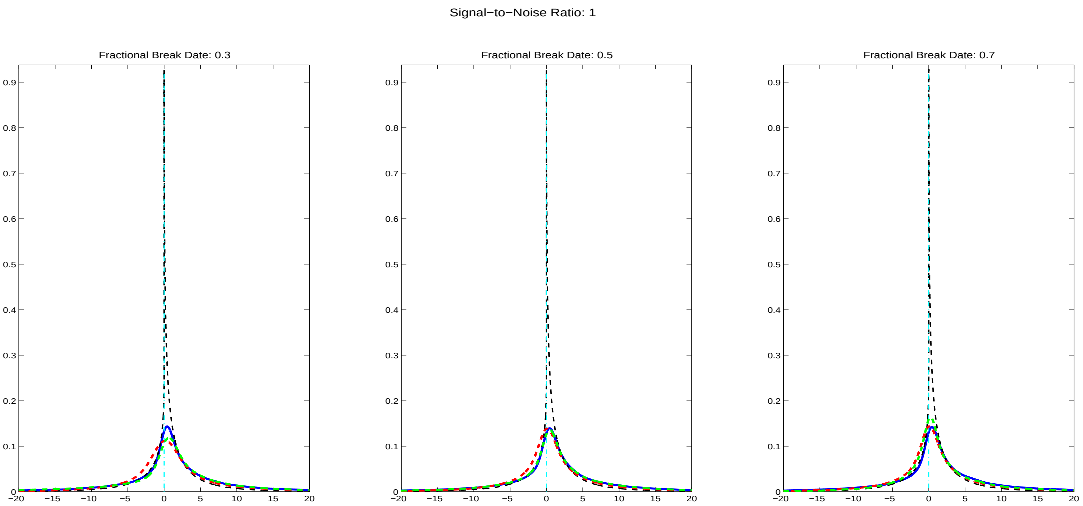

Figure S-15: The probability density of $\rho\left(\widehat{T}_b - T_b^0\right)$ for model (6.1) with $\delta^0 = 0.5$ and $\lambda_0 = 0.3$, 0.5 and 0.7 (the left, middle and right panel, respectively). The signal-to-noise ratio is $\delta^0/\sigma_{e,1} = \delta^0$ since $\sigma_{e,1}^2 = 1$ where $\sigma_{e,1}^2$ is the variance of the errors in the first regime. 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 of Bai (1997) and the red broken line break is the density of the finite-sample distribution.

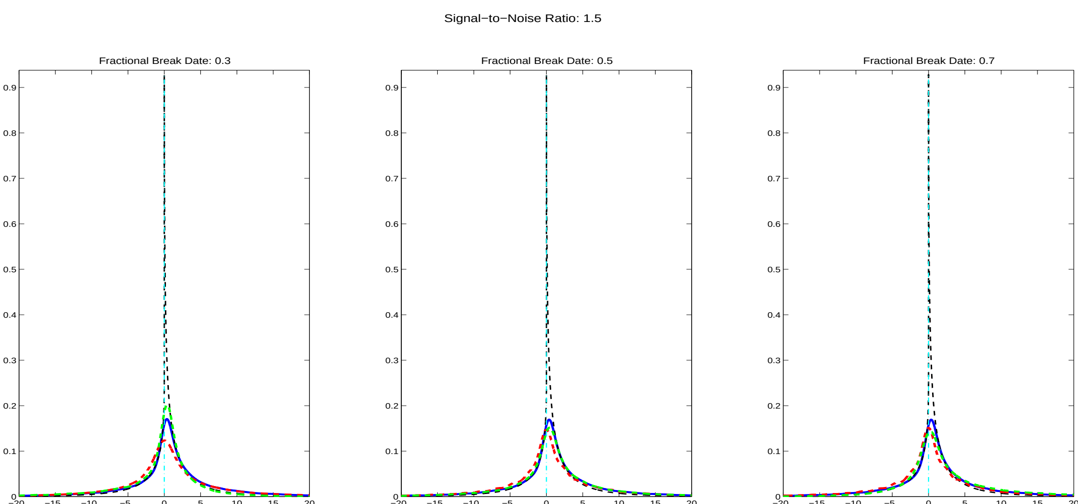

Figure S-16: The probability density of $\rho\left(\widehat{T}_b - T_b^0\right)$ for model (6.1) with $\delta^0 = 1.5$ and $\lambda_0 = 0.3$, 0.5 and 0.7 (the left, middle and right panel, respectively). The signal-to-noise ratio is $\delta^0/\sigma_{e,1} = \delta^0$ since $\sigma_{e,1}^2 = 1$ where $\sigma_{e,1}^2$ is the variance of the errors in the first regime. 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 of Bai (1997) and the red broken line break is the density of the finite-sample distribution.