Continuous Record Asymptotics for Change-Point Models∗

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

4th March 2021

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 \(T \to \infty\) 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, semimartingale.

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

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

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

In the context of a linear regression model with a single break point, we develop a continuous record asymptotic framework and inference methods for the break date. Our model is specified in continuous time but estimated with discrete-time observations using a least-squares method. We have \( T \) observations with a sampling frequency \( h \) over a fixed time horizon \([0, N]\), where \( N = Th \) denotes the time span of the data. We consider a continuous record 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.

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

We study a general change-point problem under a continuous record asymptotic framework and develop inference procedures based on the derived asymptotic distribution. We establish consistency at rate-\( T \) convergence for the least-squares estimate of the break date, assumed to occur at time \( N^0_b \). Given the fast rate of convergence, we introduce a limit theory with shrinking magnitudes of shifts and increasing variance of the residual process local to the change-point. The asymptotic distribution corresponds to the location of the extremum of a function of the (quadratic) variation of the regressors and of a Gaussian centered martingale process over some time interval. It is characterized by some notable aspects. With the time horizon \([0, N]\) fixed, we can account for the asymmetric informational content provided by the pre- and post-break sample observations, i.e., the time span and the position of the break date \( N^0_b \) 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 \( N^0_b \) and thus the distribution is asymmetric, in general. The degree of asymmetry increases as the true break point moves away from mid-sample. This holds unless the magnitude of the break is large, in which case the density is symmetric irrespective of the location of the break. This accords with simulation evidence which documents that the break point estimate is less precise.
and the coverage rates of the confidence intervals less reliable when the break is not at mid-sample
[see e.g., Chang and Perron (2018)]. When the shift magnitude is small, the density displays three
modes. As the shift magnitude increases, this tri-modality vanishes.

Our asymptotics can be seen as intermediate between the shrinkage asymptotics and more
recent approaches relying on weak identification [see e.g., Elliott, Müller, and Watson (2015)]. On
the one hand, using the usual shrinking condition of Yao (1987) and Bai (1997) for which the
break magnitude, say $\delta_T$, goes to zero at a rate slower than $O\left(T^{-1/2}\right)$ leads to underestimation
of the uncertainty about the break date. On the other hand, the weak identification condition of
Elliott and Müller (2007) for which $\delta_T$ goes to zero at a fast rate (i.e., $\delta_T = O\left(T^{-1/2}\right)$ so that
the change-point cannot be consistently estimated) leads to overstating the uncertainty. This has
opposite consequences for the confidence intervals of the break date. Confidence sets have poor
coverage probabilities when the break is small under Bai’s framework while they can be too wide
under that of Elliott and Müller (2007). In this paper, the key is not to focus our asymptotic
experiment on shrinking condition on $\delta_T$ but to make assumptions on the signal-to-noise ratio
$\delta_T/\sigma_t$, instead, where $\sigma_t$ is the volatility of the errors. We require $\delta_T$ to go to zero at a slower
rate than that of Elliott and Müller (2007)—to guarantee strong identification—and require $\sigma_t$ to
increase without bound when $t$ approaches the break date $T_0$. This offers a new characterization
of higher uncertainty without compromising strong identification and consistency of the model
parameters that are needed to conduct inference.

Despite the effort devoted to the construction of the confidence intervals for the change-point
date [see e.g., Bai and Perron (1998), Elliott and Müller (2007) and Eo and Morley (2015)], what is
still missing is a method that, for both large and small breaks, achieves both accurate coverage rates
and satisfactory average lengths. The most popular method is Bai and Perron (1998) which yields
confidence intervals that are relatively short but have good coverage only when the magnitude of
the break is very large. However, both small and large breaks are relevant for empirical work.
For example, breaks that are statistically small can be economically very large. Our simulations
calibrated to real data on GDP growth rate implies that for Bai and Perron’s confidence interval
to have accurate coverage the break size has to be unrealistically large (e.g., a break more than
4% points on yearly GDP growth). Thus, small breaks are also important in practice.

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

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

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

This paper relates to other work by the authors, namely Casini and Perron (2018, 2019a). Casini and Perron (2020a) used the asymptotic results developed in this paper and proposed a new Generalized Laplace estimator of the break date under a continuous record asymptotic framework. Casini and Perron (2020b) analyzed the Generalized Laplace method under classical asymptotics and focused on the theoretical relationship between the asymptotic distribution of frequentist and Bayesian estimators of the break point.

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

2 Model and Assumptions

We denote the transpose of a matrix $A$ by $A'$ and the $(i, j)$ elements of $A$ by $A^{(i,j)}$. We use $\| \cdot \|$ to denote the Euclidean norm of a linear space, i.e., $\| x \| = (\sum_{i=1}^{p} x_i^2)^{1/2}$ for $x \in \mathbb{R}^p$. We use $\lfloor \cdot \rfloor$ to denote the largest smaller integer function. A sequence $\{u_{kh}\}_{k=1}^{T}$ is i.i.d. (resp., i.n.d.) if the $u_{kh}$ are independent and identically (resp., non-identically) distributed. We use $\converges$, $\Rightarrow$, and $\converges^s$ 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 covariance process by \( [S, R]_t \) and their predictable counterpart by \( \langle S, R \rangle_t \). The symbol “\( \Delta \)” denotes definitional equivalence.

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

\[
Y_t = D_t \nu^0 + Z_t \delta^0_1 + e_t, \quad (t = 0, 1, \ldots, T^0) \tag{2.1}
\]
\[
Y_t = D_t \nu^0 + Z_t \delta^0_2 + e_t, \quad (t = T^0 + 1, \ldots, T),
\]

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 \( \nu^0, \delta^0_1 \) and \( \delta^0_2 \) are unknown with \( \delta^0_1 \neq \delta^0_2 \). Our main purpose is to develop inference methods for the unknown change-point date \( T^0 \) 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 \( (\Omega, \mathcal{F}, (\mathcal{F}_s)_{s \geq 0}, P) \). We observe realizations of \( \{Y_s, D_s, Z_s\} \) at discrete points of time.

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 \( \{hY_k, hD_{kh}, hZ_{kh}; k = 0, 1, \ldots, T = N/h\} \). For any process \( X \) we denote its “increments” by \( \Delta_h X_k = X_{kh} - X_{(k-1)h} \). For \( k = 1, \ldots, T \), let \( \Delta_h D_k \triangleq \mu_{D,k} h + \Delta_h M_{D,k} \) and \( \Delta_h Z_k \triangleq \mu_{Z,k} h + \Delta_h M_{Z,k} \) where the “drifts” \( \mu_{D,t} \in \mathbb{R}^q \), \( \mu_{Z,t} \in \mathbb{R}^p \) are \( \mathcal{F}_{t-h} \)-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., \( \mathbb{E} \left[ \Delta_h M_{D,t} \Delta_h M_{D,t}^\prime | \mathcal{F}_{t-h} \right] = \Sigma_{D,t-h} \Delta_t \) and \( \mathbb{E} \left[ \Delta_h M_{Z,t} \Delta_h M_{Z,t}^\prime | \mathcal{F}_{t-h} \right] = \Sigma_{Z,t-h} \Delta_t \) (\( \Delta t \) and \( h \) are used interchangeably). Let \( \lambda_0 \in (0, 1) \) denote the fractional break date (i.e., \( T^0_\lambda = \lfloor T \lambda_0 \rfloor \)). Via the Doob-Meyer Decomposition, model (2.1) can be expressed as

\[
\Delta_h Y_k \triangleq \begin{cases} (\Delta_h D_k)^\prime \nu^0 + (\Delta_h Z_k)^\prime \delta^0_1 + \Delta_h e^*_k, & (k = 1, \ldots, \lfloor T \lambda_0 \rfloor) \\ (\Delta_h D_k)^\prime \nu^0 + (\Delta_h Z_k)^\prime \delta^0_2 + \Delta_h e^*_k, & (k = \lfloor T \lambda_0 \rfloor + 1, \ldots, T) \end{cases}, \tag{2.2}
\]

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

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

**Assumption 2.1.** (i) $\mu_{D,t}$, $\mu_{Z,t}$, $\sigma_{D,t}$ and $\sigma_{Z,t}$ satisfy $P$-a.s., $\sup_{\omega \in \Omega, 0 < t \leq \tau_T} \|\mu_{D,t}(\omega)\| < \infty$, $\sup_{\omega \in \Omega, 0 < t \leq \tau_T} \|\sigma_{D,t}(\omega)\| < \infty$ 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 $0 < e_{s,t}^2 < \infty$, 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)]. Part (iii) imposes restrictions on the regressors which require them to have finite integrated covariance. We also rule out jump processes; 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.

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

An interesting issue is whether the theoretical results to be derived for model (2.2) are applicable to classical structural change models for which an increasing span of data is assumed. This requires establishing a connection between the assumptions imposed on the stochastic processes in both settings. Roughly, the classical long-span setting uses approximation results valid for weakly dependent data; e.g., ergodic and mixing processes. 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 the supplement.

**Assumption 2.3.** \( N^0_b = N \lambda_0 \) for some \( \lambda_0 \in (0, 1) \).

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

\[
\begin{align*}
y_{kh} &= x_{kh}' \beta^0 + e_{kh}, \quad \left( k = 1, \ldots, T_b^0 \right) \\
y_{kh} &= x_{kh}' \beta^0 + z_{kh}' \delta^0 + e_{kh}, \quad \left( k = T_b^0 + 1, \ldots, T \right)
\end{align*}
\]

where the true parameter \( \theta^0 = \left( \left( \beta^0 \right)', \left( \delta^0 \right)' \right)' \) 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 \( Y = (y_h, \ldots, y_{Th})', X = (x_h, \ldots, x_{Th})', e = (e_h, \ldots, e_{Th})', X_1 = (x_h, \ldots, x_{Th}, 0, \ldots, 0)', X_2 = (0, \ldots, 0, x_{(T_b+1)h}, \ldots, x_{Th})' \) and \( X_0 = (0, \ldots, 0, x_{(T_b+1)h}, \ldots, x_{Th})' \). Note that the difference between \( X_0 \) and \( X_2 \) is that the latter uses \( T_b \) rather than \( T_b^0 \). Define \( Z_1 = X_1 R, Z_2 = X_2 R \) and \( Z_0 = X R \). (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: \( \hat{T}_b^{LS} = \arg\min_{p+q \leq T_b \leq T} S_T \left( T_b \right) \). It is straightforward to show that \( \hat{T}_b^{LS} = \arg\min_{p+q \leq T_b \leq T} Q_T \left( T_b \right) \) where \( Q_T \left( T_b \right) \triangleq \tilde{\delta}'_{T_b} (Z_2'M Z_2) \tilde{\delta}_{T_b} \), \( \tilde{\delta}_{T_b} \) is the least-squares estimator of \( \delta^0 \) when regressing \( Y \) on \( X \) and \( Z_2 \), and \( M = I - X (X'X)^{-1} X' \). For brevity, we will write \( \hat{T}_b \) for \( \hat{T}_b^{LS} \) with the understanding that \( \hat{T}_b \) is a sequence indexed by \( T \) or \( h \). The estimate of the break fraction is then \( \hat{\lambda}_b = \hat{T}_b/T \).

**Remark 2.1.** In practice, applied researchers use a trimming parameter \( \pi \in (0, 1/2) \) to restrict the minimization over the subset \( [T \pi, (1 - \pi) T] \). Typical choices are \( \pi = 0.05, 0.10 \) and \( 0.15 \). While tests on structural breaks depend on \( \pi \), estimation theory does not require to specify any trimming \( \pi \) provided that a break is assumed to exist. The usual large-\( N \) shrinkage asymptotic theory is invariant to \( \pi \), a consequence of the consistency of the break fraction \( \hat{\lambda}_b \) or of the fact that the magnitude of the break is large enough asymptotically for the break to be located easily. However, if one believes that the span of the data and location of the break matter for the asymptotic properties of the estimator, then it is not difficult to see that \( \pi \) should also influence the asymptotic distribution of the estimator. Our asymptotic theory in Section 4 accommodates this property.
3 Consistency and Convergence Rate under Fixed Shifts

We now establish the consistency and convergence rate of the least-squares estimator under fixed shifts. Under the classical large-\(N\) asymptotics, related results have been established by Bai (1997) and Bai and Perron (1998). Early important results for a mean-shift appeared in Yao (1987) and Bhattacharya (1987) for an i.i.d. series, Bai (1994) for linear processes and Picard (1985) for a Gaussian autoregressive model.

Assumption 3.1. There exists an \(l_0\) such that for all \(l > l_0\), the matrices \((lh)^{-1} \sum_{k=1}^{l} x_{kh} x'_{kh}, (lh)^{-1} \sum_{k=T-l+1}^{T} x_{kh} x'_{kh}, (lh)^{-1} \sum_{k=T_0-l+1}^{T_0} x_{kh} x'_{kh}, \) and \((lh)^{-1} \sum_{k=T_0+1}^{T_0+1} x_{kh} x'_{kh}\), have minimum eigenvalues bounded away from zero in probability.

Assumption 3.2. Let \(Q_0(T_b, \theta^0) \triangleq \mathbb{E} [Q_T(T_b, \theta^0) - Q_T(T_0, \theta^0)]\). There exists a \(T_0\) such that \(Q_0(T_0, \theta^0) > \sup_{(T_b, \theta^0) \notin B} Q_0(T_b, \theta^0)\), for every open set \(B\) that contains \((T_0, \theta^0)\).

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, \(\hat{\lambda}_b \xrightarrow{P} \lambda_0\).

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 \(T\), \(P \left( T \left| \hat{\lambda}_b - \lambda_0 \right| > K \right) < \varepsilon\).

We have the same \(T\)-convergence rate as under large-\(N\) asymptotics. Let \(\theta^0 = (\beta^0)’, (\delta^0)’, (\delta_2^0)’\). The fast \(T\)-rate of convergence implies that the least-squares estimate of \(\theta^0\) is the same as when \(\lambda_0\) is known. A natural estimator for \(\theta^0\) is argmin\(_{\beta \in \mathbb{R}^p, \delta \in \mathbb{R}^p} \| Y - X\beta - \hat{Z}_2\delta \|^2\), where we use \(\hat{T}_b\) instead of \(T_b\) in the construction of \(\hat{Z}_2\). Then we have the following result, akin to an extension of corresponding results in Section 3 of Barndorf-Nielsen and Shephard (2004). As a matter of notation, let \(\Sigma^* \triangleq \{\mu_\cdot, \Sigma_\cdot, \sigma_{e,\cdot}\}_{t \geq 0}\) and denote expectation taken with respect to \(\Sigma^*\) 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 \(\Sigma^*\), \(\sqrt{T/N} \left( \hat{\beta} - \beta^0 \right), \sqrt{T/N} \left( \hat{\delta} - \delta^0 \right) \) ) \(\xrightarrow{d} \mathcal{N} (0, V)\) where \(\mathcal{N}\) denotes a mixed Gaussian distribution, with

\[
V \triangleq \nabla^{-1} \lim_{T \to \infty} T \left[ \sum_{k=1}^{T} \mathbb{E}^* (x_{kh} x'_{kh} e_{kh}^2) \sum_{k=T_0}^{T} \mathbb{E}^* (x_{kh} z_{kh}^2 e_{kh}^2) \right] \nabla^{-1},
\]

and

\[
\nabla \triangleq \lim_{T \to \infty} \left[ \sum_{k=T_0}^{T} \mathbb{E}^* (x_{kh} x'_{kh}) \sum_{k=T_0}^{T} \mathbb{E}^* (x_{kh} z_{kh}^2) \right].
\]
4 Asymptotic Distribution under a Continuous Record

We now present results about the limiting distribution of the least-squares estimate of the break date under a continuous record framework. As in the classical large-$N$ asymptotics, it depends on the exact distribution of the data and the errors for fixed break sizes [c.f., Hinkley (1971)]. This has forced researchers to consider a shrinkage asymptotic theory where the size of the shift is made local to zero as $T$ increases, an approach developed by Picard (1985) and Yao (1987). We continue with this avenue. Given the consistency result, we know that there exists some $h^*$ such that for all $h < h^*$ with high probability $\eta Th \leq \tilde{N}_h \leq (1 - \eta) Th$, for $\eta > 0$ such that $\lambda_0 \in (\eta, 1 - \eta)$. By Proposition 3.2, $\tilde{N}_h - N_h^0 = O_p(T^{-1})$, i.e., $\tilde{N}_h$ is in a shrinking neighborhood of $N_h^0$. With a certain rescaling of the objective function one can first obtain the shrinkage asymptotic distribution of Bai (1997). However, this is unsatisfactory for two reasons. First, as we show below [see also Casini and Perron (2020b; 2020a)], the shrinkage asymptotic distribution provides a poor approximation to the finite-sample distribution of the least-squares estimator. Second, the latter point also explains the poor coverage properties of the confidence intervals derived from the shrinkage asymptotic distribution when the magnitude of the break is small.

We begin with the following assumption which specifies that i) we use a shrinking condition on $\delta^0$; ii) we introduce a locally increasing variance condition on the residual process. The first is similarly used under classical large-$N$ asymptotics, while the second is new and useful in our context in order to accurately capture the relevant uncertainty in the change-point problem. We do not impose restrictions only on $\delta^0$ but also on the ratio $\delta^0/\sigma_t$ when $t$ is close to $T_h^0$. We refer to $\delta^0/\sigma_t$ as the signal-to-noise ratio. Controlling the ratio rather than just $\delta^0$ allows for a more accurate description of the uncertainty.

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

The rate $1/4$ in the conditions $\delta_h = O \left( h^{1/4} \right)$ and $\sigma_h = O \left( h^{-1/4} \right)$ is for tractability. One can show that consistency also holds for a rate faster than $1/4$, though slower than $\kappa$. However, for the derivation of the limiting distribution one needs $\delta_h/\sigma_h = O \left( h^{1/2} \right)$ and $O \left( \delta_h \right) = O \left( \sigma_h^{-1} \right)$ with $\kappa < 1/2$. The vector of scaled true parameters is $\theta_h \triangleq \left( (\beta^0)', \delta_h' \right)'$. Define

$$
\Delta_h \tilde{e}_t \triangleq \begin{cases} 
\Delta_h e_t^*, & t \notin \left( N_h^0 - \epsilon, N_h^0 + \epsilon \right) \\
\frac{1}{4} \Delta_h e_t^*, & t \in \left( N_h^0 - \epsilon, N_h^0 + \epsilon \right)
\end{cases}.
$$

(4.1)

We shall refer to $\{\Delta_h \tilde{e}_t, \mathcal{F}_t\}$ 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 $\tilde{N}_b$. Hence, in a neighborhood of $N^0_b$ in which we study the limiting behavior of the break point estimator, the rescaled criterion function is regular enough so that a feasible limit theory can be developed. The rate of convergence $T^{1-\kappa}$ is still sufficiently fast to guarantee a $\sqrt{T}$-consistent estimation of the slope parameters, as stated in the following proposition. Let $\langle Z_\Delta, Z_\Delta \rangle(v)$ be the predictable quadratic variation process of $Z_\Delta$. The process $W(v)$ is, conditionally on the $\sigma$-field $\mathcal{F}$, a two-sided centered Gaussian martingale with independent increments and variances given in Section S.B of the supplement.

**Proposition 4.1.** Under Assumption 2.1-2.3, 3.1-3.2 and 4.1, (i) $\hat{\lambda}_b \overset{P}{\rightarrow} \lambda_0$; (ii) for every $\varepsilon > 0$ there exists a $K > 0$ such that for all large $T$, $P \left( T^{1-\kappa} | \hat{\lambda}_b - \lambda_0 | > K \| \delta^0 \|^{-2} \sigma^2 \right) < \varepsilon$; and (iii) for $\kappa \in (0, 1/4]$, $\sqrt{T/N} (\hat{\lambda} - \beta^0), \sqrt{T/N} (\hat{\delta} - \delta^0)$ converge to $\mathcal{N}(0, V)$ as $T \rightarrow \infty$, with $V$ given in Proposition 3.3.

We first present a general result which shows that under Assumption 4.1 one can obtain a shrinkage asymptotic distribution similar to Bai (1997). The latter exploits the consistency of $\hat{\lambda}_b$ and the fact that mixing conditions implies that the regimes before and after $\lambda_0$ are asymptotically independent. Let $Z_\Delta \triangleq (0, \ldots, 0, z(T_b+1)_h, \ldots, z_{T^0_b h}, 0, \ldots, 0)$ if $T_b < T^0_b$ and $Z_\Delta \triangleq (0, \ldots, 0, z(T^0_b+1)_h, \ldots, z_{T_b h}, 0, \ldots, 0)$ if $T_b > T^0_b$.

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

$$T^{1-\kappa} (\hat{\lambda}_b - \lambda_0) \overset{\mathcal{L}}{\Rightarrow} \begin{cases} \text{argmax} & 2 \left( \delta^0 \right)' \mathcal{W}(v) \end{cases} (4.2)$$

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

$$T^{1-\kappa} (\hat{\lambda}_b - \lambda_0) \overset{\mathcal{L}}{\Rightarrow} \begin{cases} \text{argmax} & - \left( \delta^0 \right)' \langle Z_\Delta, Z_\Delta \rangle(v) \delta^0 + 2 \left( \delta^0 \right)' \mathcal{W}(v) \end{cases} \quad (4.3)$$

The difference between (4.2) and (4.3) is the presence of the drift (or deterministic) part $- \left( \delta^0 \right)' \langle Z_\Delta, Z_\Delta \rangle(v) \delta^0$. Without relating the magnitude of the break to the local variance condition, the order of the stochastic part dominates that of the deterministic part and so the latter vanishes asymptotically. The distributions in (4.2)-(4.3) share the same issues as Bai’s and so they do not
Hence, in order to approximate the behavior of the \( Q \) criterion function fraction and the regression coefficients so that inference is feasible. Yet, under our framework it is still possible to consistently estimate the break to shift have a first-order effect on the asymptotic analysis. For brevity, we use the notation order as \( h \) time scale simply means that we rescale the objective function in such a way that it is of higher scale, and stretch it into a time interval \( L \) and Foster (1994) and rescale “time”. For any \( L \) \( |h| \leq 0 \), \( L \) and Foster (1994) and rescale “time”. For any \( C > 0 \), let \( L_C \triangleq N_b^0 - Ch^{1-\kappa} \) and \( R_C \triangleq N_b^0 + Ch^{1-\kappa} \), where \( L_C \) and \( R_C \) are the left and right boundary points of \( D(C) \), respectively. We then have \( |R_C - L_C| = O(Ch^{1-\kappa}) \). Now, take the vanishingly small interval \( [L_C, R_C] \) on the original time scale, and stretch it into a time interval \( [T^{1-\kappa}L_C, T^{1-\kappa}R_C] \) on a new “fast time scale”. Changing time scale simply means that we rescale the objective function in such a way that it is of higher order as \( h \downarrow 0 \), i.e., it fluctuates less. This leads to an asymptotic distribution that accounts for higher uncertainty. Yet, under our framework it is still possible to consistently estimate the break fraction and the regression coefficients so that inference is feasible.

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

\[
\begin{align*}
\psi_h^{-1} \left( Q_T (T_b(v)) - Q_T \left( T_b^0 \right) \right) &=
- \delta_h \left( \sum_{k=T_b(v)+1}^{T_b^0} \frac{z_{kh} z_{kh}}{\psi_h \sqrt{\psi_h}} \right) \delta_h \pm 2 \left( \delta_h \right) \sum_{k=T_b(v)+1}^{T_b^0} \frac{z_{kh} \bar{e}_{kh}}{\sqrt{\psi_h} \sqrt{\psi_h}} + o_p \left( h^{1/2} \right),
\end{align*}
\]

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

\[
\begin{align*}
\psi_h^{-1} \left( Q_T (T_b (v)) - Q_T (T_0^b) \right) &= -\delta_h' \left( \sum_{k=T_b(v)+1}^{T_b^0} z_{\psi,kh} z'_{\psi,kh} \right) \delta_h + 2 \left( \delta^0 \right)' \sum_{k=T_b(v)+1}^{T_b^0} z_{\psi,kh} \bar{e}_{\psi,kh} + o_p \left( h^{1/2} \right),
\end{align*}
\]

with \( NT_b (v) / T = N_b (v) = N^0_b + v \), where \( z_{\psi,kh} \triangleq \frac{z_{kh}}{\sqrt{\psi_h}} \) and \( \bar{e}_{\psi,kh} \triangleq \frac{\bar{e}_{kh}}{\sqrt{\psi_h}} \). That is, because of the change of time scale, all processes in the last display are scaled up to be \( O_p (1) \) and thus behave as diffusion-like processes. On this new “fast time scale”, we have \( T^1 - \kappa R_C - T^1 - \kappa L_C = O (1) \) and \( Q_T (T_b(v)) - Q_T (T_0^b) \) is restored to be \( O_p (1) \). Observe that changing the time scale does not affect any statistic which depends on observations from \( k = 1 \) to \( k = \lfloor L_C / h \rfloor \) or from \( k = \lfloor R_C / h \rfloor \) to \( k = T \) (since these involve a positive fraction of data). However, it does affect quantities which include observations that fall in \( [T_b h, T_0^b h] \) (assuming \( T_b < T_0^b \)). In particular, on the original time scale, the processes \( \{D_t\}, \{Z_t\} \) and \( \{e_t\} \) are well-defined and scaled to be \( O_p (1) \) while \( Q_T (T_b(v)) - Q_T (T_0^b) \) (asymptotically) oscillates more rapidly than a simple diffusion-type process. On the new “fast time scale”, \( \{D_t\}, \{Z_t\} \) and \( \{e_t\} \) are not affected since they have the same order in \( [T^1 - \kappa L_C, T^1 - \kappa R_C] \) as \( h \downarrow 0 \). That is, the first conditional moments are \( O (h) \) while the corresponding moments for \( Q_T (T_b(v)) - Q_T (T_0^b) \) on \( \mathcal{D}^* (C) \) are restored to be \( O (h) \). As the continuous-time limit is approached, the rescaled criterion function \( (Q_T (T_b(v)) - Q_T (T_0^b)) / h^{1/2} \) operates on a “fast time scale” on \( \mathcal{D}^* (C) \).

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 true break date \( N^0_b \) defined on a new time scale. We first obtain the weak convergence results for the statistic \( (Q_T (T_b(v)) - Q_T (T_0^b)) / h^{1/2} \) 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

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

where \( \theta^* \triangleq (\theta_h, v) \) with \( T_b(v) \triangleq T_0^b + \lfloor v / h \rfloor \) and \( T_b(v) \) is the time index on the “fast time scale”. The normalizing factor \( \psi_h h^{1/2} \) allows us to change time scale and obtain an alternative asymptotic distribution. When \( v \) varies, \( T_b(v) \) potentially visits all integers between 1 and \( T \). Thus, on the new time scale we need to introduce the trimming parameter \( \pi \in (0, 1) \) which determines the region where \( T_b(v) \) can vary (see Remark 2.1). We have the normalizations \( T_b (v) = T \pi \) if \( T_b (v) \leq T \pi \) and \( T_b (v) = T \) \((1 - \pi) \) if \( T_b (v) \geq T \) \((1 - \pi) \). On the old time scale \( N_b (u) = N^0_b + u \) with \( v \rightarrow \psi_h^{-1} u \), so that \( N_b (u) \) is in a vanishing neighborhood of \( N^0_b \). On \( \mathcal{D}^* (C) \), we index the process \( Q_T (\theta_h, T_b(v)) - Q_T (\theta^0, T_b(v)) \) by two time subscripts: one referring to the time \( T_b \) on the
original time scale and one referring to the time elapsed since \( T_b h \) on the “fast time scale”. For simplicity, we omit the former; since the limiting distribution of the least-squares estimator will now depend on the trimming we use the notation \( \hat{\lambda}_{b,\pi} = T \hat{\lambda}_{b,\pi} \) where \( \hat{\lambda}_{b,\pi} \) is the least-squares estimator of the fractional break date associated to the fast time scale (i.e., associated to the normalizing factor \( \psi_h h^{1/2} \)). The optimization problem is not affected by the change of time scale. In fact, by Proposition 4.1, \( u = Th (\hat{\lambda} - \lambda_0) = KO_p (h^{1-\kappa}) \) on the old time scale; whereas on the new “fast time scale”, \( v = Th (\hat{\lambda}_{b,\pi} - \lambda_0) = O_p (1) \). The maximization problem is not changed because \( v/h \) can take any value in \( \mathbb{R} \). The process \( Q_T (\theta^*, T_b (v)) \rightarrow Q_T (\theta^0, T_b^0) \) is thus analyzed on a fixed horizon since \( v \) now varies over \( (N \pi - N_b^0) / \left( \| \delta^0 \|^2 \sigma^2 \right), (N (1 - \pi) - N_b^0) / \left( \| \delta^0 \|^2 \sigma^2 \right) \). Hence, redefine

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

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

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

\[
-h^{-1/2} \delta_h' \left( Z'_{\Delta} Z_{\Delta} \right) \delta_h \pm 2 h^{-1/2} \delta_h' \left( Z'_{\Delta} e \right) \overset{d}{=} - \left( \delta^0 \right)' \left( Z'_{\Delta} Z_{\Delta} \right) \delta^0 \pm 2 h^{-1/2} \left( \delta^0 \right)' h^{1/4} \left( Z'_{\Delta} h^{-1/4} e \right),
\]

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

\[
- \left( \delta^0 \right)' \left( Z'_{\Delta} Z_{\Delta} \right) \delta^0 \pm 2 \left( \delta^0 \right)' \left( h^{-1/2} Z'_{\Delta} e \right). \tag{4.5}
\]

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

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

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

\[
N \left( \hat{\lambda}_{b, \pi} - \lambda_0 \right) \xrightarrow{d} \text{argmax}_{v \in \left( N(1 - \pi) - N^0_b \right) / \left( \|\delta^0\|^2 / 2 \sigma^2 \right)} \left\{ - \left( \delta^0 \right)' \langle Z_{\Delta}, Z_{\Delta} \rangle (v) \delta^0 + 2 \left( \delta^0 \right)' W(v) \right\}.
\]

Note the differences between the results in Theorem 4.1 and in Proposition 4.2. First, on the fast time scale, \( \hat{\lambda}_{b, \pi} \) behaves as an inconsistent estimator for \( \lambda_0 \). On the original time scale \( \hat{\lambda}_b \) is not only consistent for \( \lambda_0 \) but it also enjoys a similar asymptotic distribution as in Bai (1997). Second, the asymptotic distribution of \( \hat{\lambda}_{b, \pi} \) depends on the span of the data and consequently on the trimming \( \pi \). The result in Proposition 4.2, in contrast, suggests that the span, the trimming and the location of the break are irrelevant for the limiting behavior of the estimator. This intuitively follows from the fact that under the original time scale the break date estimator is consistent. We will show in the next section that indeed the span of the data and the location of the break influence the finite-sample properties of the least-squares estimator. Consequently, Theorem 4.1 provides a more useful approximation.

Unlike Bai’s distribution, the distribution in Theorem 4.1 involves the location of the maximum of a function of the (quadratic) variation of the regressors and of a two-sided centered Gaussian martingale process over the interval \( \left( N(1 - \pi) - N^0_b \right) / \left( \|\delta^0\|^2 / 2 \sigma^2 \right) \). Notably, this domain depends on the true value of the break point \( N^0_b \) and therefore the limit distribution is asymmetric, in general. The degree of asymmetry increases as the true break point moves away from mid-sample. This holds even when the distributions of the errors and regressors are the same in the pre- and post-break regimes. The presence of the trimming confirms that the span of the (trimmed) data affects the limit distribution. It is well-known that the least-squares estimator of the break date can be sensitive to trimming [see Bai and Perron (2003) for some recommendations on the trimming choice]. Our asymptotic theory accommodates this property of the least-squares estimator while others do not.

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

We further characterize the asymptotic distribution by exploiting the ($\mathcal{F}$-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 $\mathcal{F}$.

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

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

$$
\mathcal{V}(s) = \begin{cases} 
-\frac{|s|}{2} + W^*_1(s), & \text{if } s < 0 \\
- (\beta^0)' \sum_{Z,2} \sigma^0 |s| - \frac{1}{2} + \left(\frac{(\beta^0)' \Omega_{W,2} \sigma^0}{(\beta^0)' \Omega_{W,1} \sigma^0}\right)^{1/2} W^*_2(s), & \text{if } s \geq 0.
\end{cases}
$$

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

$$
\frac{\left((\beta^0)' (Z, Z)_1 \sigma^0\right)^2}{(\beta^0)' \Omega_{W,1} \sigma^0} N \left(\hat{\lambda}_{b, \pi} - \lambda_0\right) \Rightarrow \argmax_{s \in \mathbb{R}} \left(\frac{N_s - N^0_b \left((\beta^0)' (Z, Z)_1 \delta^0\right)^2}{N_s (1 - \pi) - N^0_b \left((\beta^0)' (Z, Z)_1 \delta^0\right)^2} \mathcal{V}(s)\right). 
$$

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

5 Feasible Approximations to the Finite-Sample Distributions

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

$$\rho \triangleq \left( (\delta^0)' (Z, Z) \right)_1 \delta^0 \frac{2}{((\delta^0)' \Omega\_W \_1 \delta^0)} , \quad \xi_1 = (\delta^0)' (Z, Z) \frac{2}{\sigma^2} , \quad \xi_2 = (\delta^0)' (Z, Z) \frac{2}{\sigma^2} .$$

5.1 A Feasible Version of the Limit Distribution

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

$$Y_t = D_t \nu^0 + Z_t' \beta^0 + Z_t' \delta^0 1_{\{t > T^0_b\}} + e_t, \quad t = 1, \ldots, T, \quad (5.1)$$

where $Z_t = 0.5 Z_{t-1} + u_t$ with $u_t \sim i.i.d. \mathcal{N} (0, 1)$ independent of $e_t \sim i.i.d. \mathcal{N} (0, \sigma^2_e)$, $\sigma^2_e = 1$, $\nu^0 = 1$, $Z_0 = 0$, $D_t = 1$ for all $t$, and $T = 100$. We set $\pi = 0.05$, $T^0_b = |T \lambda_0|$ 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 $N^0_b$, $||\delta^0|| \sigma^2$, $\xi_1$, $\xi_2$ and $\rho$ at their true values. The feasible counterparts are constructed with plug-in estimates of $\xi_1$, $\xi_2$, $\rho$ and $\left( N^0_b ||\delta^0|| \sigma^2 \right) \rho$. In practice we need to use a normalization for $N$. A common choice is $N = 1$. Then $\hat{\lambda}_b = \hat{T}_b / T$ from Proposition 4.1-4.2 is a natural estimate of $\lambda_0$, using the consistency result of $\hat{\lambda}_b$ under the original time scale since the latter holds in the setting of Theorem 4.1. In practice this means that we approximate the distribution of the estimator $\hat{\lambda}_{b, \pi}$ where $\pi$ is chosen by the researcher and we plug-in the estimator $\hat{\lambda}_b$ which can be based on any trimming because of the consistency property. Here we set $\hat{\lambda}_b$ equal to the least-squares estimator based on a trimming 0.15, which is also used for the other plug-in estimates. The estimates of $\xi_1$ and $\xi_2$ are given,
respectively, by
\[
\hat{\xi}_1 = \frac{\hat{\delta}' (T - \hat{T}_b)^{-1} \sum_{k=\hat{T}_b+1}^T z_{kh} z_{kh}' \hat{\delta}}{\hat{\delta}' (\hat{T}_b)^{-1} \sum_{k=1}^T z_{kh} z_{kh}' \hat{\delta}}, \quad \hat{\xi}_2 = \frac{\hat{\delta}' (T - \hat{T}_b)^{-1} \sum_{k=\hat{T}_b+1}^T \hat{e}_{kh}^2 z_{kh} z_{kh}' \hat{\delta}}{\hat{\delta}' (\hat{T}_b)^{-1} \sum_{k=1}^T \hat{e}_{kh}^2 z_{kh} z_{kh}' \hat{\delta}},
\]
where \( \hat{\delta} \) is the least-squares estimator of \( \delta_h \) and \( \hat{e}_{kh} \) are the least-squares residuals. Use is made of the fact that the quadratic variation \( \langle Z, Z \rangle_1 \) is consistently estimated by \( \sum_{k=1}^T z_{kh} z_{kh}' / \hat{\lambda}_b \) while \( \Omega_{W,1} \) is consistently estimated by \( T \sum_{k=1}^T \hat{e}_{kh}^2 / \hat{\lambda}_b \). The method to estimate \( \lambda_0 \| \delta^0 \|^2 \sigma^{-2} \rho \) is less immediate because it involves manipulating the scaling of each of the three estimates. Let \( \hat{\vartheta} = \| \delta^0 \|^2 \sigma^{-2} \rho \). We use the following estimates for \( \hat{\vartheta} \) and \( \hat{\rho} \), respectively,
\[
\hat{\vartheta} = \rho \| \delta \|^2 \left( T^{-1} \sum_{k=1}^T \hat{e}_{kh}^2 \right)^{-1}, \quad \hat{\rho} = \frac{\left( \hat{\delta}' (\hat{T}_b)^{-1} \sum_{k=1}^T z_{kh} z_{kh}' \hat{\delta} \right)^2}{\hat{\delta}' (\hat{T}_b)^{-1} \sum_{k=1}^T \hat{e}_{kh}^2 z_{kh} z_{kh}' \hat{\delta}},
\]
Whereas we have \( \hat{\xi}_i \stackrel{p}{\rightarrow} \hat{\xi}_i \) \( (i = 1, 2) \), the corresponding approximations for \( \hat{\vartheta} \) and \( \hat{\rho} \) are given by \( \hat{\vartheta}/h \stackrel{p}{\rightarrow} \vartheta \) and \( \hat{\rho}/h \stackrel{p}{\rightarrow} \rho \). The latter results use the fact that Assumption 4.1 implies that the errors have higher volatilities and thus the squared residual \( \hat{e}_{kh}^2 \) needs to be multiplied by the factor \( h^{1/2} \). Then, \( h^{1/2} \sum_{k=1}^T \hat{e}_{kh}^2 \stackrel{p}{\rightarrow} \sigma^2 \). However, before letting \( T \rightarrow \infty \) we can apply a change in variable which results in the extra factor \( h \) canceling from the latter two estimates and the relevant quantities

Proposition 5.1. Under the conditions of Theorem 4.2, (4.7) holds when using \( \hat{\xi}_1, \hat{\xi}_2, \hat{\rho} \) and \( \hat{\vartheta} \) in place of \( \xi_1, \xi_2, \rho \) and \( \vartheta \), respectively.

The proposition implies that the limiting distribution can be simulated using plug-in estimates. This allows feasible inference about the break date. The results are presented in Figure 1-4 which also plot the classical shrinkage asymptotic distribution from Bai (1997). Here by signal-to-noise ratio we mean \( \delta^0 / \sigma_e \) which, given \( \sigma_e^2 = 1 \), equals the break size \( \delta^0 \). Unlike the shrinkage asymptotic distribution from Bai (1997), 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. The supplement shows that the quality of the approximation is good for a wide variety of models and for the case of non-stationary regimes where the distributions of the errors and regressors change across regimes.

5.2 Comparison with Other Approaches

The figures reported above have shown that the structural change problem is characterized by a high degree of uncertainty when the break magnitude is not large. The classical shrinkage asymptotics of Bai (1997) with \( \delta_T \) required to convergence to zero at a rate slower than \( O \left( T^{-1/2} \right) \) clearly
underestimates that degree of uncertainty and, as the figures show, it provides a poor approximation to the finite-sample behavior of the least-squares estimator. In Section 7 we show that this issue is responsible for the poor coverage probabilities of the confidence intervals introduced in Bai (1997) when the break magnitude is small. On the other hand, Elliott and Müller (2007) and Elliott, Müller, and Watson (2015) require $\delta T$ to go to zero at a fast rate $O(T^{-1/2})$ leading to weak identification. The latter implies that the relevant quantities in the model become inconsistent. This can be problematic for inference and indeed, their inference often suffers from the opposite problem in that confidence intervals for $\hat{T}_b$ can be too large [Casini and Perron (2020a, 2019) and Chang and Perron (2018)].

We impose conditions on the signal-to-noise ratio $\delta/\sigma$ rather than just on $\delta$. Consider a simple location model with a change $\delta$ in the mean and independent errors. What describes the uncertainty about the break in this model is the ratio $\delta/\sigma$ where $\sigma$ is the volatility of the errors. We let $\delta$ go to zero at a not too fast rate while letting $\sigma$ increase to infinity in a neighborhood of $T_b^0$. That is $(\delta_T/\sigma_t) \to 0$ at rate $O(T^{-1/2})$ in a neighborhood of $T_b^0$. Interestingly, this is the same rate Elliott and Müller used for $\delta_T \to 0$. Away from $T_b^0$, we require $(\delta_T/\sigma_t) \to 0$ at slower rate—similar to Yao (1987) and Bai (1997). The difference now is that we do not lose identification and all the parameters in the model remain consistent. Under continuous-time asymptotics, the variance of the processes is proportional to the sampling interval. This allows us to trade-off the rate of convergence at which $\lambda_0$ approaches $\lambda_0$ with the variance of the errors in a neighborhood of $T_b^0$ by letting $\sigma_t$ become large when $t$ is close to $T_b^0$ [i.e., a change of time scale as in Foster and Nelson (1994, 1996)]. This offers a new characterization of higher uncertainty without losing identification.

6 Highest Density Region-based Confidence Sets

The features of the limit and finite-sample distributions suggest that standard methods to construct confidence intervals may be inappropriate; e.g., two-sided intervals around the estimated break date based on the standard deviations of the estimate. Our approach is rather non-standard and relates to Bayesian methods. In our context, the Highest Density Region (HDR) seems the most appropriate in light of the asymmetry and, especially, the multi-modality of the distribution for small break sizes. All that is needed to implement the procedure is an estimate of the density function, using plug-in estimates as explained in Section 5. Choose some significance level $0 < \alpha < 1$ and let $\hat{P}_{T_b}$ denote the empirical counterpart of the probability distribution of $\rho N(\hat{\lambda}_{b,\pi} - \lambda_0)$ as defined in Theorem 4.2. Further, let $\hat{p}_{T_b}$ denote the density function defined by the Radon-Nikodym equation $\hat{p}_{T_b} = d\hat{P}_{T_b}/d\lambda_L$, where $\lambda_L$ 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 $(\Omega_Y, \mathcal{F}_Y, \mathbb{P}_Y)$ and taking values on the measurable space $(\mathcal{Y}, \mathcal{Y})$ is continuous and bounded. Then the $(1 - \alpha) 100\%$ Highest Density Region is a subset $S(\kappa_\alpha)$ of $\mathcal{Y}$ defined as $S(\kappa_\alpha) = \{y : f_Y(y) > \kappa_\alpha\}$ where $\kappa_\alpha$ is the largest constant that satisfies $\mathbb{P}_Y(Y \in S(\kappa_\alpha)) \geq 1 - \alpha$.

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 $T^0_b$ under a Continuous Record:** Under Assumption 2.1-2.3, 3.1-3.2 and 4.1-4.2, a $(1 - \alpha) 100\%$ confidence set for $T^0_b$ is a subset of $\{1, \ldots, T\}$ given by $C(cv_\alpha) = \{T_b \in \{1, \ldots, T\} : T_b \in S(cv_\alpha)\}$, where $S(cv_\alpha) = \{T_b : \hat{P}_{T_b} > cv_\alpha\}$ and $cv_\alpha$ satisfies $\sup_{cv_\alpha \in \mathbb{R}^+} \hat{P}_{T_b}(T_b \in S(cv_\alpha)) \geq 1 - \alpha$.

The confidence set $C(cv_\alpha)$ 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 $C(cv_\alpha)$ 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 $T^0_b$:** 1) Estimate by least-squares the break point and the regression coefficients from model (2.4); 2) Replace quantities appearing in (4.7) by consistent estimators as explained in Section 5; 3) Simulate the limiting distribution $\hat{P}_{T_b}$ from Theorem 4.2; 4) Compute the HDR of the empirical distribution $\hat{P}_{T_b}$ 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 $T^0_b$ consists in general of the union of disjoint intervals if $\hat{T}_b$ is not near the tails of the sample. One is located around the estimate of the break date, while the others are in the pre- and post-break regimes. To provide an illustration, we consider a simple example involving a single draw from a simulation experiment. Figure 5 reports the HDR of the feasible limiting distribution of $\rho(\hat{T}_{b, \pi} - T^0_b)$ for a random draw from the model described by (5.1) with parameters $\nu^0 = 1$, $\beta^0 = 0$, unit variance and autoregressive coefficient 0.6 for $Z_t$ and $\sigma^2_e = 1.2$ for the error term. We set $\lambda_0 = 0.35, 0.5$ and $\delta^0 = 0.3, 0.8, 1.5$. We use a trimming 0.15 for the plug-in estimator $\hat{T}_b$ and $\pi = 0.05$ for $\hat{T}_{b, \pi}$. As explained in Section 5.1, we could use any other trimming in place of 0.15. The results remain unchanged. 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 \( \rho (\hat{T}_{b, \pi} - T_0^b) \) is computed, it is straightforward to derive a 95% confidence set for \( T_0^b \). 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 \( \hat{T}_{b} = 70 \) and the implied 95% confidence set for \( T_0^b \) is given by \( C(\text{cv}_0.05) = \{1, \ldots, 12\} \cup \{18, \ldots, 100\} \). This includes \( T_0^b \) and the overall length is 95 observations. Table 1 reports for various method whether \( T_0^b \) is covered or not and the length of the confidence sets for this example. The length of Bai’s (1997) confidence interval is 55 but does not include \( T_0^b \). Elliott and Müller’s (2007) confidence set, denoted by \( \hat{U}_{T, \text{eq}} \) in Table 1, also does not include the true break date at the 90% confidence level, but does so at the 95% and its length is 95. Our method covers \( T_0^b \) and has a relatively short length across different \( \delta^0 \).

7 Small-Sample Properties of the HDR Confidence Sets

We now assess via simulations the finite-sample performance of the method proposed to construct confidence sets for the break date. We also make comparisons with alternative methods in the literature: Bai’s (1997) approach based on the large-\( N \) shrinkage asymptotics; Elliott and Müller’s (2007), hereafter EM, method on inverting Nyblom’s (1989) statistic; the Inverted Likelihood Ratio (ILR) approach of Eo and Morley (2015), 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 (2018) for a review and comparisons. We consider two DGPs: M1 is \( y_t = \beta^0 + \delta^0 1_{\{t > T_0^b\}} + e_t \) with \( \beta^0 = 1 \) and \( e_t \sim \text{i.i.d. } \mathcal{N}(0, 1) \); M2 is \( y_t = \delta^0 (1 - \nu^0) 1_{\{t > T_0^b\}} + \nu^0 y_{t-1} + e_t \) with \( \nu^0 = 0.8 \) and \( e_t \sim \text{i.i.d. } \mathcal{N}(0, 0.04) \). Our companion paper Casini and Perron (2021b) includes extensive simulation results. We set the significance level at \( \alpha = 0.05 \), and the break occurs at date \( \lfloor T \lambda_0 \rfloor \), where \( \lambda_0 = 0.2, 0.35, 0.5 \) and \( T = 200 \) for M1 and \( T = 100 \) for M2. The results are presented in Table 2-3. The last row in each table includes the rejection probability of a 5%-level sup-Wald test using the asymptotic critical value in Andrews (1993), which provides a measure of the magnitude of the break relative to the noise. For models with predictable processes we use the two-step procedure described in Section S.C.2.

Overall, the simulation results confirm previous findings about the performance of existing methods. Bai’s (1997) method has a coverage rate below the nominal level when the size of the break is small. Overall, our HDR method and that of EM show accurate empirical coverage rates for all DGP considered. However, EM’s method almost always displays confidence sets which are larger than those from the other approaches. Over all DGPs considered, the average length of the HDR confidence sets are 40% to 70% shorter than those obtained with EM’s approach when the size of the shift is moderate to high. The results for 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 $\delta_0^0 = 2$, the average length from the HDR method is 8.34 compared to 93.71 with EM’s. This concurs with the results in Chang and Perron (2018).

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

### 8 Conclusions

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


CASINI, A., T. DENG, and P. PERRON (2021): “Theory of Low-Frequency Contamination from Nonstationarity and Misspecification: Consequences for HAR Inference,” *arXiv pre-


——— (2021d): “Supplement to “Continuous Record Asymptotics for Change-Point Models”,” *Unpublished Manuscript, Department of Economics and Finance, University of Rome Tor Vergata*.


Perron, P. (1991): “A Continuous Time Approximation to the Unstable First-Order Autoregres-


Figure 1: The probability density of $\rho \left( \hat{T}_{h,\pi} - T^0_0 \right)$ 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 2: The probability density of $\rho \left( \hat{T}_{h,\pi} - T^0_0 \right)$ 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 3: The probability density of $\rho(\hat{T}_{b,n} - T^0_b)$ 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 4: The probability density of $\rho(\hat{T}_{b,\pi} - T^0_b)$ 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 5: Highest Density Regions (HDRs) of the feasible probability density of $\rho(\hat{T}_{b,n} - T_0^*)$ as described in Section 6. The significance level is $\alpha = 0.05$, the true break point is $\lambda_0 = 0.3$ and 0.5 (the left and right panels, respectively) and the break magnitude is $\delta_0 = 0.3$, 0.8 and 1.5 (the top, middle and bottom panels, respectively). The horizontal axis is the support of $\rho(\hat{T}_{b,n} - T_0^*)$. The red dot is the true value of the break point. The union of the black lines below the horizontal axis is the 95% HDR confidence region.
Table 1: Coverage rate and length of the confidence set for the example of Section 6

<table>
<thead>
<tr>
<th>$\delta^0$</th>
<th>Cov. $\lambda_0 = 0.35$</th>
<th>Cov. $\lambda_0 = 0.5$</th>
<th>Lgth.</th>
<th>Cov. $\lambda_0 = 0.35$</th>
<th>Cov. $\lambda_0 = 0.5$</th>
<th>Lgth.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta_0 = 0.3$</td>
<td>HDR</td>
<td>1</td>
<td>94</td>
<td>1</td>
<td>27</td>
<td>1</td>
</tr>
<tr>
<td>Bai (1997)</td>
<td>0</td>
<td>55</td>
<td>0</td>
<td>13</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>$\hat{U}_T$.neq</td>
<td>1</td>
<td>95</td>
<td>1</td>
<td>37</td>
<td>1</td>
<td>24</td>
</tr>
<tr>
<td>$\delta_0 = 0.8$</td>
<td>HDR</td>
<td>1</td>
<td>82</td>
<td>1</td>
<td>14</td>
<td>1</td>
</tr>
<tr>
<td>Bai (1997)</td>
<td>1</td>
<td>67</td>
<td>1</td>
<td>18</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>$\hat{U}_T$.neq</td>
<td>1</td>
<td>95</td>
<td>1</td>
<td>35</td>
<td>1</td>
<td>14</td>
</tr>
<tr>
<td>$\delta_0 = 1.5$</td>
<td>HDR</td>
<td>1</td>
<td>73</td>
<td>1</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>Bai (1997)</td>
<td>1</td>
<td>61</td>
<td>1</td>
<td>10</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>$\hat{U}_T$.neq</td>
<td>1</td>
<td>95</td>
<td>1</td>
<td>35</td>
<td>1</td>
<td>14</td>
</tr>
</tbody>
</table>

Coverage rate and length of the confidence sets corresponding to the example from Section 6. See also Figure 5. The significance level is $\alpha = 0.05$. Cov. and Lgth. refer to the coverage rate and average size of the confidence sets (i.e. average number of dates in the confidence sets), respectively. Cov=1 if the confidence set includes $T^0_0$ 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

<table>
<thead>
<tr>
<th></th>
<th>$\delta^0 = 0.3$</th>
<th>$\delta^0 = 0.6$</th>
<th>$\delta^0 = 1$</th>
<th>$\delta^0 = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda = 0.5$</td>
<td>HDR</td>
<td>0.938 131.35</td>
<td>0.941 69.05</td>
<td>0.943 24.02</td>
</tr>
<tr>
<td></td>
<td>Bai (1997)</td>
<td>0.842 114.24</td>
<td>0.855 51.58</td>
<td>0.911 19.75</td>
</tr>
<tr>
<td></td>
<td>$\hat{U}_{T\text{-eq}}$</td>
<td>0.946 146.23</td>
<td>0.943 76.13</td>
<td>0.948 33.45</td>
</tr>
<tr>
<td></td>
<td>ILR</td>
<td>0.954 147.25</td>
<td>0.956 78.17</td>
<td>0.965 23.87</td>
</tr>
<tr>
<td>$\lambda = 0.35$</td>
<td>HDR</td>
<td>0.939 129.02</td>
<td>0.934 63.70</td>
<td>0.939 24.23</td>
</tr>
<tr>
<td></td>
<td>Bai (1997)</td>
<td>0.855 111.45</td>
<td>0.855 49.52</td>
<td>0.914 19.39</td>
</tr>
<tr>
<td></td>
<td>$\hat{U}_{T\text{-eq}}$</td>
<td>0.933 148.74</td>
<td>0.933 75.94</td>
<td>0.933 33.08</td>
</tr>
<tr>
<td></td>
<td>ILR</td>
<td>0.946 149.81</td>
<td>0.960 77.54</td>
<td>0.964 25.63</td>
</tr>
<tr>
<td>$\lambda = 0.2$</td>
<td>HDR</td>
<td>0.941 127.29</td>
<td>0.940 62.13</td>
<td>0.942 22.06</td>
</tr>
<tr>
<td></td>
<td>Bai (1997)</td>
<td>0.863 110.12</td>
<td>0.911 53.14</td>
<td>0.931 20.20</td>
</tr>
<tr>
<td></td>
<td>$\hat{U}_{T\text{-eq}}$</td>
<td>0.950 158.98</td>
<td>0.951 97.12</td>
<td>0.950 35.26</td>
</tr>
<tr>
<td></td>
<td>ILR</td>
<td>0.956 162.32</td>
<td>0.956 96.45</td>
<td>0.965 33.31</td>
</tr>
</tbody>
</table>

The model is $y_t = \beta^0 + \delta^0 I_{\{t > |T\lambda|\}} + e_t$, $e_t \sim i.i.d. N(0, 1)$, $T = 200$. Cov. and Lgth. refer to the coverage probability and the average length of the confidence set (i.e., the average number of dates in the confidence set). sup-W refers to the rejection probability of the sup-Wald test using a 5% size with the asymptotic critical value. The number of simulations is 5,000.
Table 3: Small-sample coverage rate and length of the confidence sets for model M8

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

The model is $y_t = \delta^0 \left(1 - \nu^0\right) I_{\{t > \lfloor T\lambda_0 \rfloor\}} + \nu^0 y_{t-1} + e_t$, $e_t \sim \text{i.i.d. } \mathcal{N}(0, 0.04)$, $\nu^0 = 0.8$, $T = 100$. The notes of Table 2 apply.
Supplemental Material to

Continuous Record Asymptotics for Structural Change Models

Alessandro Casini  Pierre Perron
University of Rome Tor Vergata  Boston University

4th March 2021
First Version: 28th October 2015

Abstract

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

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

\[
-T^{1/2} \left[(\hat{\delta}') \left( \sum_{k=T_b + [v/h]}^{T_b} \hat{z}_{kh} z_{kh}' \right) \hat{\delta} - 2 \hat{\omega}_h (v) \right],
\]

where \( \hat{\omega}_h \) is a simple-size dependent sequence of Gaussian processes whose marginal distribution is characterized by \( h^{1/2} \sum_{k=T_b + [v/h]}^{T_b} \hat{z}_{kh} z_{kh}' \) which is a consistent estimate of \( \int_0^v \Omega_{Z_e,s} dW \). Thus, in the limit \( \hat{\omega}_h (v) \) has the same marginal distribution as \( W (v) \), and it follows that the limiting distribution from Theorem 4.1 can be simulated. The proposed estimator with (S.1) is valid under a continuous-record asymptotic (i.e., under Assumption 4.1 and the adoption of the “fast time scale”). It can also be shown to be valid under a fixed-shifts framework.

S.B Description of the Limiting Process in Theorem 4.1

We describe the probability setup underlying the limit process of Theorem 4.1. Note that \( Z_\Delta (v) \) is characterized by \( \hat{\delta} = \delta h^{1/2} \sum_{k=T_b + [v/h]}^{T_b} \hat{z}_{kh} z_{kh}' \) if \( T_b \leq T^0_b \). Consider an additional measurable space \( \{\Omega^*, F^*\} \) and a transition probability \( P (d\omega, d\omega^*) \) from \( \{\Omega^*, F^*\} \) into \( \{\Omega^*, F^*\} \). Next, we can define the products \( \tilde{\Omega} = \Omega \times \Omega^*, \tilde{\mathcal{F}} = \mathcal{F} \otimes \mathcal{F}^*, \tilde{\omega} (d\omega, d\omega^*) = P (d\omega) H (\omega, d\omega^*) \). This defines an extension \((\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\omega})\) of the original space \((\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, P)\). We also consider another filtration \( \{\tilde{\mathcal{F}}_t\}_{t \geq 0} \) which takes the following product form \( \tilde{\mathcal{F}}_t = \cap_{s \geq t} \mathcal{F}_s \otimes \mathcal{F}_s^* \). For the transition probability \( H \), we consider the simple form \( H (\omega, d\omega^*) = H (\omega) d\omega^* \) for some probability measure \( P^* \) on \( \{\Omega^*, \mathcal{F}^*\} \). This constitutes a “very good” product filtered extension. Next, assume that \((\Omega^*, \mathcal{F}^*, \{\mathcal{F}^*_t\}_{t \geq 0}, P^*)\) supports \( p \)-dimensional \( \{\mathcal{F}^*_t\} \)-standard independent Wiener processes \( W (v) \) \((i = 1, 2) \). Finally, we postulate the process \( \Omega_{Z_e,t} \) with entries \( \Sigma_{Z_e}^{(i,j)} \sigma_e^2 \) to admit a progressively measurable \( p \times p \) matrix-valued process \( \sigma_{Z_e} \), satisfying \( \Omega_{Z_e} = \sigma_{Z_e} \sigma_{Z_e}' \), with the property that \( \| \sigma_{Z_e} \|^2 \leq K \| \Omega_{Z_e} \| \) for some \( K < \infty \). Define the process \( \mathcal{W} (v) = \mathcal{W}^1 (v) \) if \( v \leq 0 \), and \( \mathcal{W} (v) = \mathcal{W}^2 (v) \) if \( v > 0 \), where \( \mathcal{W}^1 (v) = \int_{N^0_b}^{N^0_b + v} \sigma_{Z_e,s} dW^1_s \) and \( \mathcal{W}^2 (v) = \int_{N^0_b}^{N^0_b + v} \sigma_{Z_e,s} dW^2_s \) with components \( \mathcal{W}^{(j)} (v) = \sum_{r=1}^{\delta_{Z_e,s}} \sigma_{Z_e,r} dW^{(r)}_s \) if \( v \leq 0 \) and \( \mathcal{W}^{(j)} (v) = \sum_{r=1}^{\delta_{Z_e,s}} \sigma_{Z_e,r} dW^{(r)}_s \) if \( v > 0 \). The process \( \mathcal{W} (v) \) is well defined on the product extension \((\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}_{t \geq 0}, \tilde{\omega})\), and furthermore, conditionally on \( \mathcal{F} \), is a two-sided centered continuous Gaussian process with independent increments and (conditional) covariance

\[
\tilde{\mathcal{E}} \left( \mathcal{W}^{(u)} (v) \mathcal{W}^{(j)} (v) \right) = \Omega_{\mathcal{W}^{(u,j)}} (v) = \begin{cases} 
\Omega_{\mathcal{W}^{(u,j)}} (v), & \text{if } v \leq 0 \smallskip 
\Omega_{\mathcal{W}^{(u,j)}} (v), & \text{if } v > 0,
\end{cases}
\]

(S.1)
where $\Omega^{(u,j)}(v) = \int_{N_0}^{N_0 + v} \Omega^{(u,j)}_v ds$ and $\Omega^{(u,j)}_v = \int_{N_0}^{N_0 + v} \Omega^{(u,j)}_v ds$. Therefore, $\mathcal{W}(v)$ is conditionally on $\mathcal{F}$, a continuous martingale with “deterministic” quadratic covariation process $\Omega^\beta_{\mathcal{W}}$. The continuity of $\Omega^\beta_{\mathcal{W}}$ signifies that $\mathcal{W}(v)$ is not only conditionally Gaussian but also a.s. continuous. Precise treatment of this result can be found in Section II.7 of Jacod and Shiryaev (2003).

## S.C The Extended Model with Predictable Processes

### S.C.1 The Extended Model

The assumptions on $D_t$ and $Z_t$ specify that they are continuous semimartingale of the form (2.3). This precludes predictable processes, which are often of interest in applications; e.g., a constant and/or a lagged dependent variable. Technically, these require a separate treatment since the coefficients associated with predictable processes are not identified under a fixed-span asymptotic setting. Let $\tau_1,k = \mu_1,h + \alpha_1,h Y_{(k-1)h}$ for $k \leq \lfloor T \lambda_0 \rfloor$ and $\tau_2,k = \mu_2,h + \alpha_2,h Y_{(k-1)h}$ for $k > \lfloor T \lambda_0 \rfloor + 1, \ldots, T$. We consider the following extended model:

$$
\Delta_h Y_k \triangleq \begin{cases} 
\tau_1,k + (\Delta_h D_k)' \nu^0 + (\Delta_h Z_k)' \delta^0_{2,1} + \Delta_h e_k^*, & (k = 1, \ldots, T^0) \\
\tau_2,k + (\Delta_h D_k)' \nu^0 + (\Delta_h Z_k)' \delta^0_{2,2} + \Delta_h e_k^*, & (k = T^0 + 1, \ldots, T)
\end{cases}
$$

(S.1)

for some given initial value $Y_0$. We specify the parameters associated with the constant and the lagged dependent variable as being of higher order in $h$, or lower in $T$, as $h \downarrow 0$ so that some fixed true parameter values can be identified, i.e., $\mu_1,h \triangleq \mu_1^0 h^{-1/2}$, $\mu_2,h \triangleq \mu_2^0 h^{-1/2}$, $\mu_{\delta,h} \triangleq \mu_{\delta,h} - \mu_1,h$, $\alpha_{1,h} \triangleq \alpha_{1,h}^0 h^{-1/2}$, $\alpha_{2,h} \triangleq \alpha_{2,h}^0 h^{-1/2}$ and $\alpha_{\delta,h} \triangleq \alpha_{\delta,h}^0$. Our framework is then similar to the small-diffusion setting studied previously [cf. Ibragimov and Has’minskiï (1981), Galtchouk and Konev (2001), Laredo (1990) and Sørensen and Uchida (2003)]. With $\mu,h$ and $\alpha,h$ independent of $h$ and fixed, respectively, at the true values $\mu^0$ and $\alpha^0$, the continuous-time model is then equivalent to

$$
Y_t = Y_0 + \int_0^t \left( \mu_1^0 + \mu_3 \mathbf{1}_{\{s>N_0^\delta\}} \right) ds + \int_0^t \left( \alpha_1^0 + \alpha_3 \mathbf{1}_{\{s>N_0^\delta\}} \right) Y_s ds \\
+ D_t' \nu^0 + \int_0^t (\delta_{Z,1}^0 + \delta_{Z,1} \mathbf{1}_{\{s>N_0^\delta\}})' dZ_s + e_t^*,
$$

(S.2)

for $t \leq [0, N]$, where $Y_t = \sum_{k=1}^{[t/h]} \Delta_h Y_k$, $D_t = \sum_{k=1}^{[t/h]} \Delta_h D_k$, $Z_t = \sum_{k=1}^{[t/h]} \Delta_h Z_k$ and $e_t^* = \sum_{k=1}^{[t/h]} \Delta_h e_k^*$. 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.

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

In this section, we present asymptotic results allowing for predictable processes that include a constant and a lagged dependent variable among the regressors. Recall model (S.1). Let $\beta^0 = \left( \mu_1^0, \alpha_1^0, (\nu^0)', \left( \delta_{Z,1}^0 \right)' \right)'$, ...
\( \delta^0 = \left( \mu_2^0, \alpha_0^0, (\delta_{Z,2}^0 - \delta_{Z,1}^0) \right)' \) and \( x_{kh} = ((\mu_{1,h}/\mu_1^0) h, (\alpha_{1,h}/\alpha_1^0) Y_{(k-1)h} h, \Delta_h D_k, \Delta_h Z_k) \).

In matrix format, the model is \( Y = X\beta^0 + Z_0\delta^0 + e \), where now \( X \) is \( T \times (p + q + 2) \) and \( Z_0 = XR \), \( R \triangleq \left[(I_2, 0_{2 \times p})', (0_{(p+q) \times 2}, R)\right]' \), with \( R \) as defined in Section 2. Natural estimates of \( \beta^0 \) and \( \delta^0 \) minimize the following criterion function,

\[
\text{(S.3)} = h^{-1} \sum_{k=1}^{T} \left( (\Delta_h Y_k - \beta') \int_{(k-1)h}^{kh} X_s ds - \delta' \int_{(k-1)h}^{kh} Z_s ds \right)^2
\]

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

\[
\text{(S.3)} = h^{-1} \sum_{k=1}^{T} \left( \Delta_h Y_k - \mu_1^h h - \alpha_1^h (Y_{(k-1)h} - \nu') \Delta_h D_k
\]

\[
- \delta'_{Z,1} \Delta_h Z_k (1 \{ k \leq T_b \} - \delta'_{Z,2} \Delta_h Z_k (1 \{ k > T_b \} )^2
\]

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 \( \Delta_h \bar{Y}_k \triangleq h^{1/2} \Delta_h Y_k \) and \( \Delta_h \bar{V}_k = h^{1/2} \Delta_h V_k \left( \nu^0, \delta^0_{Z,1}, \delta^0_{Z,2} \right) \) where

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

The small-dispersion format of our model is then

\[
\Delta_h \bar{Y}_k = \left( \mu_1^0 h + \alpha_1^0 (Y_{(k-1)h} - \nu^0) \right) 1 \{ k \leq T_b^0 \}
\]

\[
+ \left( \mu_2^0 h + \alpha_2^0 (Y_{(k-1)h} - \nu^0) \right) 1 \{ k > T_b^0 \} + \Delta_h \bar{V}_k \left( \nu^0, \delta^0_{Z,1}, \delta^0_{Z,2} \right).
\]

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

\[
\frac{d\tilde{Y}_t^0}{dt} = \left( \mu_1^0 + \alpha_1^0 \tilde{Y}_t^0 \right) 1 \{ t \leq N_b^0 \} + \left( \mu_2^0 + \alpha_2^0 \tilde{Y}_t^0 \right) 1 \{ t > N_b^0 \}.
\]

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

\[
Q_T (T_b) \triangleq Q_T \left( \hat{\beta} (T_b) \right) \hat{\delta} (T_b) \hat{T}_b \right) = \hat{\delta} \left( Z_2^0 M Z_2 \right) \hat{\delta},
\]

S-3
and the LS estimates of the regression parameters are
\[ \hat{\theta} \triangleq \arg \min_{\theta \in \Theta_0} h \left( S_T \left( \beta, \delta, \hat{T}_b \right) - S_T \left( \beta^0, \delta^0, T^0_b \right) \right), \]

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

**Assumption S.C.1.** Assumption 2.3 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 S.C.1.** Consider model (S.1). Under Assumption S.C.1: (i) \( \hat{\lambda}_b \overset{D}{\to} \lambda_0 \); (ii) for every \( \varepsilon > 0 \) there exists a \( K > 0 \) such that for all large \( T \), \( P \left( T \left| \hat{\lambda}_b - \lambda_0 \right| > K \| \delta^0 \|^2 \sigma^2 \right) < \varepsilon \).

**Assumption S.C.2.** Let \( \delta_h = h^{1/4} \delta^0 \) and for \( i = 1, 2 \) \( \mu_i^h = h^{1/4} \mu_i^0 \) and \( \alpha_i^h = h^{1/4} \alpha_i^0 \), and assume that for all \( t \in (N^0_b - \epsilon, N^0_b + \epsilon) \), with \( \epsilon > 0 \) and \( T^{1-\kappa} \epsilon \to B < \infty \), \( 0 < \kappa < 1/2 \), \( \mathbb{E} \left[ (\Delta_h e_i^T)^2 \right] = \sigma_{h,i}^2 \Delta t \) \( P \)-a.s., where \( \sigma_{h,t} \triangleq \sigma_h \sigma_{e,t} \) with \( \sigma_h \triangleq h^{-1/4} \sigma \).

Furthermore, define the normalized residual \( \Delta_h \bar{e}_i \) as in (4.1). We shall derive a stable convergence in distribution for \( \bar{Q}_T \ldots \) 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 \( \sigma \)-field \( \mathcal{F} \) generated by all latent processes appearing in the model. In view of its properties, the \( \sigma \)-field \( \mathcal{F} \) admits a regular version of the \( \mathcal{F} \)-conditional probability, denoted \( H(\omega, d\omega^*) \). The limit process is then realized on the extension \( \left( \tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}_{t \geq 0}, \tilde{\mathbb{P}} \right) \) of the original filtered probability space as explained in Section S.B. We again introduce a two-sided Gaussian process \( \mathcal{W}_{Z_e} (\cdot) \) 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, \( \mathcal{W}_{Z_e} (\cdot) \) is a \( p \)-dimensional process which is \( \mathcal{F} \)-conditionally Gaussian and has \( P \)-a.s. continuous sample paths. We then have the following theorem.

**Theorem S.C.1.** Consider model (S.4). Under Assumption S.C.1-S.C.2: (i) \( \hat{\lambda}_b \overset{P}{\to} \lambda_0 \); (ii) for every \( \varepsilon > 0 \) there exists a \( K > 0 \) such that for all large \( T \), \( P \left( T^{1-\kappa} \left| \hat{\lambda}_b - \lambda_0 \right| > K \| \delta^0 \|^2 \sigma^2 \right) < \varepsilon \); (iii)
\[
N \left( \hat{\lambda}_{b, \pi} - \lambda_0 \right) \overset{\mathcal{L}}{\to} \arg \max_{v \in \left[ N \sigma - N^0_b, N \left( \sigma - N^0_b \right) \right]} \left\{ - \left( \delta^0 \right)' \Lambda (v) \delta^0 + 2 \left( \delta^0 \right)' \mathcal{W} (v) \right\},
\]

where \( \Lambda (v) \) is a process given by
\[
\Lambda (v) \triangleq \begin{cases} \Lambda_1 (v), & \text{if } v \leq 0 \\ \Lambda_2 (v), & \text{if } v > 0 \end{cases},
\]

\[
\Lambda_1 (v) \triangleq \begin{bmatrix} f_{N^0_b}^{N^0_b} ds & f_{N^0_b}^{N^0_b} \bar{Y}_s ds & 0_{1 \times p} \\ f_{N^0_b}^{N^0_b} \bar{Y}_s ds & f_{N^0_b}^{N^0_b} \bar{Y}_s^2 ds & 0_{1 \times p} \\ 0_{p \times 1} & 0_{p \times 1} & \langle Z, Z \rangle_1 (v) \end{bmatrix},
\]

\[
\Lambda_2 (v) \triangleq \begin{bmatrix} f_{N^0_b - v}^{N^0_b} ds & f_{N^0_b - v}^{N^0_b} \bar{Y}_s ds & 0_{1 \times p} \\ f_{N^0_b - v}^{N^0_b} \bar{Y}_s ds & f_{N^0_b - v}^{N^0_b} \bar{Y}_s^2 ds & 0_{1 \times p} \\ 0_{p \times 1} & 0_{p \times 1} & \langle Z, Z \rangle_1 (v) \end{bmatrix},
\]

\[
\langle \cdot, \cdot \rangle_1 (v) \triangleq \begin{bmatrix} \mathbb{E} [ \bar{Y}_s \bar{Y}_s'] - \mathbb{E} [ \bar{Y}_s'] \mathbb{E} [ \bar{Y}_s] & \mathbb{E} [ \bar{Y}_s \bar{Y}_s^2] - \mathbb{E} [ \bar{Y}_s'] \mathbb{E} [ \bar{Y}_s^2] \\ \mathbb{E} [ \bar{Y}_s \bar{Y}_s'] & \mathbb{E} [ \bar{Y}_s \bar{Y}_s^2] \end{bmatrix}.
\]
and $\Lambda_2(v)$ is defined analogously, where $(Z, Z)_1(v)$ is the $p \times p$ predictable quadratic covariation process of the pair $(Z^{(u)}_{\Delta}, Z^{(j)}_{\Delta})$, $3 \leq u, j \leq p$ and $v \leq 0$. The process $\mathcal{W}(v)$ is, conditionally on $\mathcal{G}$, a two-sided centered Gaussian martingale with independent increments.

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

$$
\mathcal{W}^{(j)}(v) = \begin{cases} 
\int_{N_0}^{N_0+v} dW_{e,s}, & j = 1, \\
\int_{N_0}^{N_0+v} \bar{Y}_dW_{e,s}, & j = 2, \\
\mathcal{W}^{(j-2)}_{Z_e}(v), & j = 3, \ldots, p + 2,
\end{cases}
$$

where $\mathcal{W}^{(i)}_{Z_e}(v) \triangleq \sum_{r=1}^{p} \int_{N_0}^{N_0+v} \sigma_{Ze,s} (\mathcal{W}^{(r)}_{s}) (i = 1, \ldots, p)$ and analogously when $v > 0$. That is, $\mathcal{W}_{Z_e}(v)$ corresponds to the process $\mathcal{W}(v)$ used for the benchmark model (and so are $W^{1*}_s$, $W^{2*}_s$ and $\Omega_{Ze,s}$ below). Its conditional covariance is given by

$$
\mathbb{E} \left( \mathcal{W}^{(u)}(v) \mathcal{W}^{(j)}(v) \right) = \Omega^{(u,j)}_{\mathcal{W}}(v) = \begin{cases} 
\Omega^{(u,j)}_{\mathcal{W},1}(v), & \text{if } v \leq 0, \\
\Omega^{(u,j)}_{\mathcal{W},2}(v), & \text{if } v > 0,
\end{cases}
$$

where $\Omega^{(u,j)}_{\mathcal{W},1}(v) = \int_{N_0}^{N_0+v} \sigma_{e,s}^2 ds$, if $u, j = 1$; $\Omega^{(u,j)}_{\mathcal{W},1}(v) = \int_{N_0}^{N_0+v} \bar{Y}_d^2 \sigma_{e,s}^2 ds$, if $u, j = 2$; $\Omega^{(u,j)}_{\mathcal{W},1}(v) = \int_{N_0}^{N_0+v} \bar{Y}_d^2 \sigma_{e,s}^2 ds$, if $1 \leq u, j \leq 2$, $u \neq j$; $\Omega^{(u,j)}_{\mathcal{W},1}(v) = 0$, if $u = 1, 2, j = 3, \ldots, p$; $\Omega^{(u,j)}_{\mathcal{W},1}(v) = \int_{N_0}^{N_0+v} \Omega^{(u-2,j-2)}_{Ze,s} ds$ if $3 \leq u, j \leq p + 2$; and similarly for $\Omega^{(u,j)}_{\mathcal{W},2}(v)$. 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 $\mathcal{W}(\cdot)$ with a Gaussian unconditional law. The case with stationary regimes is described as follows.

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

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

$$
\mathcal{Y}(s) = \begin{cases} 
-\frac{|s|}{2} + W^*_1(s), & \text{if } s < 0, \\
-\left(\frac{\delta^0}{\delta^0}' \Lambda_2 \delta^0 \right)^{1/2} + \left(\frac{\delta^0}{\delta^0}' \Omega_{\mathcal{W},1} \delta^0 \right)^{1/2} W^*_2(s), & \text{if } s \geq 0.
\end{cases}
$$

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

$$
\frac{\left(\frac{\delta^0}{\delta^0}' \Lambda_1 \delta^0 \right)^2}{\left(\frac{\delta^0}{\delta^0}' \Omega_{\mathcal{W},1} \delta^0 \right)^2} N \left(\lambda_{0,\pi} - \lambda_0\right) \Rightarrow \arg\max_{s \in \left[N_{-N_0}^0, N_{N_0}^0 \right]} \mathcal{Y}(s).
$$

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

$$
\mathcal{Y}_{sta}(s) = \begin{cases} 
-\frac{|s|}{2} + W^*_1(s), & \text{if } s < 0, \\
-\left(\frac{\delta^0}{\delta^0}' \Lambda_2 \delta^0 \right)^{1/2} + \left(\frac{\delta^0}{\delta^0}' \Lambda_1 \delta^0 \right) W^*_2(s), & \text{if } s \geq 0.
\end{cases}
$$
\[ \mathcal{Y}_{\mu,\text{sta}}(s) = \begin{cases} -\frac{|s|}{\tau} + W_1(s), & \text{if } s < 0 \\ -\frac{|s|}{\tau} + W_2(s), & \text{if } s \geq 0 \end{cases}. \]

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

\[ \left( \delta^0 \right)' \Lambda_1^0 \delta^0 \left( \lambda_{b,\pi} - \lambda_0 \right) \Rightarrow \arg\max_{s \in \left[ \frac{N_\pi - N_0^0}{\| \delta^0 \| - \frac{\delta^0}{\sigma^2}}, \frac{N(1-\pi) - N_0^0 \Lambda_1^0 \delta^0}{\| \delta^0 \| - \frac{\delta^0}{\sigma^2}} \right]} \mathcal{Y}_{\text{sta}}(s). \]

**Corollary S.C.3.** Under Assumption S.C.1-S.C.3, with \( \nu^0 = 0 \), \( \delta_{Z,i}^0 = 0 \), and \( \alpha_i^0 = 0 \) for \( i = 1, 2 \):

\[ \left( \delta^0 / \sigma \right)^2 N \left( \lambda_{b,\pi} - \lambda_0 \right) \Rightarrow \arg\max_{s \in \left[ \frac{N \pi - N_0^0}{\left( \delta^0 / \sigma \right)^2}, \frac{N(1-\pi) - N_0^0}{\left( \delta^0 / \sigma \right)^2} \right]} \mathcal{Y}_{\mu,\text{sta}}(s). \]

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

The results are similar to those in the benchmark model. However, the estimation of the regression parameters is more complicated because of the identification issues about the parameters associated with predictable processes. Nonetheless, our model specification allows us to construct feasible estimators. Given the small-dispersion specification in (S.4), we propose a two-step estimator. In fact, (S.5) essentially implies that asymptotically the evolution of the dependent variable is governed by a deterministic drift function given by \( \mu_1^0 + \alpha_1^0 \tilde{V}_t^0 \) (resp., \( \mu_2^0 + \alpha_2^0 \tilde{V}_t^0 \)) if \( t \leq N_b^0 \) (resp., \( t > N_b^0 \)). Thus, in a first step we construct least-squares estimates of \( \mu_1^0 \) and \( \alpha_1^0 \) \((i = 1, 2)\). 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 semimartingale regressors.

**Proposition S.C.2.** Under Assumption S.C.1-S.C.2, as \( h \downarrow 0 \), \( \hat{\theta} \overset{P}{\to} \theta^0 \).

The consistency of the estimate \( \hat{\theta} \) 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., \( \mu_i^0 \) and \( \alpha_i^0 \).

**S.D  Mathematical Proofs**

**S.D.1  Additional Notations**

For a matrix \( A \), the orthogonal projection matrices \( P_A, M_A \) are defined as \( P_A = A (A' A)^{-1} A' \) and \( M_A = I - P_A \), respectively. For a matrix \( A \) we use the vector-induced norm, i.e., \( \| A \| = \sup_{x \neq 0} \| A x \| / \| x \| \).
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 \([A' A]_{(i \times j)}\) (resp., \([A' A]_{(i \times j)}\)). For a random variable \( \xi \) 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 \( \overset{d}{\Rightarrow} \) signifies uniform locally in time convergence under the Skorokhod topology and recall that it implies convergence in probability. The symbol \( \overset{d}{=} \) signifies equivalence in distribution. We also use the same notations as detailed in Section 2.

### S.D.2 Preliminary Lemmas

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

\[
\begin{align*}
(Z_0' M Z_0) - (Z_0' M Z_2) (Z_2' M Z_2)^{-1} (Z_2' M Z_0) & \geq R' (X'_\Delta X_\Delta) (X'_2 X_2)^{-1} (X_0' X_0) R, \quad T_b < T_0^0 \\
(Z_0' M Z_0) - (Z_0' M Z_2) (Z_2' M Z_2)^{-1} (Z_2' M Z_0) & \geq R' (X'_\Delta X_\Delta) (X' X - X'_2 X_2)^{-1} (X' X - X_0' X_0) R, \quad T_b \geq T_0^0.
\end{align*}
\]

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

**Lemma S.D.2.** Let \( X_t \) (resp., \( \tilde{X}_t \)) be a \( q \) (resp., \( p \))-dimensional Itô continuous semimartingale defined on \([0, N]\). Let \( \Sigma_t \) denote the time \( t \) instantaneous covariation between \( X_t \) and \( \tilde{X}_t \). Choose a fixed number \( \epsilon > 0 \) and \( \varpi \) satisfying \( 1/2 - \epsilon \geq \varpi \geq \epsilon > 0 \). Further, let \( B_T \overset{\Delta}{=} [N/h - T^\varpi] \). Define the moving average of \( \Sigma_t \) as \( \Sigma_{kh} \overset{\Delta}{=} (T^\varpi h)^{-1} \int_{kh}^{(k+1)h} \xi ds \), and let \( \hat{\Sigma}_{kh} \overset{\Delta}{=} (T^\varpi h)^{-1} \sum_{i=1}^{[T^\varpi]} \Delta h X_{k+i} \Delta h \tilde{X}_{k+i} \). Then,

\[
\sup_{1 \leq k \leq B_T} \| \hat{\Sigma}_{kh} - \Sigma_{kh} \| = o_p(1).\]

Furthermore, for each \( k \) and some \( K > 0 \) with \( N - K > kh > K \),

\[
\sup_{1 \leq k \leq B_T} \| \hat{\Sigma}_{kh} - \Sigma_{kh} \| = o_p(1).
\]

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

\[
\hat{\Sigma}_{kh} - \Sigma_{kh} \overset{\Delta}{=} \frac{1}{T^\varpi h} \sum_{i=1}^{[T^\varpi]} \int_{(k+i-1)h}^{(k+i)h} \left(X_s - X_{(k+i-1)h}\right) d\tilde{X}_s + \frac{1}{T^\varpi h} \sum_{i=1}^{[T^\varpi]} \int_{(k+i-1)h}^{(k+i)h} \left(\tilde{X}_s - \tilde{X}_{(k+i-1)h}\right) dX_s.
\]

For any \( l \geq 1 \), \( \| \hat{\Sigma}_{kh} - \Sigma_{kh} \|_l \leq K_l T^{-\varpi/2} \), which follows from standard estimates for continuous Itô semimartingales. By a maximal inequality,

\[
\left\| \sup_{1 \leq k \leq B_T} \| \hat{\Sigma}_{kh} - \Sigma_{kh} \|_l \right\| \leq K_l T^{1/2} T^{-\varpi/2},
\]

S-7
which goes to zero choosing \( l > 2/\nu \). This proves the first claim. For the second, note that for \( l \geq 1 \),
\[
\sup_{T \leq t \leq T' \leq T_1 - \epsilon} \left| \sum_{k=1}^{\lfloor t/h \rfloor} \left( X_{kh}^i - X_{(k-1)h}^i \right) \left( X_{kh}^j - X_{(k-1)h}^j \right) \right| \leq K l (T(1-2\epsilon)/T - \epsilon/2),
\]
and choose \( l > (2 - 4\epsilon)/\epsilon \) to verify the claim. □

**S.D.3 Preliminary Results**

As it is customary in related contexts, we use a standard localization argument as explained in Section 1.d in Jacod and Shiryaev (2003), and thus we can replace Assumption 2.1-2.2 with the following stronger assumption.

**Assumption S.D.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 \( \{\mu \cdot t\}_{t \geq 0} \) is bounded càdlàg or càglàd.

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

**S.D.3.1 Approximate Variation, LLNs and CLTs**

We review some basic definitions about approximate covariation and more general high-frequency statistics. Given a continuous-time semimartingales \( X = (X^i)_{1 \leq i \leq d} \in \mathbb{R}^d \) 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
\[
[X^i, X^j]_t = \lim_{T \to \infty} \sum_{k=1}^{\lfloor t/h \rfloor} \left( X_{kh}^i - X_{(k-1)h}^i \right) \left( X_{kh}^j - X_{(k-1)h}^j \right),
\]
where \( \lim \) 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
\[
\sum_{k \geq 1} \left( h X_{kh}^i - h X_{(k-1)h}^i \right) \left( h X_{kh}^j - h X_{(k-1)h}^j \right),
\]
which consistently estimates the increments of the quadratic covariation \([X^i, X^j]_t\). It is an ex-post estimator of the covariability between the components of \( X \) over the time interval \([0, t]\). More precisely, as \( h \downarrow 0 \):
\[
\sum_{k \geq 1} \left( X_{kh}^i - X_{(k-1)h}^i \right) \left( X_{kh}^j - X_{(k-1)h}^j \right) \to \int_0^t \sum_{XX,s}^{(i,j)} ds,
\]
where \( \sum_{XX,s}^{(i,j)} \) is referred to as the spot (not integrated) volatility.

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

---

\(^1\)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.D.3. Under Assumption S.D.1, we have as $h \downarrow 0$, $T \to \infty$ with $N$ fixed and for any $1 \leq i, j \leq p$, 
\begin{align*}
(i) & \left| Z'_2(e)_{i,1} \right| \xrightarrow{p} 0 \text{ where } (Z'_2(e)_{i,1} = \sum_{k=1}^{T} z_{kh}^i e_{kh}; \\
(ii) & \left| Z'_0(e)_{i,1} \right| \xrightarrow{p} 0 \text{ where } (Z'_0(e)_{i,1} = \sum_{k=1}^{T} z_{kh}^i e_{kh}; \\
(iii) & \left| Z'_2 Z_2 \right|_{i,j} - \int_{(T_h+1)^k}^{T_h} \Sigma_{Z_{i,j}} ds \xrightarrow{p} 0 \text{ where } (Z'_2 Z_2)_{i,j} = \sum_{k=1}^{T} z_{kh}^i z_{kh}^j; \\
(iv) & \left| Z'_0 Z_0 \right|_{i,j} - \int_{(T_h+1)^k}^{T_h} \Sigma_{Z_{i,j}} ds \xrightarrow{p} 0 \text{ where } (Z'_0 Z_0)_{i,j} = \sum_{k=1}^{T} z_{kh}^i z_{kh}^j.
\end{align*}
For the following estimates involving $X$, we have, for any $1 \leq r \leq p$ and $1 \leq l \leq q + p$,
\begin{align*}
(vi) & \left| X(e)_{i,1} \right| \xrightarrow{p} 0 \text{ where } (X(e)_{i,1} = \sum_{k=1}^{T} x_{kh}^i e_{kh}; \\
(vii) & \left| X'_2 X_{r,l} \right| - \int_{(T_h+1)^k}^{T_h} \Sigma_{Z_{i,j}} ds \xrightarrow{p} 0 \text{ where } (X'_2 X)_{r,l} = \sum_{k=1}^{T} z_{kh}^i x_{kh}^j; \\
(viii) & \left| X'_0 X_{u,d} \right| - \int_{(T_h+1)^k}^{T_h} \Sigma_{Z_{i,j}} ds \xrightarrow{p} 0 \text{ where } (X' X)_{u,d} = \sum_{k=1}^{T} x_{kh}^i x_{kh}^j.
\end{align*}

Lemma S.D.4. Under Assumption S.D.1, we have as $h \downarrow 0$, $T \to \infty$ with $N$ fixed, $|N^0 - N_0| > \gamma > 0$ and for any $1 \leq i, j \leq p$,
(i) with $(Z'_\Delta Z_{\Delta})_{i,j} = \sum_{k=T_0}^{T_h} z_{kh}^i z_{kh}^j$ we have
\begin{align*}
\begin{cases}
\left| (Z'_\Delta Z_{\Delta})_{i,j} - \int_{(T_h+1)^k}^{T_h} \Sigma_{Z_{i,j}} ds \right| \xrightarrow{p} 0, & \text{if } T_h < T_0^0 \bigg; \\
\left| (Z'_\Delta Z_{\Delta})_{i,j} - \int_{(T_h+1)^k}^{T_h} \Sigma_{Z_{i,j}} ds \right| \xrightarrow{p} 0, & \text{if } T_h > T_0^0.
\end{cases}
\end{align*}

and for $1 \leq r \leq p + q$
(ii) with $(Z'_\Delta X_{\Delta})_{i,r} = \sum_{k=T_0}^{T_h} z_{kh}^i x_{kh}^j$ we have
\begin{align*}
\begin{cases}
\left| (Z'_\Delta X_{\Delta})_{i,r} - \int_{(T_h+1)^k}^{T_h} \Sigma_{Z_{i,r}} ds \right| \xrightarrow{p} 0, & \text{if } T_h < T_0^0 \bigg; \\
\left| (Z'_\Delta X_{\Delta})_{i,r} - \int_{(T_h+1)^k}^{T_h} \Sigma_{Z_{i,r}} ds \right| \xrightarrow{p} 0, & \text{if } T_h > T_0^0.
\end{cases}
\end{align*}

Next, we turn to the central limit theorems, they all feature a limiting process defined on an extension of the original probability space $(\Omega, \mathcal{F}, P)$. 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 $(\Omega, \mathcal{F}, P)$. We accomplish this in the usual way. We first fix the original probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, P)$. Consider an additional measurable space $(\Omega^*, \mathcal{F}^*)$ and a transition probability $Q(\omega, d\omega^*)$ from $(\Omega, \mathcal{F})$ into $(\Omega^*, \mathcal{F}^*)$. Next, we can define the products $\tilde{\Omega} = \Omega \times \Omega^*$, $\tilde{\mathcal{F}} = \mathcal{F} \otimes \mathcal{F}^*$ and $\tilde{P}(d\omega, d\omega^*) = P(d\omega)Q(\omega, d\omega^*)$. This defines the extension $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$ of the original space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, P)$. Any variable or process defined on either $\Omega$ or $\Omega^*$ is extended in the usual to $\tilde{\Omega}$ as follows: for example, let $Y_t$ be defined on $\Omega$. Then we say that $Y_t$ is extended in the usual way to $\tilde{\Omega}$ by writing $Y_t(\omega, \omega^*) = Y_t(\omega)$. Further, we identify $\mathcal{F}_t$ with $\mathcal{F}_t \otimes \{0, \Omega^*\}$, so that we have a filtered space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathcal{F}}_t \geq 0, \tilde{P})$. Finally, as for the filtration, we can consider another filtration $\{\tilde{\mathcal{F}}_t\}_{t \geq 0}$ taking the product form $\tilde{\mathcal{F}}_t = \cap_{s \geq t} \mathcal{F}_s \otimes \mathcal{F}^*_s$, where $\{\tilde{\mathcal{F}}_t\}_{t \geq 0}$ is a filtration on $(\Omega^*, \mathcal{F}^*)$. As for the transition probability $Q$ we can consider the simple form $Q(\omega, d\omega^*) = P^*(d\omega^*)$ for some probability measure on $(\Omega^*, \mathcal{F}^*)$. This defines the way a product filtered extension $(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}_{t \geq 0}, \tilde{P})$ of the original filtered space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, P)$ is constructed in this paper. Assume that the auxiliary probability space $(\Omega^*, \mathcal{F}^*, \{\mathcal{F}^*_t\}_{t \geq 0}, P^*)$ supports a $p^2$-dimensional
standard Wiener process $W^t$ which is adapted to $\{\tilde{F}_t\}$. We need some additional ingredients in order to describe the limiting process. We choose a progressively measurable “square-root” process $\sigma_\tau^*$ of the $\mathcal{M}_p^\mathbb{P}\times\mathbb{P}^2$-valued process $\tilde{S}_{Z,s}$, whose elements are given by $\tilde{S}_{(ij,kl)}^{(j,k)} = \Sigma_{Z,s}^{(i,k)} \Sigma_{Z,s}^{(j,l)}$. Due to the symmetry of $\Sigma_{Z,s}$, the matrix with entries $(\sigma_{Z,s}^*)_{ij}^{(j,k)} + (\sigma_{Z,s}^*)_{ji}^{(j,k)}) / \sqrt{2}$ is a square-root of the matrix with entries $\tilde{S}_{Z,s}^{(i,k)} + \Sigma_{Z,s}$. Then the process $\mathcal{Y}_t$ with components $\mathcal{Y}_t^{(r,j)} = 2^{-1/2} \sum_{k,l=1}^p \int_0^t \left( \sigma_{Z,s}^{(r,j,k)} + \sigma_{Z,s}^{(j,r,k)} \right) dW^t_{s}$ is, conditionally on $\tilde{F}$, a continuous Gaussian process with independent increments and (conditional) covariance $\mathbb{E} \left( \mathcal{Y}_t^{(r,j)} (v) \mathcal{Y}_t^{(k,l)} (v) | \tilde{F} \right) = \int_{0}^{h} \mathbb{E} \left( \Sigma_{Z,s}^{(r,k)} \Sigma_{Z,s}^{(j,l)} + \Sigma_{Z,s}^{(r,l)} \Sigma_{Z,s}^{(j,k)} \right) ds$, where $v \leq 0$. The CLT of interest is as follows.

**Lemma S.D.5.** Let $Z$ be a continuous Itô semimartingale satisfying Assumption S.D.1. Then, $(Nh)^{-1/2} \left( Z^t_{2} Z_{T} - (Z, Z)_{T_{n}} - (Z, Z)_{H_{n+1}} \right) \overset{\mathcal{L}}{\rightarrow} \mathcal{Y}$.

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

#### S.D.4.1 Additional Notation

In some of the proofs we face a setting in which $N_b$ is allowed to vary within a shrinking neighborhood of $N_0$. Some estimates only depend on observations in this window. For example, assume $T_b < T_0$ and consider the case where $N_b$ is allowed to vary within a shrinking neighborhood of $N_0$, this sum approximates a local window of asymptotically shrinking size. Introduce a sequence of integers $\{l_T\}$ that satisfies $l_T \to \infty$ and $l_T h \to 0$. Below when we shall establish a $T^{1-\kappa}$-rate of convergence of $\hat{\lambda}_b$ toward $\lambda_0$, considering the case where $N_b - N_0 = T^{-(1-\kappa)}$ for some $\kappa \in (0, 1/2)$. Hence, define

$$\hat{S}_X \left( T_b, T_0 \right) \triangleq \sum_{k=T_b+1}^{T_0} x_{kh} x'_{kh},$$

where now $l_T = \lfloor T^k \rceil \to \infty$ and $l_T h = h^{1-\kappa} \to 0$. Note that $1/h^{1-\kappa}$ is the rate of convergence and the interpretation for $\hat{S}_X \left( T_b, T_0 \right)$ is that it involves asymptotically an infinite number of observations falling in the shrinking (at rate $h^{1-\kappa}$) block $\{ (T_b - l_h, T_b] \}$. Other statistics involving the regressors and errors are defined similarly:

$$\hat{S}_{Xe} \left( T_b, T_0 \right) \triangleq \sum_{k=T_b+1}^{T_0} x_{kh} e_{kh} = \sum_{k=T_b+1}^{T_0} x_{kh} e_{kh},$$

$$\hat{S}_{Ze} \left( T_b, T_0 \right) \triangleq \sum_{k=T_b+1}^{T_0} z_{kh} e_{kh}.$$

Further, we let $\hat{S}_{Xe} \left( T_b, T_0 \right) \triangleq h^{-(1-\kappa)} \int_{0}^{T_0} \sum_{N_b}^{N_b} X_{e,s} ds$ and analogously when $Z$ replaces $X$. We also define

$$\hat{S}_{h,X} \left( T_b, T_0 \right) \triangleq h^{-(1-\kappa)} \int_{0}^{T_0} \sum_{k=T_b+1}^{T_0} x_{kh} x'_{kh}.$$ 

The proofs of Section 4 are first given for the case where $\mu, t$ from equation (2.3) are identically zero. In the last step, this is relaxed. Furthermore, throughout the proofs we reason conditionally on the processes.
μₜ and Σ⁰ (defined in Assumption 2.2) so that they are treated as if they were deterministic. This is a natural strategy since the processes μₜ are of higher order in h and they do not play any role for the asymptotic results [cf. Barndorf-Nielsen and Shephard (2004)].

S.D.4.2 Proof of Proposition 3.1

Proof. The concentrated sample objective function evaluated at \( \hat{T}_b \) is \( Q_T(\hat{T}_b) = \hat{\delta}_{T_b}'(Z'_2MZ_2) \hat{\delta}_{T_b} \). We have

\[
\hat{\delta}_{T_b} = (Z'_2MZ_2)^{-1}(Z'_2MY) = (Z'_2MZ_2)^{-1}(Z'_2MZ_0) \delta^0 + (Z'_2MZ_2)^{-1}Z_2Me,
\]

and \( \hat{\delta}_{T_b^0} = (Z'_0MZ_0)^{-1}(Z'_0MY) = \delta^0 + (Z'_0MZ_0)^{-1}(Z'_0Me) \). Therefore,

\[
Q_T(T_b) - Q_T(T_b^0) = \hat{\delta}_{T_b}'(Z'_2MZ_2) \hat{\delta}_{T_b} - \hat{\delta}_{T_b^0}'(Z'_0MZ_0) \hat{\delta}_{T_b^0} = \delta^0 \{ (Z'_0MZ_2)(Z'_2MZ_2)^{-1}(Z'_2MZ_0) - Z'_0MZ_0 \} \delta^0 + g_e(T_b),
\]

where

\[
g_e(T_b) = 2(\delta^0)'(Z'_0MZ_2)(Z'_2MZ_2)^{-1}Z_2Me - 2(\delta^0)'(Z'_0Me) + e'MZ_2(Z'_2MZ_2)^{-1}Z_2Me - e'MZ_0(Z'_0MZ_0)^{-1}Z'_0Me.
\]

Denote

\[
X_\Delta \triangleq X_2 - X_0 = (0, \ldots, 0, x(T_h+1), \ldots, x(T_hh), 0, \ldots, ),
\]

\[
X_\Delta \triangleq - (X_2 - X_0) = (0, \ldots, 0, x(T_h+1), \ldots, x(T_hh), 0, \ldots, ),
\]

\[
X_\Delta \triangleq 0,
\]

for \( T_b < T_b^0 \)

for \( T_b > T_b^0 \)

for \( T_b = T_b^0 \).

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

\[
r(T_b) \triangleq \frac{(\delta^0)'\{ (Z'_0MZ_2) - (Z'_0MZ_0)(Z'_2MZ_2)^{-1}(Z'_2MZ_0) \} \delta^0}{|T_b - T_b^0|}.
\]

We arbitrarily define \( r(T_b) = (\delta^0)' \delta^0 \) when \( T_b = T_b^0 \). We write (S.7) as

\[
Q_T(T_b) - Q_T(T_b^0) = -|T_b - T_b^0|r(T_b) + g_e(T_b), \quad \text{for all } T_b.
\]

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

\[
P \left( |\hat{\lambda}_b - \lambda_0| > K \right) = P \left( |\hat{T}_b - T_b^0| > TK \right)
\]

\[
\leq P \left( \sup_{|T_b - T_b^0| > TK} |g_e(T_b)| \geq \inf_{|T_b - T_b^0| > TK} |T_b - T_b^0|r(T_b) \right)
\]

\[
\leq P \left( \sup_{p \leq T_b \leq T - p} |g_e(T_b)| \geq TK \inf_{|T_b - T_b^0| > TK} r(T_b) \right)
\]

S.11
\[ P \left( r_T^{-1} \sup_{p \leq T_b \leq T-p} |g_e(T_b)| \geq K \right), \]

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

\[ \sup_{p \leq T_b \leq T-p} |g_e(T_b)| = o_p(1). \quad (S.15) \]

The first term of \( g_e(T_b) \) is

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

Lemma S.D.5 implies that for any \( 1 \leq i \leq p, (Z_2e)_i, / \sqrt{h} = O_p(1) \) and for any \( 1 \leq i \leq q + p, (Xe)_i, / \sqrt{h} = O_p(1) \). These hold because they both involve a positive fraction of the data. Furthermore, from Lemma S.D.3, we also have that \( Z_2'MZ_2 \) and \( Z_2'MZ_2 \) 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.D.3. By Assumption (2.1)-(iii) \( (Z_2'MZ_2)^{-1/2} Z_2Me \) is \( O_p(1) \) 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 \( \left| \lambda_b - \lambda_0 \right| > K \). Therefore for some \( B > 0 \), these arguments combined with Lemma S.D.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 \( \varepsilon \). This concludes the proof since \( \varepsilon > 0 \) was arbitrarily chosen. \( \square \)

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

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

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

When multiplied by \( T \), we have

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

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

\[ Tr(T_b) \geq \left( \delta^0 \right)' R' X_\Delta X_\Delta / \left(N_b - N_b \right) \left( X_2'X_2 \right)^{-1} \left( X_0'X_0 \right) R \delta^0, \]

and by standard estimates for Itô semimartingales, \( X_\Delta X_\Delta = O_p(1) \) (i.e., use the Burkhölder-Davis-Gundy inequality and recalling that \( \left| \lambda_b - N_b^0 \right| > BN \)). Hence, we conclude that \( Tr(T_b) \geq \left( \delta^0 \right)' R' O_p(1) / N O_p(1) R \delta^0 \geq \kappa > 0 \), where \( \kappa \) is some positive constant. The last inequality follows whenever \( X_\Delta X_\Delta \) is positive defi-
that this implies \( R'X'_\Delta X_\Delta (X'_0X_0)^{-1}X'_0X_0 \) \( R \) can be rewritten as \( R' \left[ (X'_0X_0)^{-1} + (X'_\Delta X_\Delta)^{-1} \right] R \). According to Lemma S.D.3, \( X'_2X_2 \) is \( O_p(1) \). The same argument applies to \( X'_0X_0 \), which together with the fact that \( R \) has full common rank in turn implies that we can choose a \( B > 0 \) such that \( r_B = \inf_{|T_b - T_b^0| > B} Tr (T_b) \) satisfies \( P (r_B \geq \kappa) \leq 1 - \varepsilon \). The case with \( T_b > T_b^0 \) is similar and omitted. □

S.D.4.3 Proof of Proposition 3.2

Proof. Given the consistency result, one can restrict attention to the local behavior of the objective function for those values of \( T_b \) in \( B_T \equiv \{ T_b : T_\eta \leq T_b \leq T (1 - \eta) \} \), where \( \eta > 0 \) satisfies \( \eta \leq \lambda_0 \leq 1 - \eta \). By Proposition 3.1, the estimator \( \hat{T}_b \) will visit the set \( B \) with large probability as \( T \to \infty \). That is, for any \( \varepsilon > 0 \), \( P \left( \hat{T}_b \notin B_T \right) < \varepsilon \) for sufficiently large \( T \). We show that for large \( T \), \( \hat{T}_b \) eventually falls in the set \( B_{K,T} \equiv \{ T_b : |N_b - N_b^0| \leq KT^{-1} \} \), for some \( K > 0 \). For any \( K > 0 \), define the intersection of \( B_T \) and the complement of \( B_{K,T} \) by \( D_{K,T} \equiv \{ T_b : N_\eta \leq N_b \leq N (1 - \eta), |N_b - N_b^0| > KT^{-1} \} \). Notice that

\[
\left\{ |\hat{\lambda}_b - \lambda_0| > KT^{-1} \right\} = \\
\left\{ |\hat{\lambda}_b - \lambda_0| > KT^{-1} \cap \hat{\lambda}_b \in (\eta, 1 - \eta) \right\} \\
\cup \left\{ |\hat{\lambda}_b - \lambda_0| > KT^{-1} \cap \hat{\lambda}_b \notin (\eta, 1 - \eta) \right\} \\
\subseteq \left\{ |\hat{\lambda}_b - \lambda_0| > K \left( T^{-1} \right) \cap \hat{\lambda}_b \in (\eta, 1 - \eta) \right\} \cup \left\{ \hat{\lambda}_b \notin (\eta, 1 - \eta) \right\},
\]

and so

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

and for large \( T \),

\[
P \left( |\hat{\lambda}_b - \lambda_0| > KT^{-1} \right) \leq \varepsilon \quad \text{and} \quad P \left( |\hat{T}_b - T_b^0| > K \cap \hat{\lambda}_b \in (\eta, 1 - \eta) \right)
\]

\[
\leq \varepsilon + P \left( \sup_{T_b \in D_{K,T}} Q_T (T_b) \geq Q_T (T_b^0) \right).
\]

Therefore it is enough to show that the second term above is negligible as \( h \downarrow 0 \). Suppose \( T_b < T_b^0 \). Since \( \hat{T}_b = \arg \max Q_T (T_b) \), it is enough to show that \( P \left( \sup_{T_b \in D_{K,T}} Q_T (T_b) \geq Q_T (T_b^0) \right) < \varepsilon \). Note that this implies \( |T_b - T_b^0| > KN^{-1} \). Therefore, we have to deal with a setting where the time span in \( D_{K,T} \) 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

\[
P \left( \sup_{T_b \in D_{K,T}} h^{-1} g_e (T_b) / |T_b - T_b^0| \geq \inf_{T_b \in D_{K,T}} h^{-1} r (T_b) \right) < \varepsilon.
\]

By Lemma S.D.1,

\[
\inf_{T_b \in D_{K,T}} r (T_b) \geq \inf_{T_b \in D_{K,T}} \left( g^0 \right)' R' X'_\Delta X_\Delta (X'_2X_2)^{-1} (X'_0X_0) R g^0.
\]

S-13
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 \(i, j \leq q + p\), we have \(\mathbb{E} \left[ x_{kh}^{(i)} x_{kh}^{(j)} \right] \rightarrow 0\) as \(n \rightarrow \infty\), 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 \((h |T_b - T_b^0|)^{-1} \sum_{k=1}^{T_b^0} x_{kh} x_{kh}' = A > 0\) for all \(T_b \in D_{K,T}\).

This shows that \(\inf_{T_b \in D_{K,T}} h^{-1} r(T_b)\) 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,

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

for some \(B > 0\). Consider the terms of \(g_e(T_b)\) in (S.11). When \(T_b \in D_{K,T}\), \(Z_2\) involves at least a positive fraction \(N_\eta\) of the data. From Lemma S.D.3, as \(h \downarrow 0\), it follows that

\[
h^{-1} \left( T_b^0 - T_b \right)^{-1} \varepsilon' M Z_2 (Z_2' M Z_2)^{-1} Z_2 M e
= \left( T_b^0 - T_b \right)^{-1} h^{-1} O_p \left( h^{1/2} \right) O_p \left( h^{1/2} \right) = O_p \left( \frac{1}{T_b^0 - T_b} \right),
\]

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

\[
\left( \delta^0 \right)' (Z_0' M Z_2) (Z_0' M Z_2)^{-1} Z_2 M e
= \left( \delta^0 \right)' \left( Z_0' M Z_2 \pm Z_\Delta \right) (Z_0' M Z_2)^{-1} Z_2 M e
\]

\[
= \left( \delta^0 \right)' Z_0' M e \pm \left( \delta^0 \right)' Z_\Delta M e
\]

\[
\pm \left( \delta^0 \right)' \left( Z_\Delta' M Z_2 \right) (Z_2' M Z_2)^{-1} Z_2 M e, \tag{S.18}
\]

from which it follows that

\[
\left| 2 \left( \delta^0 \right)' (Z_0' M Z_2) (Z_0' M Z_2)^{-1} Z_2 M e - 2 \left( \delta^0 \right)' (Z_0' M e) \right|
= \left| \left( \delta^0 \right)' Z_\Delta M e \right| + \left| \left( \delta^0 \right)' \left( Z_\Delta' M Z_2 \right) (Z_0' M Z_0)^{-1} (Z_2 M e) \right| \tag{S.19}
\]

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

\[
h^{-1} \left( \delta^0 \right)' (Z_\Delta' M Z_2) = h^{-1} \left( \delta^0 \right)' (Z_\Delta' Z_2) = h^{-1} \left( \delta^0 \right)' (Z_\Delta' X_\Delta (X' X)^{-1} X' Z_2)
= \left( \delta^0 \right)' \left( \frac{Z_\Delta' X_\Delta}{h} \right) (X' X)^{-1} X' Z_2 \right) .
\]

Consider \(Z_\Delta' Z_\Delta\). By the same reasoning as above, whenever \(T_b \in D_{K,T}\), \((Z_\Delta' Z_\Delta) / h (T_b^0 - T_b) = O_p(1)\) for \(K\) large enough. The term \(Z_\Delta' X_\Delta / h (T_b^0 - T_b)\) is also \(O_p(1)\) uniformly. Thus, it follows from Lemma
We can write
\[ \frac{Z' \Delta Me}{(T_b - T_b) h} = \frac{1}{(T_b^0 - T_b) h} \sum_{k=T_b+1}^{T_b^0} z_{kh} \epsilon_{kh} \]

and have zero mean and finite second moments. Hence, by the Hájek-Rényi inequality [see S.D.5 that the second term of (S.19) is \( O_p\left(h^{1/2}\right) \). Next, note that \( Z' \Delta Me = Z' \Delta e - Z' \Delta X (X'X)^{-1} X'e \). We can write
\[ \frac{Z' \Delta Me}{(T_b - T_b) h} = \frac{1}{(T_b^0 - T_b) h} \sum_{k=T_b+1}^{T_b^0} z_{kh} \epsilon_{kh} \]

\[ \quad - \frac{1}{(T_b^0 - T_b) h} \left( \sum_{k=T_b+1}^{T_b^0} z_{kh} x'_{kh} \right) (X'X)^{-1} (X'e). \]

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

(S.20)

when \( K \) is large. Thus,
\[ \frac{1}{(T_b^0 - T_b) h} g_e(T_b) = \frac{1}{(T_b^0 - T_b) h} \left( \delta^0 \right)' 2Z' \Delta e + \frac{O_p(1)}{T_b^0 - T_b} + O_p\left(h^{1/2}\right). \]

(S.21)

We can now prove (S.17) using (S.21). To this end, we need a \( K > 0 \), such that
\[ P \left( \sup_{T_b \in \mathbf{D}_{K,T}} \left\| \left( \delta^0 \right)' \frac{2}{h} \frac{1}{T_b^0 - T_b} \sum_{k=T_b+1}^{T_b^0} z_{kh} \epsilon_{kh} \right\| > B \frac{4N}{4N} \right) \]

\[ \leq P \left( \sup_{T_b \leq T_b^0 - K N^{-1}} \left\| \frac{1}{h} \frac{1}{T_b^0 - T_b} \sum_{k=T_b+1}^{T_b^0} z_{kh} \epsilon_{kh} \right\| > B \frac{8N \delta^0 \| \delta^0 \|}{8N \delta^0 \| \delta^0 \|} \right) < \varepsilon. \]

(S.23)

Note that \( |T_b - T_b^0| \) is bounded away from zero in \( \mathbf{D}_{K,T} \). Observe that \( \left(z_{kh}/\sqrt{h}\right)\left(\epsilon_{kh}/\sqrt{h}\right) \) are independent in \( k \) and have zero mean and finite second moments. Hence, by the Hájek-Rényi inequality [see Lemma A.6 in Bai and Perron (1998)],
\[ P \left( \sup_{T_b \leq T_b^0 - K N^{-1}} \left\| \frac{1}{T_b^0 - T_b} \sum_{k=T_b+1}^{T_b^0} z_{kh} \epsilon_{kh} \right\| > B \frac{8 \| \delta^0 \|}{8N \delta^0 \| \delta^0 \|} N \right) \]

\[ \leq A \frac{64 \| \delta^0 \|^2 N^2}{B^2} \frac{1}{K N^{-1}}, \]

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

S.D.4.4 Proof of Proposition 3.3

We focus on the case with \( T_b \leq T_0 \). The arguments for the other case are similar and omitted. From Proposition 3.1 the distance \( |\hat{\lambda}_b - \lambda_0| \) can be made arbitrary small. Proposition 3.2 gives the associated rate of convergence: \( T(\hat{\lambda}_b - \lambda_0) = O_p(1) \). Given the consistency result for \( \hat{\lambda}_b \), we can apply a restricted search. In particular, by Proposition 3.2, for large \( T > \mathcal{T} \), we know that \( \{T_b \notin \mathbf{D}_{K,T}\} \), or equivalently...
of Proposition 3.1-3.2 the error in replacing $T^0_b$ with $\hat{T}_b$ is stochastically small and thus it does not affect the estimation of the parameters $\beta^0$, $\delta^0_1$ and $\delta^0_2$. Toward this end, we first find a lower bound on the convergence rate for $\hat{\lambda}_b$ 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 $\hat{T}_b$ in order to construct $Z_2$ and define $\hat{Z}_0 \triangleq Z_2$.

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

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

$$h^{-1/2}X'(Z_0 - \hat{Z}_0) = h^{1/2-\gamma} \frac{1}{h^{1-\gamma}} \sum_{T^0_b-[T^\gamma]} x_{kh} z_{kh}.$$  

Notice that, as $h \downarrow 0$, the number of terms in the sum on the right-hand side, for all $T > T$, increases to infinity at rate $1/h^\gamma$. Since $\hat{N}_b$ approaches $N^0_b$ at rate $T^{-(1-\gamma)}$, the quantity $X'(Z_0 - \hat{Z}_0)/h^{1-\gamma}$ is a consistent estimate of the so-called instantaneous or spot covariation between $X$ and $Z$ at time $N^0_b$.

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 $h \downarrow 0$,

$$X'_\Delta Z_\Delta/h^{1-\gamma} \overset{P}{\rightarrow} \Sigma_{XX,N^0_b},$$  

(S.24)

which implies that $h^{-1/2}X'(Z_0 - \hat{Z}_0) = O_p\left(h^{1/2-\gamma}\right)$. This shows that the order of the error in replacing $Z_0$ by $Z_2 = \hat{Z}_0$ goes to zero at a enough fast rate. That is, by definition we can write $Y = X\beta^0 + \hat{Z}_0 \delta^0 + (Z_0 - \hat{Z}_0) \delta^0 + \epsilon$, from which it follows that $X'\hat{Z}_0 = X'Z_0 + o_p(1)$, $X'(Z_0 - \hat{Z}_0) \delta^0 = o_p(1)$ and $Z_0'(Z_0 - \hat{Z}_0) \delta^0 = o_p(1)$. To see this, consider for example

$$X'(\hat{Z}_0 - Z_0) = \sum_{T^0_b-[T^\gamma]} x_{kh} z_{kh} = h^{1-\gamma} \sum_{T^0_b-[T^\gamma]} x_{kh} z_{kh} = h^{1-\gamma} O_p(1),$$

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

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

Proof of Proposition 3.3. By standard arguments,

$$\sqrt{T} \begin{bmatrix} \hat{\beta} - \beta^0 \\ \hat{\delta} - \delta^0 \end{bmatrix} = \begin{bmatrix} X'X & X'\hat{Z}_0 \\ \hat{Z}_0X & \hat{Z}_0\hat{Z}_0 \end{bmatrix}^{-1} \sqrt{T} \begin{bmatrix} X'e + X'(Z_0 - \hat{Z}_0) \delta^0 \\ \hat{Z}_0'e + \hat{Z}_0'(Z_0 - \hat{Z}_0) \delta^0 \end{bmatrix},$$

S-16
from which it follows that
\[
\begin{bmatrix}
X'X & X'\hat{Z}_0 \\
\hat{Z}_0'X & \hat{Z}_0'\hat{Z}_0
\end{bmatrix}^{-1} \frac{1}{h^{1/2}} X' \left(Z_0 - \hat{Z}_0\right) \delta^0 = O_p(1) o_p(1) = o_p(1),
\]
and a similar reasoning applies to \( \hat{Z}_0' \left(Z_0 - \hat{Z}_0\right) \delta^0 \). All other terms involving \( \hat{Z}_0 \) can be treated in analogous fashion. In particular, the \( O_p(1) \) result above follows from Lemma S.D.3-S.D.4. The rest of the arguments (including mixed normality) follows from Barndorff-Nielsen and Shephard (2004) and are omitted. □

**S.D.4.5 Proof of Proposition 4.1**

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

\[
Q_T(T_b) - Q_T(T_b^0) = \delta'_{T_b} (Z_2'MZ_2) \hat{\delta}_{T_b} - \delta'_{T_b^0} (Z_2'MZ_0) \hat{\delta}_{T_b^0}
\]

\[
= \delta'_{h} \left\{ (Z_2'MZ_2) (Z_2'MZ_2)^{-1} (Z_2'MZ_0) - Z_2'MZ_0 \right\} \delta_h 
+ g_e(T_b),
\]

where

\[
g_e(T_b) = 2\delta'_{h} (Z_2'MZ_2) (Z_2'MZ_2)^{-1} Z_2Me - 2\delta'_{h} (Z_2'Me)
+ \epsilon'MZ_2 (Z_2'MZ_2)^{-1} Z_2Me - \epsilon'MZ_0 (Z_2'MZ_0)^{-1} Z_0'Me.
\]

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

\[
P\left( |\hat{\lambda}_b - \lambda_0| > K \right)
\leq P\left( \sup_{|T_b - T_b^0| > TK} h^{-1/2} |g_e(T_b)| \geq \inf_{|T_b - T_b^0| > TK} h^{-1/2} |T_b - T_b^0| r(T_b) \right)
\]

\[
\leq P\left( \sup_{p \leq T_b \leq T - p} h^{-1/2} |g_e(T_b)| \geq TK \inf_{|T_b - T_b^0| > TK} h^{-1/2} r(T_b) \right)
\]

\[
= P\left( r_T^{-1} \sup_{p \leq T_b \leq T - p} h^{-1/2} |g_e(T_b)| \geq K \right),
\]

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

\[
\sup_{p \leq T_b \leq T - p} h^{-1/2} |g_e(T_b)| = o_p(1).
\]

Consider the first term of \( g_e(T_b) \):

\[
2\delta'_{h} (Z_2'MZ_2) (Z_2'MZ_2)^{-1/2} (Z_2'MZ_2)^{-1/2} Z_2Me
\leq 2h^{1/4} \left( \delta^0 \right)' (Z_2'MZ_2) (Z_2'MZ_2)^{-1/2} (Z_2'MZ_2)^{-1/2} Z_2Me.
\]

S-17
For any \(1 \leq j \leq p\), \((Z_2\tilde{e})_{j,1}/\sqrt{h} = O_p(1)\) by Theorem S.D.5, and similarly, for any \(1 \leq i \leq q + p\), \((X\tilde{e})_{i}/\sqrt{h} = O_p(1)\). Furthermore, from Lemma S.D.3 we also have that \(Z_2' M Z_2\) and \(Z_0' M Z_2\) are \(O_p(1)\). Therefore, the supremum of \((Z_2' M Z_2)(Z_2' M Z_2)^{-1/2}\) over all \(T_b\) is such that

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

by Lemma S.D.3. By Assumption 2.1-(iii) \((Z_2' M Z_2)^{-1/2} Z_2 M\tilde{e}\) is \(O_p(1) O_p(\sqrt{h})\) uniformly, which implies that (S.32) is \(O_p(\sqrt{h})\) uniformly over \(p \leq T_b \leq T - p\). In view of Assumption 4.1 [recall (4.1)], we need to study the behavior of \((X'\tilde{e})_{j,1}\) for \(1 \leq j \leq p + q\). Note first that \(|\hat{\lambda}_b - \lambda_0| > K\) or \(N > |\hat{N}_b - N_0| > KN\). Then, by Itô formula, proceeding as in the proof of Lemma S.D.2, we have a standard result for the local volatility of a continuous Itô semimartingale; namely that for some \(A > 0\) (recall the condition \(T^{1-\epsilon} \to B > 0\)),

\[
\left\| \mathbb{E} \left( \frac{1}{\epsilon} \sum_{T_b = [T^\epsilon]}^{T_b + [T^\epsilon]} x_{kh} \tilde{e}_{kh} - \frac{1}{\epsilon} \int_{N_b^0 - \epsilon}^{N_b^0} \sum_{s \leq t} \|X_{e,s}d| \mathcal{F}_{(T_b^0 - 1)\epsilon} \right) \right\| \leq Ah^{1/2}.
\]

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

\[
X'\tilde{e} = \sum_{k=1}^{T_b^0 - [T^\epsilon]} x_{kh} \tilde{e}_{kh} + h^{-1/4} \sum_{k=T_b^0 - [T^\epsilon] + 1}^{T_b^0} x_{kh} \tilde{e}_{kh} + \sum_{k=T_b^0 + [T^\epsilon] + 1}^{T} x_{kh} \tilde{e}_{kh} = O_p \left( h^{1/2} \right) + h^{-1/4} O_p \left( h^{1-\epsilon + 1/2} \right) + O_p \left( h^{1/2} \right) = O_p \left( h^{1/2} \right).
\]

(S.33)

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

\[
2h^{-1/2} h^{1/4} \left( \delta^0 \right)' (Z_2' M Z_2)(Z_2' M Z_2)^{-1/2} (Z_2' M Z_2)^{-1/2} Z_2 M e
\]

\[
= 2h^{-1/2} h^{1/4} \left\| \delta^0 \right\| O_p \left( h^{1/2} \right) = O_p \left( h^{1/2} \right).
\]

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

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

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

\[
h^{-1/2} e' M Z_2 (Z_2' M Z_2)^{-1} Z_2 M e
\]

\[
= h^{-1/2} BO_p \left( h^{1/2} \right) O_p \left( h^{1/2} \right) = O_p \left( h^{1/2} \right).
\]

(S.34)

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

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

**Proof.** Assume \(N_b \leq N_0\), and observe that \(r_T \geq r_{B,h}\) for an appropriately chosen \(B\). From the first inequality result in Lemma S.D.1,
By Lemma S.D.1, approximation is uniform, establishing that

\[ Th^{-1/2} r (T_b) \geq Th^{-1/2} h^{1/2} \left( \delta_0 \right)^{\prime} \left( X_{\Delta}^{\prime} X_{\Delta} / T_0 - T_b \right) (X_2 X_2' - 1) (X_0 X_0) R \delta_0 \]

\[ = \left( \delta_0 \right)^{\prime} R' \left( X_{\Delta}^{\prime} X_{\Delta} / (N_h^0 - N_b) \right) (X_2 X_2' - 1) (X_0 X_0) R \delta_0. \]

Note that \( B < h (T_0^0 - T_b) < N \). Then

\[ Th^{-1/2} r (T_b) \geq \left( \delta_0 \right)^{\prime} R' \left( X_{\Delta}^{\prime} X_{\Delta} / N \right) (X_2 X_2' - 1) (X_0 X_0) R \delta_0 > A \]

by the same argument as in Lemma S.D.6. Following the same reasoning as in the proof of Lemma S.D.6 we can choose a \( B > 0 \) such that \( r_{B,h} = \inf_{|T_b - T_0^0| > T_B} Th^{-1/2} r (T_b) \) satisfies \( P (r_{B,h} \geq A) \leq 1 - \varepsilon \). □

**Proof of part (ii) of Proposition 4.1.** Suppose \( T_b < T_0^0 \). Let

\[ D_{K,T} = \left\{ T_b : N \eta \leq N_b \leq N (1 - \eta), \ |N_b - N_0^0| > K (T^{-1})^{\prime} \right\}. \]

It is enough to show that \( P \left( \sup_{T_b \in D_{K,T}} Q_T (T_b) \geq Q_T (T_0^0) \right) < \varepsilon \). The difficulty is again to control the estimates that depend on \( |N_b - N_0^0| \). We need to show

\[ P \left( \sup_{T_b \in D_{K,T}} h^{-3/2} g_e (T_b, \delta_h) \geq \inf_{T_b \in D_{K,T}} h^{-3/2} r (T_b) \right) < \varepsilon. \]

By Lemma S.D.1,

\[ \inf_{T_b \in D_{K,T}} r (T_b) \geq \inf_{T_b \in D_{K,T}} \delta_h \left( X_{\Delta}^{\prime} X_{\Delta} / T_0 - T_b \right) (X_2 X_2' - 1) (X_0 X_0) R \delta_h \]

and since \( |T_b - T_0^0| > KT^\kappa \), it is important to consider \( X_{\Delta}^{\prime} X_{\Delta} = \sum_{k = T_0^0 + 1}^{T_0} x_{kh} x_{kh}'. \) We shall apply asymptotic results for the local approximation of the covariation between processes. Consider

\[ \frac{X_{\Delta}^{\prime} X_{\Delta}}{h (T_0^0 - T_b)} = \frac{1}{h (T_0^0 - T_b)} \sum_{k = T_0^0 + 1}^{T_0} x_{kh} x_{kh}'. \]

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

\[ \frac{1}{h (T_0^0 - T_b)} \sum_{k = T_0^0 + 1}^{T_0} x_{kh} x_{kh}' \xrightarrow{P} \Sigma_{XX,N_0^0}, \quad (S.35) \]

since \( |N_b - N_0^0| \) shrinks at a rate no faster than \( K h^{1-\kappa} \) and \( 1/K h^{1-\kappa} \to \infty \). By Lemma S.D.2 this approximation is uniform, establishing that

\[ h^{-3/2} \inf_{T_b \in D_{K,T}} \left( \delta_h \right)^{\prime} R' \left( X_{\Delta}^{\prime} X_{\Delta} / T_0 - T_b \right) (X_2 X_2' - 1) (X_0 X_0) R \delta_h \]

\[ = \inf_{T_b \in D_{K,T}} \left( \delta_0 \right)^{\prime} R' \left( X_{\Delta}^{\prime} X_{\Delta} / h (T_0^0 - T_b) \right) (X_2 X_2' - 1) (X_0 X_0) R \delta_0, \]
is bounded away from zero. Thus, it is sufficient to show

\[
P \left( \sup_{T_b \in D_{K,T}} \frac{h^{-3/2} g_{e}(T_b, \delta_h)}{|T_b - T_0^0|} \geq B \right) < \varepsilon, \tag{S.36}
\]

for some \( B > 0 \). Consider the terms of \( g_{e}(T_b) \) in (S.29). Using \( Z_2 = Z_0 \pm Z_\Delta \), we deduce for the first term,

\[
\delta'_{h} (Z'_0 M Z_2) \left( Z'_2 M Z_2 \right)^{-1} Z_2 M e
\]

\[
= \delta'_{h} \left( (Z'_2 + Z_\Delta) M Z_2 \right) \left( Z'_2 M Z_2 \right)^{-1} Z_2 M e
\]

\[
= \delta_{h} Z'_0 M e + \delta'_{h} Z'_\Delta M e + \delta_{h} \left( Z'_\Delta M Z_2 \right) \left( Z'_2 M Z_2 \right)^{-1} Z_2 M e. \tag{S.37}
\]

First, we can apply Lemma S.D.3 [(vi)-(viii)], together with Assumption 2.1-(iii), to the terms that do not involve \( |N^0_b - N^0_0| \). Let us focus on the third term,

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

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

\[
Z'_\Delta Z_\Delta / K h^{1-\kappa} = O_p \left( 1 \right), \tag{S.39}
\]

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

\[
\frac{1}{K} h^{-(1-\kappa)-1/2} \left( \delta'_{h} \right)' \left( Z'_\Delta M Z_2 \right) \left( Z'_2 M Z_2 \right)^{-1} Z_2 M e \tag{S.40}
\]

\[
\leq \frac{1}{K} h^{-1/4} \left( \delta'_{h} \right)' \left( \frac{Z'_\Delta M Z_2}{h^{1-\kappa}} \right) \left( Z'_2 M Z_2 \right)^{-1} Z_2 M e
\]

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

where \( (Z'_2 M Z_2)^{-1} = O_p \left( 1 \right) \). So the right-hand side of (S.40) is less than \( \varepsilon/4 \) in probability. Therefore, for the second term of (S.37),

\[
K^{-1} h^{-(1-\kappa)-1/2} \delta'_{h} Z'_\Delta M e
\]

\[
= h^{-1/2} \frac{K h^{1-\kappa} \delta'_{h}}{h^{1-\kappa}} \sum_{k=T_b+1}^{T_b^0} z_{kh} e_{kh}
\]

\[
- h^{-1/2} \frac{h^{1-\kappa} \delta'_{h}}{h^{1-\kappa}} \sum_{k=T_b+1}^{T_b^0} z_{kh} x'_{kh} \left( X'X \right)^{-1} \left( X'e \right)
\]

\[
\leq h^{-1/2} \frac{K h^{1-\kappa} \delta'_{h}}{h^{1-\kappa}} \sum_{k=T_b+1}^{T_b^0} z_{kh} e_{kh}
\]

\[
- B \frac{h^{-1/4}}{K} h^{1-\kappa} \left( \delta'_{h} \right) \left( \sum_{k=T_b+1}^{T_b^0} z_{kh} x'_{kh} \right) \left( X'X \right)^{-1} \left( X'e \right)
\]

S-20
Thus, using (S.37), (S.28) is such that
\[
2\delta_h' Z_0'Me \pm 2\delta_h' (Z_2'MZ_2) (Z_2'MZ_2)^{-1} Z_2Me - 2\delta_h' (Z_0'Me) \leq \frac{h^{-1/2}}{Kh^{1-\kappa}} \sum_{k=T_b+1}^{T_b} z_{kh} e_{kh} - h^{-1/4} O_p(1) O_p\left(h^{1/2}\right),
\]
(S.41)

in view of (S.40) and (S.41). Next, consider (S.29). We can use the decomposition \(Z_2 = Z_0 \pm Z_\Delta\) and show that all terms involving the matrix \(Z_\Delta\) are negligible. To see this, consider the first term when multiplied by \(K^{-1} h^{-(3/2-\kappa)}\),
\[
K^{-1} h^{-(3/2-\kappa)} e'M Z_2 (Z_2'MZ_2)^{-1} Z_2Me
= K^{-1} h^{-(3/2-\kappa)} e'M Z_0 (Z_2'MZ_2)^{-1} Z_2Me
\pm K^{-1} h^{-(3/2-\kappa)} e'M Z_\Delta (Z_2'MZ_2)^{-1} Z_2Me.
\]
(S.42)

By the same argument as in (S.33), \(Z_2'Me = O_p\left(h^{1/2}\right)\). Then, using the Burkholder-Davis-Gundy inequality, estimates for the local volatility of continuous \(\hat{I}\) semimartingales yield
\[
\bar{e}'MZ_\Delta = \bar{e}'Z_\Delta - \bar{e}'X (X'X)^{-1} X'Z_\Delta
= O_p\left(Kh^{1/2+1-\kappa}\right) - O_p\left(h^{1/2}\right) O_p(1) O_p\left(Kh^{1-\kappa}\right).
\]

Thus, the second term in (S.42) is such that
\[
K^{-1} h^{-(3/2-\kappa)} e'M Z_\Delta (Z_2'MZ_2)^{-1} Z_2Me
\]
(S.43)

\[
= B \left(K^{-1} h^{-(3/2-\kappa)}\right) O_p\left(Kh^{1-\kappa+1/2}\right) O_p(1) O_p\left(h^{1/2}\right)
= BO_p\left(h^{1/2}\right).
\]

Next, let us consider (S.29). The key here is to recognize that, on \(D_{K,T}\), \(T_b\) 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),
\[
\bar{e}'MZ_0 (Z_0'MZ_0)^{-1} Z_0Me = O_p\left(h^{1/2}\right) O_p(1) O_p\left(h^{1/2}\right) = O_p\left(h\right).
\]

By the fact that \(Z_0 = Z_2 \pm Z_\Delta\) applied repeatedly in (S.42), and noting that the cross-product terms involving \(Z_\Delta\) 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
\[
e'M Z_0 (Z_0'MZ_2 \pm Z_\Delta'MZ_2)^{-1} Z_0Me - e'M Z_0 (Z_0'MZ_0)^{-1} Z_0Me
= e'M Z_0 \left[\left(Z_0'MZ_2 \pm Z_\Delta'MZ_2\right)^{-1} - (Z_0'MZ_0)^{-1}\right] Z_0Me.
\]

On \(D_{K,T}\), \(|N_b - N_b^0| = O_p\left(Kh^{1-\kappa}\right)\), and so each term involving \(Z_\Delta\) 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)]
\[ h^{-3/2 - \kappa} K^{-1} e' M Z_0 \left[ (Z_0' M Z_2 \pm Z_\Delta M Z_2)^{-1} - (Z_0' M Z_0)^{-1} \right] Z_0' M e \]
\[ = h^{-3/2 - \kappa} K^{-1} O_p (h) \left[ (Z_0' M Z_2 \pm Z_\Delta M Z_2)^{-1} - (Z_0' M Z_0)^{-1} \right] \]
\[ = h^{-3/2 - \kappa} K^{-1} O_p (h) \]
\[ \times \left[ (Z_0' M Z_0 \pm Z_0' M Z_\Delta \pm Z_\Delta M Z_2)^{-1} - (Z_0' M Z_0)^{-1} \right] \]
\[ = h^{-3/2 - \kappa} K^{-1} O_p (h) o_p \left( h^{1 - \kappa} \right) = O_p \left( h^{1/2} \right) o_p (1). \]

Therefore, (S.29) is stochastically small uniformly in \( T_b \in D_{K, T} \) when \( T \) is large. Altogether, we have
\[ h^{-1/2} \frac{g_e (T_b)}{|T_b - T_b^0|} \leq 2 \frac{h^{-1/2} \delta'_h}{K h^{1 - \kappa}} \sum_{k=1}^{T_b^0} z_{kh} e_{kh} \]
\[ - h^{-1/4} O_p (1) O_p \left( h^{1/2} \right) + O_p \left( h^{-1/4} \right). \]

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 \( T_b \),
\[ E \left[ \left\| \sum_{b, h} \right( T_2, T_b^0) - \sum_{b, h} \left( T_2, T_b^0 \right) \right\| \mathcal{F}_{T_b h} \right] \leq B h^{1/2}, \quad (S.44) \]
for some \( B > 0 \). Let \( R_{1, h} = \sum_{k=T_b^0 - (B+1) T^\kappa}^{T_b^0} z_{kh} e_{kh} \), \( R_{2, h} (T_b) = \sum_{k=T_b^0 + 1}^{T_b^0 - (B+1) T^\kappa} z_{kh} e_{kh} \) and note that \( \sum_{k=T_b^0 + 1}^{T_b^0 - (B+1) T^\kappa} z_{kh} e_{kh} = R_{1, h} + R_{2, h} (T_b) \). Then, for any \( C > 0 \),
\[ P \left( \sup_{T_b < T_b^0 - K T^\kappa} \frac{1}{K h^{1 - \kappa}} \left\| R_{1, h} \right\| > 4^{-1} C \left\| \delta_0 \right\| \right)^{1/2} \]
\[ \leq P \left( \sup_{T_b < T_b^0 - K T^\kappa} \frac{1}{K h^{1 - \kappa}} \left\| R_{2, h} (T_b) \right\| > 4^{-1} C \left\| \delta_0 \right\| \right)^{1/2} \]
\[ \leq P \left( \frac{1}{K h^{1 - \kappa}} \left\| R_{2, h} (T_b) \right\| > 4^{-1} C \left\| \delta_0 \right\| \right)^{1/2} \]
\[ \leq \frac{(4 (B + 1) \left\| \delta_0 \right\|)^r}{4 h^{-r/4} K B T^r E \left( \frac{1}{(B + 1) K h^{1 - \kappa}} \left\| R_{2, h} (T_b) \right\| \right)^r} \]

S-22
\[ \leq C_r (B + 1) B^{-1} \left\| \delta^0 \right\| h^{-r/4} T^\kappa h^{r/2} \leq C_r \left\| \delta^0 \right\| h^{r/2 - \kappa - r/4} \to 0, \]

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

\[
P \left( \frac{1}{Kh^{1-\kappa}} \left\| R_{1,h} \right\| > 2^{-1} C \left\| \delta^0 \right\|^{-1} h^{1/2} \right) \]
\[
\leq P \left( \left( B + 1 \right) K \left\| R_{1,h} \right\| > 2^{-1} C \left\| \delta^0 \right\|^{-1} h^{1/2} \right)
\]
\[
> \frac{C}{4} \left\| \delta^0 \right\|^{-1} h^{1/2}
\]
\[
\leq P \left( (B + 1) K^{-1} O_\varepsilon(1) > 2^{-1} C \left\| \delta^0 \right\|^{-1} \right) \to 0,
\]

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

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

\[
\sqrt{T} \begin{bmatrix} \tilde{\beta} - \beta_0 \\ \delta - \delta_h \end{bmatrix} = \begin{bmatrix} X'X & X'\tilde{Z}_0 \\ \tilde{Z}_0X & \tilde{Z}_0'\tilde{Z}_0 \end{bmatrix}^{-1} \sqrt{T} \begin{bmatrix} X'e + X'(Z_0 - \tilde{Z}_0) \delta_h \\ \tilde{Z}_0'e + \tilde{Z}_0'(Z_0 - \tilde{Z}_0) \delta_h \end{bmatrix},
\]

so that we have to show

\[
\begin{bmatrix} X'X & X'\tilde{Z}_0 \\ \tilde{Z}_0X & \tilde{Z}_0'\tilde{Z}_0 \end{bmatrix}^{-1} \frac{1}{h^{1/2}} X'(Z_0 - \tilde{Z}_0) \delta_h \xrightarrow{P} 0,
\]

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),

\[
X'e = \sum_{k=1}^{T_b-[T^\kappa]} x_{kh}\tilde{e}_{kh} + h^{-1/4} \sum_{k=T_b-[T^\kappa]+1}^{T_b+[T^\kappa]} x_{kh}\tilde{e}_{kh} + \sum_{k=T_b+[T^\kappa]+1}^{T} x_{kh}\tilde{e}_{kh}
\]
\[
\triangleq R_{1,h} + R_{2,h} + R_{3,h}.
\]

By Theorem S.D.5, \( h^{-1/2} R_{1,h} \xrightarrow{P} \mathcal{N} (0, V_1) \), where \( V_1 \triangleq \lim_{T \to \infty} T \sum_{k=1}^{T_b-[T^\kappa]} E \left( x_{kh}x_{kh}'\tilde{e}_{kh}^2 \right) \). Similarly, \( h^{-1/2} R_{3,h} \xrightarrow{P} \mathcal{N} (0, V_3) \), where \( V_3 \triangleq \lim_{T \to \infty} T \sum_{k=T_b+[T^\kappa]+1}^{T} E \left( x_{kh}x_{kh}'\tilde{e}_{kh}^2 \right) \). If \( \kappa \in (0, 1/4) \), \( h^{-1-\kappa} \sum_{k=T_b-[T^\kappa]+1}^{T_b+[T^\kappa]} x_{kh}\tilde{e}_{kh} \xrightarrow{P} \Sigma_{X_e,N^b_0} \) by Theorem 9.3.2 in Jacod and Protter (2012) and so \( h^{-1/2} R_{2,h} = h^{-3/4} \sum_{k=T_b-[T^\kappa]}^{T_b+[T^\kappa]} x_{kh}\tilde{e}_{kh} \xrightarrow{P} 0 \). If \( \kappa = 1/4 \), then \( h^{-1/2} R_{2,h} \to \Sigma_{X_e,N^b_0} \) in probability again by Theorem 9.3.2 in Jacod and Protter (2012). Since by Assumption 2.1-(iv) \( \Sigma_{X_e,t} = 0 \) 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. \( \Box \)

S-23
S.D.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.D.9.** The following identity holds

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

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

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

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

$$
Q_T(T_b) - Q_T(T_0) = -\delta_h' \left\{ (Z_0' M Z_2) (Z_2' M Z_2)^{-1} (Z_2' M Z_2) \right\} \delta_h + g_e(T_b, \delta_h),
$$

where

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

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

$$
\frac{1}{h^{1-\kappa}} \sum_{k=T_b^0+u T^\kappa}^{T_b^0} x_{kh} x_{kh}^t P_{\kappa} \sum_{X,N_b^0}.
$$

(S.49)

Since $Z_0' X = Z_0' X_{\Delta}$, we will use this result also for $Z_0' X/h^{1-\kappa}$. With the notation of Section S.D.4.1 [recall (S.6)], by the Burkhölder-Davis-Gundy inequality, we have that standard estimates for the local volatility yield,

$$
\left\| E \left( \hat{\Sigma}_{XX} \left( T_b, T_b^0 \right) - \Sigma_{XX,(T_b^0-1)\delta h} \right) \mathcal{F}_{(T_b^0-1)\delta h} \right\| \leq B h^{1/2}.
$$

(S.50)

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

$$
\psi_h^{-1} Z_0' X (X'X)^{-1} X' Z_0 = O_p(1) X' Z_{\Delta}.
$$

(S.51)

and

$$
Z_0' M Z_2 = Z_0' Z_{\Delta} - Z_0' X (X'X)^{-1} X' Z_2 = O_p(\psi_h) - O_p(\psi_h) O_p(1) O_p(1).
$$

(S.52)
Now, expand the first term of (S.46),

\[ \delta_h Z'_\Delta M Z \delta_h = \delta_h Z'_\Delta Z \delta_h - \delta_h Z'_\Delta X (X'X)^{-1} X'Z \delta_h. \]  

(S.53)

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

\[ \psi^{-1}_h \delta_h Z'_\Delta M Z \delta_h = \psi^{-1}_h \delta_h Z'_\Delta Z \delta_h - \psi^{-1}_h \delta_h Z'_\Delta X (X'X)^{-1} X'Z \delta_h. \]  

(S.54)

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

\[ \| \delta^0 \|^2 h^{1/2} \frac{Z'_\Delta X}{\psi \delta} (X'X)^{-1} X'Z = \| \delta^0 \|^2 h^{1/2} O_p(1) X'Z, \]  

(S.55)

uniformly in \(T_h(u)\). Therefore,

\[ \psi^{-1}_h \delta_h Z'_\Delta M Z \delta_h = \psi^{-1}_h \delta_h Z'_\Delta Z \delta_h - \| \delta^0 \|^2 h^{1/2} O_p(1) O_p(\psi_h). \]  

(S.56)

The last equality shows that the second term of \(\delta' Z'_\Delta M Z \delta\) 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 \(\psi^{-1}_h\) is, uniformly in \(T_h(u)\),

\[ \psi^{-1}_h \delta_h \left( Z'_\Delta M Z_2 \right) \left( Z'_2 M Z_2 \right)^{-1} \left( Z'_2 M Z \right) \delta'_h = \| \delta^0 \|^2 h^{1/2} O_p(1) O_p(1) O_p(\psi_h), \]

where we have used the fact that \(Z'_\Delta M Z_2 / \psi_h = O_p(1)\) [cf. (S.52)]. Hence, the second term of (S.46), when multiplied by \(\psi^{-1}_h\), is \(O_p\left(h^{3/2-\kappa}\right)\) uniformly in \(T_h\). Finally, let us consider \(g_c (T_h, \delta_h)\). Recall that \(\tilde{e}_{kh}\) defined in (4.1) is i.n.d. with zero mean and conditional variance \(\sigma^2_{e,k-1} h\). 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,

\[ \psi^{-1}_h \left[ e' M Z_2 \left( Z'_2 M Z_2 \right)^{-1} Z_2 M e - e' M Z_0 \left( Z'_0 M Z_0 \right)^{-1} Z'_0 M e \right] = o_p(c_T h). \]

Next, consider the first two terms of \(g_c (T_h, \delta_h)\). Using \(Z'_0 M Z_2 = Z'_2 M Z_2 \pm Z'_\Delta M Z_2\), it is easy to show that

\[ 2h^{1/4} \left( \delta^0 \right)' \left( Z'_0 M Z_2 \right) \left( Z'_2 M Z_2 \right)^{-1} Z_2 M e - 2h^{1/4} \left( \delta^0 \right)' \left( Z'_0 M e \right) \]

\[ = 2h^{1/4} \left( \delta^0 \right)' Z'_\Delta M e + 2h^{1/4} \left( \delta^0 \right)' Z'_\Delta M Z_2 \left( Z'_2 M Z_2 \right)^{-1} Z'_2 M e. \]  

(S.57)

Note that, uniformly in \(T_h(u)\),

\[ \psi^{-1}_h h^{1/4} \left( \delta^0 \right)' Z'_\Delta M Z_2 \]

\[ = h^{1/4} \left( \delta^0 \right)' Z'_\Delta Z + \left( \delta^0 \right)' h^{1/4} \frac{Z'_\Delta X}{\psi_h} (X'X)^{-1} X'Z \]

\[ = h^{1/4} \left( \delta^0 \right)' \frac{Z'_\Delta Z}{\psi_h} + \left( \delta^0 \right)' h^{1/4} O_p(1) \]

S-25
\[ = h^{1/4} \| \delta^0 \| O_p(1) + \| \delta^0 \| h^{1/4} O_p(1), \]

where we have used (S.49) and the fact that \((X'X)^{-1} \) and \(X'Z_2\) are each \(O_p(1)\). Recall the decomposition in (S.33):

\[ X'e = O_p \left(h^{1-\kappa+1/4} \right) + O_p \left(h^{1/2} \right). \]  
(S.58)

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

\[ \psi_h^{-1} 2h^{1/4} \left( \delta^0 \right)' Z'_\Delta M Z_2 (Z'_2 M Z_2)^{-1} Z'_2 M e \]
\[ = h^{1/4} \left( \delta^0 \right)' O_p(1) O_p(1) \left[ O_p \left( h^{1-\kappa+1/4} \right) + O_p \left( h^{1/2} \right) \right] \]
\[ = \left( \delta^0 \right)' h^{1/4} O_p(1) O_p \left(h^{1/2} \right) \]
\[ = \left( \delta^0 \right)' \delta^0 \| h^{1/4} O_p(1) O_p \left(h^{1/2} \right) \]
\[ = \left( \delta^0 \right)' \delta^0 \| \| h^{3/4} O_p \left(h^{3/4} \right). \]

The first term of (S.57) can be decomposed further as follows

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

Then, when multiplied by \(\psi_h^{-1}\), the second term above is, uniformly in \(T_b\),

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

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

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

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

\[ \psi_h^{-1} \left( Q_T \left(T_b \right) - Q_T \left(T_b^0 \right) \right) \]
\[ = -\delta_h \left( Z'_\Delta Z/\psi_h \right) \delta_h \pm 2\delta_h \left( Z'_\Delta e / \psi_h \right) \]
\[ + O_p \left( h^{3/2-\kappa} \right) + O_p \left( h^{3/4} \right) + \| \delta^0 \| O_p \left( h^{3/4} \right) + o_p \left( c_T h \right). \]

Therefore, for \(T\) large enough,

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

This concludes the proof of Lemma 4.1. \( \square \)

**S.D.4.7 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 \(Q_T \left(T_b(v)\right) - Q_T \left(T_b^0 \right)\) is determined by that of \(-\delta_h \left( Z'_\Delta Z/\psi_h \right) \delta_h \pm 2\delta_h \left( Z'_\Delta e / \psi_h \right)\). Next, we rescale time with \(s \mapsto t \triangleq \psi_h^{-1} s\) on \(\mathcal{D}(C)\). This is achieved by rescaling the criterion function \(Q_T \left(T_b(u)\right) - Q_T \left(T_b^0 \right)\) by the factor \(\psi_h^{-1}\). First, note that the processes \(Z_t\) and \(e^*_t\) [recall (2.3) and (4.1)] are rescaled as follows on
$D(\mathcal{C})$. Let $Z_{\psi,s} \triangleq \psi^{-1/2}_h Z_s$, $W_{\psi,e,s} \triangleq \psi^{-1/2}_h W_{e,s}$ and note that

$$dZ_{\psi,s} = \psi^{-1/2}_h \sigma_{Z,s} dW_{Z,s}, \quad dW_{\psi,e,s} = \psi^{-1/2}_h \sigma_{e,s} dW_{e,s}, \quad \text{with } s \in D(\mathcal{C}). \quad (S.59)$$

For $s \in [N_0^0 - Ch^{1-\kappa}, N_0^0 + Ch^{1-\kappa}]$, let $v = \psi^{-1}_h (N_0^0 - s)$ and, by using the properties of $W_{e,s}$ and the fact that $\sigma_{Z,s}$, $\sigma_{e,s}$ are $\mathcal{F}_s$-measurable, we have

$$dZ_{\psi,t} = \sigma_{Z,t} dW_{Z,t}, \quad dW_{\psi,e,t} = \sigma_{e,t} dW_{e,t}, \quad \text{with } t \in D^* (\mathcal{C}). \quad (S.60)$$

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

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

which by (S.59)-(S.60) is such that

$$\psi^{-1}_h Z'_\Delta Z_{\Delta} = \sum_{k=T_b^0 + [v/h]}^{T_b^0} z_{kh} z'_{kh}, \quad v \in D^* (\mathcal{C}). \quad (S.61)$$

Using the same argument:

$$\psi^{-1}_h Z'_\Delta \tilde{e} = \sum_{k=T_b^0 + [v/h]}^{T_b^0} z_{kh} \tilde{e}_{kh}, \quad v \in D^* (\mathcal{C}). \quad (S.62)$$

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

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

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

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

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

$$Q_T (\theta^*) = - (\delta')' \left( \sum_{k=T_b+1}^{T_b^0} z_{kh} z'_{kh} \right) (\delta') + 2 (\delta')' \left( h^{-1/2} \sum_{k=T_b+1}^{T_b^0} z_{kh} \tilde{e}_{kh} \right). \quad (S.63)$$

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

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

S-27
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 \(\mathbb{D}(\mathbb{D}^s(C), \mathbb{R})\). 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 \(J_{Z,h}(v) \xrightarrow{u.p.} [Z, Z]_1(v)\), where \([Z, Z]_1(v) \triangleq [Z, Z]_h\lfloor v/h \rfloor - [Z, Z]_h[v/h]_1\), and recall by Assumption 2.2 that \([Z, Z]_1(v)\) is equivalent to \((Z, Z)_1(v)\) where \((Z, Z)_1(v) = (Z, Z)_{h[v/h]}(v)\).

Next, let \(\mathcal{W}_h(v) = h^{-1/2}J_{e,h}(v)\) and \(\mathcal{W}_1(v) = \int_{N_0^h}^{N_1^h} \sigma_{Z,e,a}dW_s^a\) where \(W_s^a\) is defined in Section S.B. By Theorem 5.4.2 in Jacod and Protter (2012) we have \(\mathcal{W}_h(v) \xrightarrow{L^2} \mathcal{W}_1(v)\) under the Skorokhod topology. Note that the both limit processes \([Z, Z]_1(v)\) and \(\mathcal{W}_1(v)\) are continuous. This restores the compatibility of the Skorokhod topology with the natural linear structure of \(\mathbb{D}(\mathbb{D}^s(C), \mathbb{R})\). For \(v \leq 0\), the finite-dimensional stable convergence in law for \(\overline{Q}_T(\cdot)\) then follows: \(\overline{Q}_T(0^+) \xrightarrow{L^2} - (\delta^0)'(Z, Z)_1(v)\delta^0 + 2(\delta^0)'\mathcal{W}_1(v)\). Next, we verify the asymptotic stochastic equicontinuity of the sequence of processes \(\{\overline{Q}_T(\cdot), T \geq 1\}\). For \(1 \leq i \leq p\), let \(\zeta_{h,k}^{(i)} \triangleq \varepsilon_{kh}^{(i)}\), \(\zeta_{h,k}^{*} \triangleq \varepsilon_{kh}^{(i)}\), and \(\zeta_{Z,h,k}^{*} \triangleq \varepsilon_{kh}^{(i)}\). For \(1 \leq i, j \leq p\), let \(\zeta_{Z,h,k}^{(i,j)} \triangleq \varepsilon_{kh}^{(i,j)} - \varepsilon_{kh}^{(i)}\). For \(1 \leq i, j \leq p\), let \(\zeta_{Z,h,k}^{(i,j)} \triangleq \varepsilon_{kh}^{(i,j)} - \varepsilon_{kh}^{(i)}\). Then, we have the following decomposition for \(\overline{Q}_T(\theta^*) \triangleq \overline{Q}_T(\theta^*) + (\delta^0)'(Z, Z)_1(v)\delta^0 (v \leq 0, \text{and defined analogously for } v > 0)\),

\begin{align}
\overline{Q}_T(\theta^*) = \sum_{r=1}^{4} \overline{Q}_{r,T}(\theta^*),
\end{align}

where \(\overline{Q}_{1,T}(\theta^*) \triangleq - (\delta^0)'(\sum_k \zeta_{Z,h,k}^*\delta^0), \overline{Q}_{2,T}(\theta^*) \triangleq - (\delta^0)'(\sum_k \zeta_{Z,h,k}^* \delta^0), \overline{Q}_{3,T}(\theta^*) \triangleq (\delta^0)'(h^{-1/2} \sum_k \zeta_{h,k}^*), \text{and } \overline{Q}_{4,T}(\theta^*) \triangleq (\delta^0)'(h^{-1/2} \sum_k \zeta_{h,k}^*); \text{where } \sum_k \text{ stands for } \sum_{T_{b,v}^0}^{T_{b,v}^0 + \lfloor v/h \rfloor}. \text{Then,}

\begin{align}
\sup_{(\theta, v) \in \overline{D}^s(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}{\rightarrow} 0,
\end{align}

which follows from Jacod and Rosenbaum (2013) given that \(\Sigma_{Z,e,k} = 0\) 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 \((\theta, v), (\bar{\theta}, \bar{v})\) with \(v < \bar{v}\left( \text{the other cases can be proven similarly}\right)\) and \(\bar{\delta} = \delta + c_{p \times 1}, \text{where } c_{p \times 1} \text{ is a } p \times 1 \text{ vector with each entry equals to } c \in \mathbb{R}, \text{with } 0 < c \leq \tau < \infty, \text{then}

\begin{align}
\left| \overline{Q}_{1,T}(\theta^*) - \overline{Q}_{1,T}(\bar{\theta}^*) \right| &= \bar{\delta}' \left( \sum_{k=T_{b,v}^0 + 1}^{T_{b,v}^0 + \lfloor \bar{v}/h \rfloor} \zeta_{Z,h,k}^* \right) \bar{\delta} - \delta' \left( \sum_{k=T_{b,v}^0 + 1}^{T_{b,v}^0 + \lfloor v/h \rfloor} \zeta_{Z,h,k}^* \right) \delta \\
&= \left| c_{p \times 1}' \left( \sum_{k=T_{b,v}^0}^{T_{b,v}^0 + \lfloor \bar{v}/h \rfloor} \zeta_{Z,h,k}^* \right) c_{p \times 1} + \delta' \left( \sum_{k=T_{b,v}^0 + 1}^{T_{b,v}^0 + \lfloor v/h \rfloor} \zeta_{Z,h,k}^* - \sum_{k=T_{b,v}^0 + 1}^{T_{b,v}^0 + \lfloor v/h \rfloor} \zeta_{Z,h,k}^* \right) \delta \right|
\end{align}
\[
\begin{align*}
&\leq K \left( \sum_{k=T_0^b+1}^{T_b(v)} \left\| \zeta_{T_b(v), k}^* \right\|_p \left\| c_{p \times 1} \right\|^2 + \sum_{k=T_0^b+1}^{T_b(v)} \left\| \zeta_{T_b(v), k}^* - \sum_{k=T_0^b+[v/h]} T_b(v) \left( \sum_{k=T_0^b+[v/h]} \zeta_{T_b(v), k}^* \right) \right\|_p \left\| \delta \right\|^2 \right) \\
&\leq K \left( \left( p c^2 \sum_{k=T_0^b+1}^{T_b(v)} \left\| \zeta_{T_b(v), k}^* \right\| + \sum_{k=T_0^b+[v/h]} T_b(v) \left( \sum_{k=T_0^b+[v/h]} \zeta_{T_b(v), k}^* \right) \right) \left\| \delta \right\|^2 \right).
\end{align*}
\]

By Itô’s formula \( \left\| \zeta_{T_b(v), k}^* \right\| = O \left( h^{3/2} \right), \) and so
\[
\left| \bar{Q}_{1,T} \left( \theta^* \right) - \bar{Q}_{1,T} \left( \tilde{\theta}^* \right) \right| \leq K \left( c^2 h^{-1} O_p \left( h^{3/2} \right) O (\tau) + \left\| \delta \right\|^2 h^{-1} O_p \left( h^{3/2} \right) O (\tau) \right) \\
\leq K \left( c^2 O_p \left( h^{1/2} \right) O (\tau) + \left\| \delta \right\|^2 O_p \left( h^{1/2} \right) O (\tau) \right),
\]

which goes to zero uniformly in \( \theta^* \in \Theta \) as \( r \to 0. \) Next, consider \( \bar{Q}_{2,T} \left( \theta^* \right) \) and observe that for any \( r \geq 1, \) standard estimates for Itô semimartingales yields \( \mathbb{E} \left( \left\| \zeta_{T_b(v), k}^* \right\|_p \left| \mathcal{F}_{k-1} \right. \right) \leq K_r h^r \). Then, by using a maximal inequality and choosing \( r > 2, \)
\[
\left( \mathbb{E} \left[ \sup_{(\theta, v) \in \mathcal{D}(C)} \left| \bar{Q}_{2,T} \left( \theta^* \right) \right| \right]^r \right)^{1/r} \leq K_r \left\| \delta \right\|^2 h^{-2/r} \leq K_r h^{1-2/r} \to 0,
\]
and thus we can use Markov’s inequality together with the latter result to verify that \( \bar{Q}_{2,T} \left( \theta^* \right) \) is stochastically equicontinuous. Turning to \( \bar{Q}_{4,T} \left( \theta^* \right), \)
\[
\begin{align*}
&\left| \bar{Q}_{4,T} \left( \theta^* \right) - \bar{Q}_{4,T} \left( \theta^* \right) \right| \\
&= \left| \delta' \left( h^{-1/2} \sum_{k=T_0^b+[v/h]} T_b(v) \left( \sum_{k=T_0^b+[v/h]} \zeta_{T_b(v), k}^* \right) \right) - \delta' \left( h^{-1/2} \sum_{k=T_0^b+[v/h]} T_b(v) \left( \sum_{k=T_0^b+[v/h]} \zeta_{T_b(v), k}^* \right) \right) \right| \\
&= \left| c_{p \times 1} \left( h^{-1/2} \sum_{k=T_0^b+[v/h]} T_b(v) \left( \sum_{k=T_0^b+[v/h]} \zeta_{T_b(v), k}^* \right) \right) \\
&+ \delta' \left( h^{-1/2} \sum_{k=T_0^b+[v/h]} T_b(v) \left( \sum_{k=T_0^b+[v/h]} \zeta_{T_b(v), k}^* \right) \right) \right| \\
&\leq K \left( h^{-1/2} \sum_{k=T_0^b+[v/h]} T_b(v) \left( \sum_{k=T_0^b+[v/h]} \zeta_{T_b(v), k}^* \right) \right) \left\| \delta \right\|^2 \\
&+ \left( h^{-1/2} \sum_{k=T_0^b+[v/h]} T_b(v) \left( \sum_{k=T_0^b+[v/h]} \zeta_{T_b(v), k}^* \right) \right) \left\| \delta \right\|^2 \\
&\leq K \left( p c h^{-1/2} \sum_{k=T_0^b+[v/h]} T_b(v) \left( \sum_{k=T_0^b+[v/h]} \zeta_{T_b(v), k}^* \right) \right) \left\| \delta \right\|^2 \\
&\leq K \left( p c h^{-1/2} \sum_{k=T_0^b+[v/h]} T_b(v) \left( \sum_{k=T_0^b+[v/h]} \zeta_{T_b(v), k}^* \right) \right) \left\| \delta \right\|^2.
\end{align*}
\]

By the Burkholder-Davis-Gundy inequality, \( \left\| \zeta_{T_b(v), k}^* \right\| \leq K h^{3/2} \) (recall \( \Sigma_{Z,t} = 0 \) for all \( t \geq 0 \), so that
\[
\left| \bar{Q}_{4,T} \left( \theta^* \right) - \bar{Q}_{4,T} \left( \theta^* \right) \right| \leq K \left( c^2 h^{-1/2} h^{-1} h^{3/2} O (\tau) \right)
\]

S-29
By a change in variable $v$ function $f$ $W$ $\epsilon > 0$ where

\[ \text{drift} N \]

Proof. S.D.4.8 Proof of Theorem 4.2

□

of Theorem 4.1 is complete.

In view of Section S.D.4.9, a result which shows the negligibility of the drift term, the proof then the main assertion of the theorem follows from the continuous mapping theorem for the argmax functional. In view of Section S.D.4.9, a result which shows the negligibility of the drift term, the proof of Theorem 4.1 is complete. □

S.D.4.8 Proof of Theorem 4.2

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

\[ Q_T (\theta, v) \xrightarrow{L^2} \]

\[ \mathcal{H} (v) = \begin{cases} - (\delta^0)' (Z, Z_1) (v) \delta^0 + 2 \left( (\delta^0)' \Omega_{W,1} (\delta^0) \right)^{1/2} W_1^* (v), & \text{if } v \leq 0 \\ - (\delta^0)' (Z, Z_2) (v) \delta^0 + 2 \left( (\delta^0)' \Omega_{W,2} (\delta^0) \right)^{1/2} W_2^* (v), & \text{if } v > 0. \end{cases} \]

By a change in variable $v = \vartheta^{-1} s$ with $\vartheta = \left( (\delta^0)' (Z, Z_1) (\delta^0) \right)^2 / (\delta^0)' \Omega_{W,1} (\delta^0)$, we can show that

\[ \text{argmax} v \in \left[ N_{\pi-N_0_b-N_0 (1-\pi-N_0) b} \right] \mathcal{H} (v) \]

\[ \overset{d}{=} \text{argmax} s \in \left[ N_{\pi-N_0_b-N_0 (1-\pi-N_0) b} \right] \mathcal{Y} (s), \]

where

\[ \mathcal{Y} (s) = \begin{cases} -\frac{|s|}{2} + W_1^* (s), & \text{if } s < 0 \\ -\frac{(\delta^0)' (Z, Z_2) \delta^0 |s|}{(\delta^0)' (Z, Z_1) \delta^0} + \left( (\delta^0)' \Omega_{W,2} (\delta^0) \right)^{1/2} W_2^* (s), & \text{if } s \geq 0, \end{cases} \]

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

\[ \text{argmax} v \in \left[ N_{\pi-N_0_b-N_0 (1-\pi-N_0) b} \right] \mathcal{H} (v) \]
\[
\arg\max_{s \in \left[ N^{1-s-N_0^{\theta}} \left( \frac{(\delta^0)^t}{\delta^0} \right)^2 \frac{N(1-s-N_0^{\theta})}{\|\delta^0\|^{-2\pi^2} (\delta^0)^t \Omega_{\Omega}^{-1}(\delta^0)} \right]^{-1} \mathcal{Y}(s),
\]
and finally by the continuous mapping theorem for the argmax functional,
\[
\arg\max_{s \in \left[ N^{1-s-N_0^{\theta}} \left( \frac{(\delta^0)^t}{\delta^0} \right)^2 \frac{N(1-s-N_0^{\theta})}{\|\delta^0\|^{-2\pi^2} (\delta^0)^t \Omega_{\Omega}^{-1}(\delta^0)} \right]^{-1} \mathcal{Y}(s).
\]

This concludes the proof. \(\square\)

S.D.4.9 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 \(\mu_x\) 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
\[
d\psi_{s} = \psi_{h^{-1/2}} \mu_{Z,s} ds + \psi_{h^{-1/2}} \sigma_{Z,s} dW_{Z,s}, \quad dW_{\psi,s} = \psi_{h^{-1/2}} \sigma_{e,s} dW_{e,s}, \quad (S.68)
\]

with \(s \in \mathcal{D}(\mathcal{C})\). Given \(s \mapsto t = \psi_{h^{-1}} s\), we have \(\psi_{h^{-1}} \mu_{Z,s} ds = \psi_{h^{-1}} \mu_{Z,s} dt + \sigma_{Z,s} dW_{Z,s}\) and
\[
\psi_{h^{-1}} \sigma_{Z,s} dW_{Z,s} = \psi_{h^{-1}} \sigma_{e,s} dW_{e,s} + \psi_{h^{-1}} \sigma_{Z,s} dW_{Z,s}.
\]

Thus, the change of time scale effectively makes the drift \(\mu_{Z,s} ds\) of even higher order. We show a stronger result in that we demonstrate its negligibility even in the case \(\vartheta = 0\); hence, we show that the limit law of (S.63) remains the same when \(\mu_{x}\) are nonzero. We set for any \(1 \leq i \leq p\) and \(1 \leq j \leq q+1\), \(\mu_{Z,i} \triangleq \int_{(k-1)h}^{kh} \mu_{x,i} ds\), \(\mu_{X,i} \triangleq \int_{(k-1)h}^{kh} \mu_{X,i} ds\), \(z_{0,kh} \triangleq \sum_{r=1}^{q} \int_{(k-1)h}^{kh} \sigma_{X,i}^2 dW_{Z}^{(r)} x_{0,kh} \triangleq \sum_{r=1}^{q+p} \int_{(k-1)h}^{kh} \sigma_{X,i}^2 dW_{X}^{(r)}\).

Recall that \(\mu_{x,k} \) is \(O(h)\) uniformly in \(k\), and note that \(\mu_{Z,k} x_{0,kh} + \mu_{Z,k} z_{0,kh} \) follows a Gaussian law with zero mean and variance of order \(O(h^3)\). Also note that on \(\mathcal{D}(\mathcal{C})\), \(T_{b}^{0} - b - 1 < 1/h\), where \(a_{h} < b_{h}\) if for some \(c \geq 1\), \(b_{h}/c < a_{h} \leq c b_{h}\). Then,
\[
\sum_{k=T_{b}^{0}}^{T_{b}^{0}} z_{0,kh} x_{0,kh} = \sum_{k=T_{b}^{1}}^{T_{b}^{0}} \mu_{Z,k}^0 \mu_{X,k} + \sum_{k=T_{b}^{0}}^{T_{b}^{0}} \mu_{Z,k} x_{0,kh}^0 + \sum_{k=T_{b}^{0}}^{T_{b}^{0}} \mu_{Z,k} z_{0,kh}^0 + \sum_{k=T_{b}^{0}}^{T_{b}^{0}} z_{0,kh} x_{0,kh}^0 = o(h^{1/2}) + o_h(h^{1/2}) + \sum_{k=T_{b}^{0}}^{T_{b}^{0}} z_{0,kh} x_{0,kh}^0.
\]
Therefore, conditionally on $\Sigma^0 = \{\mu_{s,t}, \sigma_{s,t}\}_{t \geq 0}$, the limit law of

$$Q_T (\theta^*) = - \left(\delta^0\right) \left( \sum_{k=T_b+1}^{T_b^0} z_{kh} z_{kh}^I \right) \delta^0 + 2 \left(\delta^0\right) \left( h^{-1/2} \sum_{k=T_b+1}^{T_b^0} z_{0,kh} \tilde{e}_{kh} \right),$$

is the same as the limit law of

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

which completes the proof of Theorem 4.1. □

S.D.4.10 Proof of Proposition 4.2

Proof. By Lemma 4.1,

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

Divide both sides by $h$ to yield,

$$h^{-1} \left( Q_T (T_b) - Q_T (T_b^0) \right) = - h^{1/2} \left( \delta^0 \right) \left( \frac{Z' e}{\sqrt{h}} \right) + o_p \left( h^{1/2-\kappa} \right).$$

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

$$h^{-1+\kappa/2} \left( Q_T (T_b) - Q_T (T_b^0) \right) = - h^{1/2} \left( \delta^0 \right) \left( \frac{Z' e}{\sqrt{h}} \right) + o_p \left( h^{1/2-\kappa/2} \right).$$

Note that $T_b = T_b^0 + \lfloor vT^\kappa \rfloor$. Then,

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

The continuous mapping theorem then yields the desired result. □

S.D.4.11 Proof of Proposition 5.1

Proof. Replace $\xi_1, \xi_2, \rho$ and $\vartheta$ in (4.7) by their corresponding estimates $\xi_1, \xi_2, \rho$ and $\vartheta$, respectively. Multiply both sides of (4.7) by $h^{-1}$ and apply a change in variable $v = s/h$. Consider the case $s < 0$. On the “fast time scale” $W^*$ is replaced by $W_{1,h} (s) = W_{1,h}^* (sh)$ $(s < 0)$ where $W_{1,h}^* (s)$ is a sample-size dependent Wiener process. It follows that

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

S-32
A similar argument can be applied for \( s \geq 0 \). Let \( \hat{Y}(s) \) denote our estimate of \( Y(s) \) constructed with the proposed estimates in place of the population parameters. Then,

\[
\hat{Y}(s) = \arg\max_{v \in \left[\pi - \lambda_0 \vartheta, (1 - \pi - \lambda_0) \vartheta \right]} \hat{Y}(v),
\]

which is equal to the right-hand side of (4.7). Recall that

\[
\psi = \frac{\delta^0}{\alpha_0} \times \left( \left( \delta^0 \right)^T \left( Z, Z^\prime, \delta^0 \right) \right)^2 / \left( \left( \delta^0 \right)^T \Omega_{\psi,1} \left( \delta^0 \right) \right).
\]

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

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

The steps are similar to those used for the case when the model does not include predictable processes. However, we need to rely occasionally on different asymptotic results since the latter processes have distinct statistical properties. Recall that the dependent variable \( \Delta_h Y_k \) in model (S.2) is the increment of a discretized process which cannot be identified as an ordinary diffusion. However, its normalized version, \( \tilde{Y}_{(k-1)h} \), 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.\(^3\) 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:

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

by a standard approximation for Riemann sums and recalling that \( \tilde{Y}_{(k-1)h} \) is scaled to be \( O_p(1) \). Then it is easy to see that \( \left\{ \tilde{Y}_{(k-1)h} \right\}_{k=1}^T \) 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.D.1.

**Assumption S.D.2.** Assumption S.C.1 holds, the process \( \left\{ \tilde{Y}_t, D_t, Z_t \right\}_{t \geq 0} \) takes value in some compact set and the processes \( \left\{ \mu_t, \sigma_t \right\}_{t \geq 0} \) (except \( \left\{ \mu^h_t \right\}_{t \geq 0} \)) are bounded.

Recall the notation \( M = I - X (X^\prime X)^{-1} X^\prime \), where now

\[
X = \begin{bmatrix}
  h^{1/2} & Y_0 h & \Delta_h D_1^l & \Delta_h Z_1^l \\
h^{1/2} & Y_1 h & \Delta_h D_2^l & \Delta_h Z_2^l \\
  \vdots & \vdots & \vdots & \vdots \\
h^{1/2} & Y_{T h} h & \Delta_h D_T^l & \Delta_h Z_T^l
\end{bmatrix}_{T \times (q+p+2)}.
\]

\(^3\)For an introduction to the terminology used in this sub-section, we refer the reader to first chapters in Jacod and Shiryaev (2003).
Thus, $X'X$ is a $(q + p + 2) \times (q + p + 2)$ matrix given by $\begin{bmatrix} a_1 & a_2 & a_3 & a_4 \end{bmatrix}$, where

$$ a_1 = \left[ \begin{array}{c} \sum_{k=1}^{T} h \frac{1}{2} \sum_{j=1}^{T} (Y_{(k-1)h})_{hk} \\ \sum_{k=1}^{T} h \frac{1}{2} (\Delta_{h} D_k) \\ \sum_{k=1}^{T} h \frac{1}{2} (\Delta_{h} Z_k) \end{array} \right], \quad a_2 = \left[ \begin{array}{c} h^{1/2} \sum_{k=1}^{T} (Y_{(k-1)h})_{hk} \\ \sum_{k=1}^{T} (\Delta_{h} D_k) (Y_{(k-1)h})_{hk} \\ \sum_{k=1}^{T} (\Delta_{h} Z_k) (Y_{(k-1)h})_{hk} \end{array} \right], $$

$$ a_3 = \left[ \begin{array}{c} \sum_{k=1}^{T} h \frac{1}{2} (\Delta_{h} D_k)^{\prime} \\ \sum_{k=1}^{T} (\Delta_{h} D_k) (Y_{(k-1)h})_{hk} \\ X'_D X_D \\ X'_Z X_Z \end{array} \right], \quad a_4 = \left[ \begin{array}{c} \sum_{k=1}^{T} h \frac{1}{2} (\Delta_{h} Z'_k) \\ \sum_{k=1}^{T} (\Delta_{h} Z_k) (Y_{(k-1)h})_{hk} \\ X'_D X_Z \\ X'_Z X_Z \end{array} \right], $$

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

$$ X'e = \left[ \begin{array}{c} \sum_{k=1}^{T} h^{1/2} e_{kh} \\ \sum_{k=1}^{T} (Y_{(k-1)h})_{hk} e_{kh} \\ \sum_{k=1}^{T} \Delta_{h} D_k e_{kh} \\ \sum_{k=1}^{T} \Delta_{h} Z_k e_{kh} \end{array} \right]. \quad (S.70) $$

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.D.10.** For $1 \leq i \leq 2$, $3 \leq j \leq p + 2$ and $\gamma > 0$, the following estimates are asymptotically negligible: $\sum_{k=[s/h]}^{[t/h]} z^{(1)}_{kh} \sum_{k=[s/h]}^{[t/h]} z^{(j)}_{kh} u \rightarrow \rho_{n} \Rightarrow 0$, for all $N > t > s + \gamma > s > 0$.

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

$$ P \left( \sum_{k=[s/h]}^{[t/h]} | z^{(1)}_{kh} z^{(j)}_{kh} |^2 > \eta \right) \leq \frac{K}{\eta} \mathbb{E} \left( \sum_{k=[s/h]}^{[t/h]} h^2 \left( \Delta_{h} M^{(j)}_{Z,k} \right)^2 \right) \leq \frac{K}{\eta} hO_p(t-s) \rightarrow 0, $$

where the second inequality follows from the Burkholder-Davis-Gundy inequality with parameter $r = 2$. This shows that the array $\left\{ z^{(i)}_{kh} z^{(j)}_{kh} \right\}$ 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 $z^{(2)}_{kh} z^{(i)}_{kh} = (Y_{(k-1)h})_{hk} (\Delta_{h} M^{(j)}_{Z,k})$, and recall that $\tilde{Y}_{(k-1)h} = h^{1/2} Y_{(k-1)h} = O_p(1)$. Thus, the same proof remains

S-34
valid for the case \( i = 2 \). The assertion of the lemma follows. \( \square \)

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

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

\[
Q_T(T_b) - Q_T(T_0) = -|T_b - T_0^0| d(T_b) + g_e(T_b), \quad \text{for all} \ T_b, \quad (S.71)
\]

where

\[
d(T_b) \triangleq \frac{(\delta^0)' \left\{ (Z_0'MZ_0) - (Z_0'MZ_2)(Z_2'MZ_2)^{-1}(Z_2'MZ_0) \right\} \delta^0}{|T_b - T_0^0|}, \quad (S.72)
\]

and we arbitrarily define \( d(T_b) = (\delta^0)' \delta^0 \) when \( T_b = T_0^0 \). Let \( d_T = T \inf_{|T_b - T_0^0| > TK} d(T_b) \); it is positive and bounded away from zero by Lemma S.D.11 below. Then

\[
P \left( |\hat{\lambda}_b - \lambda_0| > K \right) = P \left( |\hat{T}_b - T_0^b| > TK \right)
\]

\[
\leq P \left( \sup_{|T_b - T_0^b| > TK} |g_e(T_b)| \geq \inf_{|T_b - T_0^b| > TK} |T_b - T_0^b| d(T_b) \right)
\]

\[
\leq P \left( \sup_{p+2 \leq T_b \leq T - p - 2} |g_e(T_b)| \geq TK \inf_{|T_b - T_0^b| > TK} d(T_b) \right)
\]

\[
= P \left( d_T^{-1} \sup_{p+2 \leq T_b \leq T - p - 2} |g_e(T_b)| \geq K \right). \quad (S.73)
\]

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

\[
2 \left( \delta^0 \right)' (Z_0'MZ_2)(Z_2'MZ_2)^{-1/2} (Z_2'MZ_2)^{-1/2} Z_2 M e. \quad (S.74)
\]

For the stochastic regressors, Theorem S.D.5 implies that for any \( 3 \leq j \leq p + 2, \ (Z_2 e)_{j,1}/\sqrt{h} = O_p(1) \) and for any \( 3 \leq i \leq q + p + 2, \ (X e)_{i,1}/\sqrt{h} = O_p(1) \), 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,

\[
\sup_{T_b} (Z_0'MZ_2)(Z_2'MZ_2)^{-1} (Z_2'MZ_0) \leq Z_0'MZ_0 = O_p(1),
\]

by Lemma S.D.3. Next, note that the first two elements of the vector \( X' e \) and \( Z_2'e \) are \( O_p \left( h^{1/2} \right) \) [recall (S.70)]. By Assumption 2.1-(iii) and the inequality

\[
\sup_{T_b} \| (Z_2'MZ_2)^{-1/2} Z_2 M e \| \leq \sup_{T_b} \| (Z_2'MZ_2)^{-1/2} \| \sup_{T_b} \| Z_2 M e \|,
\]

we have that \( (Z_2'MZ_2)^{-1/2} Z_2 M e \) 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.74) 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.D.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 \( \hat{\lambda}_b \overset{P}{\to} \lambda_0. \) \( \square \)
Assuming Under Assumption 2.1-(iii) and in view of (S.69),

Let Lemma S.D.1.

The result follows choosing $X$ such that

$$ \text{Lemma S.D.11.} \quad \text{Let } B > 0 \text{ and following the same steps as in Lemma S.D.6 (but replacing } R \text{ by } R \text{)}$$

$$ Td(T_b) \geq T \left( \delta^0 \right)^{\prime} R \frac{X'_{\Delta} X_{\Delta}}{T_0^b - T_b} (X'_{0}X_0) \text{ R} \left( \delta^0 \right)$$

$$ = \left( \delta^0 \right)^{\prime} R \frac{X'_{\Delta} X_{\Delta}}{B} (X'_{2}X_2) -1 (X'_{0}X_0) \text{ R} \left( \delta^0 \right).$$

Under Assumption 2.1-(iii) and in view of (S.69), $X'_{\Delta} X_{\Delta}$ is positive definite: for the $p \times p$ lower-right sub-block apply Lemma S.D.3 as in the proof of Lemma S.D.6, whereas for the remaining elements of $X'_{\Delta} X_{\Delta}$ the result follows from the convergence of approximations to Riemann sums. Note that $X'_{2}X_2$ and $X'_{0}X_0$ are $O_p(1)$. It follows that

$$ Td(T_b) \geq \left( \delta^0 \right)^{\prime} R \frac{X'_{\Delta} X_{\Delta}}{N} (X'_{2}X_2) -1 (X'_{0}X_0) \text{ R} \delta^0 \geq \kappa > 0.$$ 

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

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

$$ D_{K,T} = \{ T_b : N\eta \leq N_0 \leq N(1 - \eta), |N_0 - N_b| > KT^{-1} \},$$

and observe that it is enough to show that $P \left( \sup_{T_b \in D_{K,T}} Q_T(T_b) \geq Q_T(T_b) \right) < \varepsilon$, or

$$ P \left( \sup_{T_b \in D_{K,T}} h^{-1} g_e(T_b) \geq \inf_{T_b \in D_{K,T}} h^{-1} |T_b - T_b| d(T_b) \right) < \varepsilon. \quad (S.75)$$

By Lemma S.D.1,

$$ \inf_{T_b \in D_{K,T}} d(T_b) \geq \inf_{T_b \in D_{K,T}} \left( \delta^0 \right)^{\prime} R \frac{X'_{\Delta} X_{\Delta}}{T_0^b - T_b} (X'_{2}X_2) -1 (X'_{0}X_0) \text{ R} \delta^0.$$

For the $(q + p) \times (q + p)$ lower right sub-block of $X'_{\Delta} X_{\Delta}$ the arguments of Proposition 3.2 apply: $\left( h(T_0^b - T_b) \right)^{-1}$

$[X'_{\Delta} X_{\Delta}]_{i \times j, (q + p) \times (q + p)}$ is bounded away from zero for all $T_b \in D_{K,T}$ by choosing $K$ large enough (recall $|T_0^b - T_b| > K$), where $[A]_{i \times j}$ 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

$$ \sum_{i=0}^{T_0^b - T_b} \left( h^{1/2} \right)^{2} \text{ Thus } \left( h(T_0^b - T_b) \right)^{-1} \sum_{k=T_b}^{T_0^b} \left( h^{1/2} \right)^{2} > 0.$$ 

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

$$ \inf_{T_b \in D_{K,T}} Td(T_b) \text{ is bounded away from zero. It remains to show that } \sup_{T_b \in D_{K,T}} (h(T_0^b - T_b)^{-1} g_e(T_b) \text{ is small when } T \text{ is large. Recall that the terms } Z_2 \text{ and } Z_0 \text{ involve a positive fraction } N\eta \text{ of the data. We can apply Lemma S.D.3 to those elements which involve the stochastic regressors only, whereas the other terms are dealt with directly using the definition of } X' \text{ in (S.70). Consider the first term of } g_e(T_b).$$

Using the same steps which led to (S.19), we have

$$ \left| 2 \left( \delta^0 \right)^{\prime} (Z_0'MZ_2)(Z_2'MZ_2)^{-1} Z_2M_e - 2 \left( \delta^0 \right)^{\prime} (Z_0'Me) \right|$$

$$ = \left| \left( \delta^0 \right)^{\prime} Z'_\Delta Me + \left| \left( \delta^0 \right)^{\prime} (Z'_\Delta M Z_2)(Z'_2 M Z_2)^{-1} (Z_2M_e) \right| \right. \quad (S.76)$$

We can apply Lemma S.D.3 to the terms that do not involve $|N_0 - N_b|$ but only stochastic regressors.
Next consider the first term of

\[
\left( h \left( T^0_b - T_b \right) \right)^{-1} (\delta^0)' \left( Z'_\Delta M Z_2 \right) = \frac{(\delta^0)' \left( Z'_\Delta Z_\Delta \right)}{h \left( T^0_b - T_b \right)} - \frac{(\delta^0)' \left( Z'_\Delta X_\Delta \right)}{h \left( T^0_b - T_b \right)} \left( X'X \right)^{-1} X'Z_2).
\]

Applying the same manipulations as those used above for the \( p \times p \) lower right sub-block of \( Z'_\Delta Z_\Delta \), we have \( (h \left( T^0_b - T_b \right))^{-1} \left[ Z'_\Delta Z_\Delta \right]_{(\cdot, p \times p)} = O_p(1) \), since there are \( T^0_b - T_b \) 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 \( Z'_\Delta X_\Delta \left/ (h \left( T^0_b - T_b \right)) \right. \). Hence, as \( h \downarrow 0 \) the second term above is \( O_p(1) \). Next, consider the upper left 2 \( \times \) 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 \( h^{1/2}Y_{(k-1)h} = \tilde{Y}_{(k-1)h} = O_p(1) \) by Assumption S.D.2),

\[
\sum_{k=T_{b}+1}^{T^0_b} \left( Y_{(k-1)h} \right)^2 = \sum_{k=T_{b}+1}^{T^0_b} \left( \tilde{Y}_{(k-1)h} h^{1/2} \right)^2 \leq C \sum_{k=T_{b}+1}^{T^0_b} h,
\]

where \( C = \sup_k \left| \tilde{Y}_{(k-1)h} \right| \) is a fixed constant given the localization in Assumption S.D.2. Thus, when multiplied by \( (h \left( T^0_b - T_b \right))^{-1} \), 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 \( 3 \leq j \leq p + 2 \),

\[
\frac{1}{h \left( T^0_b - T_b \right)} \sum_{k=T_{b}+1}^{T^0_b} \tilde{z}_{kh} z_{kh} = \frac{1}{h \left( T^0_b - T_b \right)} \sum_{k=T_{b}+1}^{T^0_b} \left( \tilde{Y}_{(k-1)h} h^{1/2} \right) \tilde{z}_{kh} = \frac{1}{T^0_b - T_b} \sum_{k=T_{b}+1}^{T^0_b} \tilde{Y}_{(k-1)h} \tilde{z}_{kh} \sqrt{\frac{h}{h}}.
\]

Since \( \tilde{z}_{jkh} / \sqrt{h} \) is i.i.d. with zero mean and finite variance and \( \tilde{Y}_{(k-1)h} \) is \( O_p(1) \) by Assumption S.D.2, Assumption 3.1 implies that we can find a \( K \) large enough such that the right hand side is \( O_p(1) \) uniformly in \( T_b \). The same argument applies to \( (Z'_\Delta Z_\Delta)_{1,j} \), \( 3 \leq j \leq p + 2 \). This shows that the term \( (Z'_\Delta X_\Delta \left/ (h \left( T^0_b - T_b \right)) \right. \left( X'X \right)^{-1} X'Z_2 \) is bounded and so is \( Z'_\Delta X_\Delta \left/ (h \left( T^0_b - T_b \right)) \right. \) using the same reasoning. Thus, \( (h \left( T^0_b - T_b \right))^{-1} \delta^0)' \left( Z'_\Delta M Z_2 \right) = O_p(1) \). By the same arguments as before, we can use Theorem S.D.5 to show that the second term of (S.76) is \( O_p \left( h^{1/2} \right) \) when multiplied by \( (h \left( T^0_b - T_b \right))^{-1} \) since the last term involves a positive fraction of the data. Now, expand the \((p + 2)\)-dimensional vector \( Z'_\Delta M e \) as

\[
\frac{Z'_\Delta M e}{h \left( T^0_b - T_b \right)} = \frac{1}{h \left( T^0_b - T_b \right)} \sum_{k=T_{b}+1}^{T^0_b} \tilde{z}_{kh} e_{kh} - \frac{1}{h \left( T^0_b - T_b \right)} \left( \sum_{k=T_{b}+1}^{T^0_b} \tilde{z}_{kh} x_{kh} \right) \left( X'X \right)^{-1} (X'e).
\]
The arguments for the last $p$ elements are the same as above and yield [recall (S.20)]

\[
\frac{[Z'_\Delta Me]_{\{p\}}}{h (T^0_b - T_b)} = o_p \left( K^{-1} \right) - O_p \left( 1 \right) O_p \left( h^{1/2} \right),
\]

where we recall that by Assumption 2.1-(iv) $\Sigma_{Z_{e,N}^0} = 0$. Note that the convergence is uniform over $T_b$ by Lemma S.D.2. We now consider the first two elements of $Z'_\Delta e$:

\[
\sum_{k=T_b+1}^{T_b^0} z_{kh} e_{kh} \leq \sum_{k=T_b+1}^{T_b^0} h^{1/2} y_{(k-1)h} h^{1/2} e_{kh} \leq A \sum_{k=T_b+1}^{T_b^0} \left| \bar{y}_{(k-1)h} h^{1/2} e_{kh} \right|,
\]

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

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

where $C = \sup_k \left| \bar{y}_{(k-1)h} \right|$ is finite by Assumption S.D.2. Choose $K$ large enough such that the probability that the right-hand side is larger than $B/3N$ is less than $\varepsilon$. For the first element of $Z'_\Delta e$ the argument is the same and thus $P \left( \left( h \left( T^0_b - T_b \right) \right)^{-1} \sum_{k=T_b+1}^{T_b^0} z_{kh} \left| e_{kh} \right| > B/3N \right) \leq \varepsilon$, 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

\[
\sum_{k=1}^{T} x_{kh} \left( 1 \right) e_{kh} = \sum_{k=1}^{T} h^{1/2} e_{kh} = h^{1/2} O_p \left( 1 \right),
\]

(S.77)

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

\[
\frac{1}{h \left( T^0_b - T_b \right)} g_e \left( T_b \right) = \frac{1}{h \left( T^0_b - T_b \right)} \left( \delta^{(0)} \right)' 2Z'_\Delta e + O_p \left( 1 \right).
\]

(S.78)

In order to prove

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

we can use (S.78). To this end, we shall find a $K > 0$, such that

\[
P \left( \sup_{T_b \leq T^0_b - \frac{K}{N}} \left( T_b^0 - T_b \right)^{-1} \sum_{k=T_b+1}^{T_b^0} \left| z_{kh} \left( 1 \right) e_{kh} \right| > \frac{B}{3N} \right) \leq P \left( \sup_{T_b \leq T^0_b - \frac{K}{N}} \left( T_b^0 - T_b \right)^{-1} \sum_{k=T_b+1}^{T_b^0} \frac{e_{kh}}{\sqrt{h}} > \frac{B}{6 |\mu_0^b| N} \right) < \varepsilon/3.
\]

(S.79)

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

\[
P \left( \sup_{T_b \leq T^0_b - \frac{K}{N}} \left( T_b^0 - T_b \right)^{-1} \left| \sum_{k=T_b+1}^{T_b^0} \frac{e_{kh}}{\sqrt{h}} \right| > \frac{B}{6 |\mu_0^b| N} \right) \leq A \frac{36 \left( \mu_0^b \right)^2 N^2}{B^2} \frac{1}{KN^{1.2}}.
\]
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),

$$P \left( \sup_{T_b \leq T^0_b - \frac{K}{N}} \frac{1}{h (T^0_b - T_b)} \left\| 2 \left( \delta^0 \right)' \sum_{k=T_b+1}^{T^0_b} [Z'_{\Delta}e]_{\{k\}} \right\| > \frac{B}{3N} \right) < \frac{\varepsilon}{3},$$

since for each $j = 3, \ldots, p$, $\left\{ z_{kh} e_{kh} / h \right\}$ is i.n.d. with finite variance, and thus the result is implied by the Hájek-Rényi inequality for large $K$. Using the latter results into (S.78), we have

$$P \left( \sup_{T_b \leq T^0_b - \frac{K}{N}} \frac{1}{h (T^0_b - T_b)} \left\| 2 \left( \delta^0 \right)' \sum_{k=T_b+1}^{T^0_b} z_{kh} e_{kh} \right\| > \frac{B}{N} \right) < \varepsilon,$$

which verifies (S.75) and thus proves our claim. □

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

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

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

$$\left( Q_T (T_b) - Q_T \left( T^0_b \right) \right) / \psi_h = -\delta_h \left( Z'_{\Delta}Z_{\Delta} / \psi_h \right) \delta_h \pm 2 \delta_h \left( Z'_{\Delta}\bar{e} / \psi_h \right) + O_p \left( h^{3/4\Lambda_1-\kappa/2} \right).$$

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

$$Q_T (T_b) - Q_T \left( T^0_b \right) = -\delta_h \left\{ Z'_{\Delta}MZ_{\Delta} + (Z'_{\Delta}MZ_2) (Z'_{\Delta}MZ_2)^{-1} (Z'_{\Delta}MZ_2) \right\} \delta_h$$

$$+ g_e (T_b, \delta_h).$$

We can expand the first term of (S.80) as

$$\delta_h Z'_{\Delta}MZ_{\Delta} \delta_h = \delta_h Z'_{\Delta}Z_{\Delta} \delta_h - \delta_h A \delta_h,$$

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.D.10 provides the formal statement of the result for $Z'_{\Delta}Z_{\Delta}$. Hence, the result carries over to $Z'_{\Delta}X$ with no changes. By symmetry so is the $2 \times p$ upper-right block. This allows us to treat the $2 \times 2$ upper-left block and the $p \times p$ lower-right block of statistics such as $A$ separately. By Lemma S.D.3, $(X'X)^{-1} = O_p (1)$. Using Proposition 4.1-(ii), we let $N_b - N^0_b = K \psi_h$. By the Burkholder-Davis-Gundy inequality, we have standard estimates for local volatility so that

$$\left\| \mathbb{E} \left( \bar{\Sigma}_{ZX} (T_b, T^0_b) - \Sigma_{ZX, (T^0_b-1)h} (T^0_b-1)h \right) \right\| \leq Kh^{1/2},$$

with $3 \leq i \leq p+2$ and $3 \leq j \leq q+p+2$ which in turn implies $[Z'_{\Delta}X_{\Delta}]_{\{i,p+q\}} = O_p \left( 1 / (h (T^0_b - T_b)) \right)$. 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

$$(Z'_\Delta X)_{2,2} = \sum_{k=T_b+1}^{T_b^0} z_{kh}^{(2)} x_{kh}^{(2)} = \sum_{k=T_b+1}^{T^0_b} \left( \bar{Y}_{(k-1)h} \right)^2 h.$$  

By a localization argument (cf. Assumption S.D.2), $\bar{Y}_{(k-1)h}$ is bounded. Then, since the number of summands grows at a rate $T^\kappa$, we have $(Z'_\Delta X)_{2,2} = O_p \left( Kh^{1-\kappa} \right)$. The same proof can be used for $(Z'_\Delta X)_{1,1}$, which gives $(Z'_\Delta X)_{1,1} = O_p \left( Kh^{1-\kappa} \right)$. Thus, in view of (S.82), we conclude that (S.81) when divided by $\psi_h$ is such that

$$\delta'_h Z'_\Delta M Z_\Delta \delta_h / \psi_h = \delta'_h Z'_\Delta X \delta_h / \psi_h - \delta'_h Z'_\Delta X (X'X)^{-1} X'Z_\Delta \delta_h / \psi_h$$

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

(S.82)

For the second term of (S.80), we have

$$\psi_h^{-1} h^{1/2} \left( \delta^0 \right)' \left\{ (Z'_\Delta M Z_2) (Z'_2 M Z_2) (Z'_2 M Z_\Delta) \right\} \delta^0$$

$$= \psi_h^{-1} h^{1/2} \| \delta_0 \|^2 O_p (\psi_h) O_p (1) O_p (\psi_h) \leq K \psi_h^{-1} h^{1/2} O_p \left( h^{2(1-\kappa)} \right)$$

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

$$g_e (T_b, \delta_h) = 2\delta'_h \left( Z'_0 M Z_2 \right) (Z'_2 M Z_2)^{-1} Z_2 M \epsilon - 2\delta'_h \left( Z'_0 M \epsilon \right)$$

$$+ e' M Z_2 (Z'_2 M Z_2)^{-1} Z_2 M \epsilon - Z_0 (Z'_0 M Z_0)^{-1} Z'_0 M \epsilon.$$  

(S.84)

Recall (S.70), and $\sum_{k=T_b+1}^{T^0_b} x_{kh} e_{kh} = h^{-1/4} \sum_{k=T_b+1}^{T^0_b} x_{kh} e_{kh}$. Introduce the following decomposition,

$$(X' e)_{2,1} = \sum_{k=1}^{T^0_b-[T^\kappa]} x_{kh}^{(2)} e_{kh} + h^{-1/4} \sum_{k=T^0_b-[T^\kappa]+1}^{T^0_b+[T^\kappa]} x_{kh}^{(2)} e_{kh} + \sum_{k=T^0_b+[T^\kappa]+1}^{T} x_{kh}^{(2)} e_{kh},$$

where $\bar{e}_{kh} \sim \text{i.n.d.} \mathcal{N} \left( 0, \sigma^2_{-1-kh} \right)$. The first and third terms are $O_p \left( h^{1/2} \right)$ in view of (S.77). The term in the middle is $h^{3/4} \sum_{k=T^0_b-[T^\kappa]+1}^{T^0_b+[T^\kappa]} \bar{Y} (k-1)h h^{-1/2} \epsilon_{kh}$, which involves approximately $2T^\kappa$ summands. Since $\bar{Y} (k-1)h$ is bounded by the localization procedure,

$$h^{3/4} \sum_{k=T^0_b-[T^\kappa]+1}^{T^0_b+[T^\kappa]} \bar{Y} (k-1)h^{3/4} \epsilon_{kh} = h^{3/4} T^\kappa \epsilon_{kh} / \sqrt{h} = h^{3/4} T^\kappa / 2 O_p (1),$$

or $h^{-1/4} \sum_{k=T^0_b-[T^\kappa]}^{T^0_b+[T^\kappa]} x_{kh}^{(2)} \epsilon_{kh} = h^{3/4-\kappa/2} O_p (1)$. 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 $Z'_0 M Z_2 = Z'_2 M Z_2 + Z'_\Delta M Z_2$, we have

$$2h^{1/4} \left( \delta^0 \right)' \left( Z'_0 M Z_2 \right) (Z'_2 M Z_2)^{-1} Z_2 M \epsilon - 2h^{1/4} \left( \delta^0 \right)' \left( Z'_0 M \epsilon \right)$$

$$= 2h^{1/4} \left( \delta^0 \right)' Z'_\Delta M \epsilon + 2h^{1/4} \left( \delta^0 \right)' Z'_2 M Z_2 (Z'_2 M Z_2)^{-1} Z'_2 M \epsilon.$$  

(S.85)
The last term above when multiplied by $ψ_h^{-1}$ is such that

$$ψ_h^{-1} 2h^{1/4} \left( δ_0 \right)' Z_Δ M Z_2 (Z_2' M Z_2)^{-1} Z_2' M e = \left\| δ_0 \right\| O_p \left( h^{1/5}/4 - \kappa/2 \right),$$

where we have used the fact that $Z_Δ M Z_2 / ψ_h = O_p \left( 1 \right)$. For the first term of (S.85),

$$2h^{1/4} \left( δ_0 \right)' Z_Δ M e / ψ_h$$

$$= 2h^{1/4} \left( δ_0 \right)' Z_Δ e / ψ_h - 2h^{1/4} \left( δ_0 \right)' Z_Δ X (X' X)^{-1} X' e / ψ_h$$

$$= 2h^{1/4} \left( δ_0 \right)' Z_Δ e - 2 \left( δ_0 \right)' O_p \left( 1 \right) O_p \left( h^{1/5}/4 - \kappa/2 \right).$$

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

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

This leads to

$$g_e \left( T_b, δ_h \right) / ψ_h = 2h^{1/4} \left( δ_0 \right)' Z_Δ e / ψ_h + O_p \left( h^{3/4}/4 - \kappa/2 \right)$$

$$+ \left\| δ_0 \right\| O_p \left( h^{3/4}/4 - \kappa/2 \right) + o_p \left( h^{1/2} \right),$$

for sufficiently small $c_T$. This together with (S.82) and (S.83) yields,

$$ψ_h^{-1} \left( Q_T \left( T_b \right) - Q_T \left( T_b^0 \right) \right) = -δ_h \left( Z_Δ Z / ψ_h \right) δ_h$$

$$± 2δ_h \left( Z_Δ e / ψ_h \right) + O_p \left( h^{3/4}/4 - \kappa/2 \right) + o_p \left( h^{1/2} \right),$$

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

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

$$dZ^{(1)}_{ψ,s} = ψ^{-1/2} (ds)^{1/2} = (ds)^{1/2},$$

$$dZ^{(2)}_{ψ,s} = ψ^{-1/2} Y_{s-} ds = ψ^{-1/2} Y_{s-} (ds)^{1/2} = Y_{s-} (ds)^{1/2},$$

where the first equality in the second term above follows from $Y_{(k-1)h} = h^{1/2} Y_{(k-1)h}$ on the old time scale. $N_0^0 (v)$ varies on the time horizon $\left[ N_0^0 - |v|, N_0^0 + |v| \right]$ as implied by $D^* \left( C \right)$, as defined in Section 4. Again, in order to avoid clutter, we suppress the subscript $ψ$. We then have equation (S.61)-(S.62). Consider $T_b \leq T_b^0$ (i.e., $v \leq 0$). By Lemma S.D.12, there exists a $T$ such that for all $T > T$, $h^{-1/2} (Q_T \left( T_b \right) - Q_T \left( T_b^0 \right))$ is

$$Q_T \left( \theta^* \right) = -h^{-1/2} δ_h Z_Δ Z_Δ δ_h + h^{-1/2} 2δ_h Z_Δ e + o_p \left( 1 \right)$$

$$= -\left( δ_0 \right)' \left( \sum_{k=T_b}^{T_b^0} z_{kh} \bar{z}_{kh} \right) δ_0$$

$$+ 2 \left( δ_0 \right)' \left( h^{-1/2} \sum_{k=T_b}^{T_b^0} z_{kh} \bar{e}_{kh} \right) + o_p \left( 1 \right),$$

S-41
and note that this relationship corresponds to (S.63). 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 \( \mathcal{D}(\mathcal{D}^*, \mathbb{R}) \),

\[
J_{Z,h}(v) \triangleq \sum_{k=T_b(v)+1}^{T_b} z_{kh} \zeta_{kh}, \quad J_{e,h}(v) \triangleq \sum_{k=T_b(v)+1}^{T_b} z_{kh} \bar{e}_{kh},
\]

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

\[
\overline{Q}_T(\theta, v) = \left( \delta_p^0 \right)' R_{1,h}(v) \delta_p^0 + \left( \delta_Z^0 \right)' R_{2,h}(v) \delta_Z^0 + 2 \left( \delta^0 \right)' \left( \frac{1}{\sqrt{h}} \sum_{k=T_b+1}^{T_b^0} z_{kh} \bar{e}_{kh} \right),
\]

where

\[
R_{1,h}(v) \triangleq \sum_{k=T_b(v)+1}^{T_b^0} \left[ \frac{h}{Y_{(k-1)h}} \frac{Y_{(k-1)h}^{3/2}}{(Y_{(k-1)h})^2} \right], \quad R_{2,h}(v) \triangleq \left[ Z_{\Delta}^0 Z_{\Delta} \right]_{\{p \times p\}},
\]

and \( \delta^0 \) has been partitioned accordingly; that is, \( \delta_p^0 = (\mu_p^0, \alpha_p^0)' \) is the vector of parameters associated with the predictable regressors whereas \( \delta_Z^0 \) is the vector of parameters associated with the stochastic martingale regressors in \( Z \). By ordinary results for convergence of Riemann sums,

\[
\left( \delta_p^0 \right)' R_{1,h}(v) \delta_p^0 \overset{u.c.p.}{\Rightarrow} \left( \delta_p^0 \right)' \begin{bmatrix} N_b^0 - N_b & \int_{N_b^0 + v}^{N_b^0 + v} \tilde{Y}_s ds \\ \int_{N_b^0 + v}^{N_b^0 + v} Y_s ds & \int_{N_b^0 + v}^{N_b^0 + v} \tilde{Y}_s^2 ds \end{bmatrix} \delta_p^0.
\]

(S.87)

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

\[
R_{2,h}(v) \overset{u.c.p.}{\Rightarrow} \langle Z_\Delta, Z_\Delta \rangle (v).
\]

(S.88)

We now turn to examine the asymptotic behavior of the second term in (S.86) on \( \mathcal{D}^* \). We use the following steps. First, we present a stable central limit theorem for each component of \( Z_{\Delta}^0 e \). 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 \( Z_{\Delta}^0 e \),

\[
\frac{1}{\sqrt{h}} \sum_{k=T_b(v)+1}^{T_b^0} \alpha_p^0 \zeta_{(2)kh} = \frac{1}{\sqrt{h}} \sum_{k=T_b(v)+1}^{T_b^0} \alpha_p^0 \left( Y_{(k-1)h} \right) \bar{e}_{kh},
\]

and using \( \tilde{Y}_{(k-1)h} = h^{1/2} Y_{(k-1)h} \) [recall that \( \tilde{Y}_{(k-1)h} \) is bounded by the localization Assumption S.D.2] we
then have
\[
\frac{1}{\sqrt{h}} \sum_{k=T_b(v)+1}^{T_b^0} \alpha_0^0 \left(Y_{(k-1)h} \bar{e}_{kh}\right) \equiv \frac{1}{\sqrt{h}} \sum_{k=T_b(v)+1}^{T_b^0} \alpha_0^0 \left(\bar{Y}_{(k-1)h}\right) \bar{e}_{kh}
\]
\[
\underset{u.c.p.}{\Rightarrow} \int_{N_b^0}^{N_b^0 + v} \alpha_0^0 \bar{Y}_s dW_{e,s},
\]

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:

\[
\frac{1}{\sqrt{h}} \sum_{k=T_b(v)+1}^{T_b^0} \mu_0^0 \bar{z}_{kh} \bar{e}_{kh} \underset{u.c.p.}{\Rightarrow} \int_{N_b^0}^{N_b^0 + v} \mu_0^0 dW_{e,s}.
\]

(S.89)

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

\[
2 \left( \delta_0^0 \right)' \left( \frac{1}{\sqrt{h}} \sum_{k=T_b(v)+1}^{T_b^0} \bar{z}_{kh} \bar{e}_{kh} \right),
\]

where we have partitioned \(z_{kh}\) as \(z_{kh} = \left[ h^{1/2} Y_{(k-1)h} \bar{z}_{kh} \right]'\). Then, note that the small-dispersion asymptotic re-parametrization implies that \(\bar{z}_{kh} \bar{e}_{kh}\) corresponds to \(z_{kh} \bar{e}_{kh}\) from Theorem 4.1. Hence, we shall apply the same arguments as in the proof of Theorem 4.1 since (S.90) is simply \(2 \left( \delta_0^0 \right)'\) times \(W_h(v) = h^{-1/2} J_{e,h}(v)\), where \(J_{e,h}(v) \triangleq \sum_{k=T_b(v)+1}^{T_b^0} \bar{z}_{kh} \bar{e}\) with \((T_b(v) + 1)h \leq t_v < (T_b(v) + 2)h\). By Theorem 5.4.2 in Jacod and Protter (2012), \(W_h(v) \overset{L^\infty}{\Rightarrow} W_{Z_e}(v)\). Since the convergence of the drift processes \(R_{1,h}(v)\) and \(R_{2,h}(v)\) occur in probability locally uniformly in time while \(W_h(v)\) 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 \(Q_T(\theta, \cdot)\), for each \(\theta\). It remains to verify stochastic equicontinuity. As for the terms in \(R_{1,h}(v)\), we can decompose \((\alpha_\delta)^2 \left( \sum_{k=T_b(v)+1}^{T_b^0} \bar{z}_{k+1}^2 \right) - \left( \int_{N_b^0 + v}^{N_b^0} \bar{Y}_s ds \right)\) as \(Q_{0,T}(\theta, v) + Q_{7,T}(\theta, v)\), where \(Q_{0,T}(\theta, v) \triangleq (\alpha_\delta)^2 \left( \sum_k \zeta_{2,h,k}^* \right)\) and \(Q_{7,T}(\theta, v) \triangleq (\alpha_\delta)^2 \left( \sum_k \zeta_{2,h,k}^* \right)\), with

\[
\zeta_{2,h,k}^* \triangleq \left( z_{kh}^2 \right) - \left( \int_{(k-1)h}^{kh} \bar{Y}_s ds \right) - 2 \bar{Y}_{(k-1)h} \int_{(k-1)h}^{kh} \bar{Y}_s ds
\]
\[
+ 2 \mathbb{E} \left[ Y_{(k-1)h} \left( \bar{Y}_{(k-1)h} \cdot h - \int_{(k-1)h}^{kh} \bar{Y}_s ds \right) \right] \bigg| \mathcal{F}_{(k-1)h} \bigg)
\]
\[
\triangleq L_{1,h,k} + L_{2,h,k},
\]
and

\[
\zeta_{2,h,k}^* = 2 \bar{Y}_{(k-1)h} \bar{Y}_{(k-1)h} \cdot h - \int_{(k-1)h}^{kh} \bar{Y}_s ds
\]
\[
- \mathbb{E} \left[ \left( \bar{Y}_{(k-1)h} - \int_{(k-1)h}^{kh} \bar{Y}_s ds \right) \right] \bigg| \mathcal{F}_{(k-1)h} \bigg).
\]
Then, we have the following decomposition for $\mathcal{Q}_T^r (\theta^*) \triangleq \mathcal{Q}_T (\theta^*) + (\delta_0) \Lambda (v) \delta^0$ (if $v \leq 0$ and defined analogously for $v > 0$): $\mathcal{Q}_T^r (\theta^*) = \sum_{r=1}^{\infty} \mathcal{Q}_{r,T} (\theta, v)$, where $\mathcal{Q}_{r,T} (\theta, v)$, $r = 1, \ldots, 4$ are defined in (S.64) and $\mathcal{Q}_{5,T} (\theta, v) \triangleq (\mu_1^2)^2 (\sum_k \zeta_{1,k})$, $\mathcal{Q}_{6,T} (\theta, v) \triangleq \mu_2^2 \left( h^{-1/2} \sum_k \xi_{1,k} \right)$, $\mathcal{Q}_{9,T} (\theta, v) \triangleq (\alpha_1^2)^2 \left( h^{-1/2} \sum_k \xi_{2,k} \right)$ where $\zeta_{1,k} \triangleq \left( \sum_{k} \zeta_{1,k}^2 \right)^{1/2} - h$, $\xi_{1,k} \triangleq h^{1/2} \delta_{kh}$ and $\xi_{2,k} \triangleq \left( \sum_{k} \xi_{2,k}^2 \right)^{1/2} h^{1/2} \delta_{kh}$. Moreover, recall that $\sum_k$ replaces $\sum_{k=n+1}^{s}$ for $N \in D^*$ (C). Let us consider $\mathcal{Q}_{6,T} (\theta, v)$ first. For $s \in [(k-1)/h, kh]$, by the Burkholder-Davis-Gundy inequality

$$
\mathbb{E} \left[ \left| \mathcal{Y}_{k-1,h} \right| \mathcal{F}_{(k-1)h} \right] \leq K_h,
$$

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

$$
\mathbb{E} \left( \sup_{(\theta, v)} \left( \alpha_1^2 \sum_k L_{2,h} \right)^r \right)^{1/r} \leq K_r \left( \sup_{(\theta, v)} \left( \alpha_1^2 \sum_k h^r \right)^{1/r} \right) = K_r h^{r-1}. \tag{S.91}
$$

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

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

where the second inequality follows from the Burkholder-Davis-Gundy inequality. Thus, using a maximal inequality as in (S.91), we have for $r > 1$

$$
\mathbb{E} \left( \sup_{(\theta, v)} \left( \alpha_1^2 \sum_k L_{1,h} \right) \right)^{1/r} = K_r h^{r-1}. \tag{S.92}
$$

(S.91) and (S.92) imply that $\mathcal{Q}_{6,T} (\theta, v)$ is stochastically equicontinuous. Next, note that $\mathcal{Q}_{7,T} (\theta, v)$ is a sum of martingale differences times $h^{1/2}$ (recall the definition of $\Delta h \delta_{kh}$ and $V = \Delta h \delta_{kh}$). Therefore by Assumption S.D.2, for any $0 \leq s < t \leq N$, $V_t - V_s = O_p (1)$ uniformly and therefore,

$$
\sup_{(\theta, v)} \left| \mathcal{Q}_{7,T} (\theta, v) \right| \leq KO_p \left( h^{1/2} \right) \tag{S.93}
$$

Given (S.87) and (S.91)-(S.93), we deduce that

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

As for the term involving $R_{1,h} (v)$, it is easy to see that $\sup_{(\theta, v)} \left| \mathcal{Q}_{5,T} (\theta, v) \right| \rightarrow 0$. 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 $\{ \delta \cdot \mathcal{W}_h \}$ follows from the same property as those of $\{ \mathcal{Q}_{3,T} (\theta, v) \}$ and $\{ \mathcal{Q}_{4,T} (\theta, v) \}$. The stochastic equicontinuity of

$$
(\delta \cdot \mathcal{W}_h (\theta, v) - (Z, Z) (v) ) \delta Z,
$$

also follows from the same proof. Recall $\mathcal{Q}_{1,T} (\theta, v) + \mathcal{Q}_{2,T} (\theta, v)$ as defined in (S.64). Thus, stochastic equicontinuity follows from (S.66) and the equation right before that. Next, let us consider $\mathcal{Q}_{9,T} (\theta, v)$. We use the alternative definition (ii) of stochastic equicontinuity in Andrews (1994). Consider any sequence $\{ (\theta, v) \}$ and $\{ (\tilde{\theta}, \tilde{v}) \}$ (we omit the dependence on $h$ for simplicity). Assume $N_b \leq N^0_b \leq \tilde{N}_b$ (the other

S-44
Thus, the sequence main assertion from the continuous mapping theorem for the argmax functional.

\[Q_{0,T}(\theta, v) - Q_{0,T}(\bar{\theta}, \bar{v}) = \left| \alpha_{\delta} \sum_{k=T_{b}(v)+1}^{T_{b}^{0}} \bar{Y}_{(k-1)h}\bar{e}_{kh} - \bar{\alpha}_{\delta} \sum_{k=T_{b}^{0}}^{T_{b}(\bar{v})} \bar{Y}_{(k-1)h}\bar{e}_{kh} \right| \]

\[\leq \left| \alpha_{\delta} \right| \left( \sum_{k=T_{b}(v)+1}^{T_{b}^{0}} \bar{Y}_{(k-1)h}\bar{e}_{kh} \right) \leq \left| \alpha_{\delta} \right| \left( \sum_{k=T_{b}^{0}}^{T_{b}(\bar{v})} \bar{Y}_{(k-1)h}\bar{e}_{kh} \right) + \left| \bar{\alpha}_{\delta} \right| \left( \sum_{k=T_{b}^{0}}^{T_{b}(\bar{v})} \bar{Y}_{(k-1)h}\bar{e}_{kh} \right) \]  \hspace{1cm} (S.94)

For the second term, by the Burkholder-Davis-Gundy inequality for any \(r \geq 1\),

\[E \left[ \sup_{0 \leq u \leq d_{h}} \left| \sum_{k=T_{b}^{0}}^{T_{b}(\bar{v})} \bar{Y}_{(k-1)h}\bar{e}_{kh} \right|^{r} \right] \leq K_{r} (Nd_{h})^{r/2} \frac{1}{N_{dh}} \left( \sum_{k=T_{b}^{0}}^{T_{b}(\bar{v})} \int_{(k-1)h}^{kh} \left( \bar{Y}_{s} \right)^{2} ds \right)^{r/2} \leq K_{r} d_{h}^{r/2}. \]

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 \(B > 0\)

\[\left( \frac{1}{N_{dh}} \left( \sum_{k=T_{b}^{0}}^{T_{b}(\bar{v})} \int_{(k-1)h}^{kh} \left( \bar{Y}_{s} \right)^{2} ds \right)^{r/2} \right)^{1/r} \leq K_{r} d_{h}^{1/2} P (Nd_{h} > B) \rightarrow 0. \]

The argument for the first term in (S.94) is analogous. By Markov’s inequality and combining the above steps we have that for any \(\varepsilon > 0\) and \(\eta > 0\) there exists some \(T\) such that for \(T > T\),

\[P \left( \left| Q_{0,T}(\theta, v) - Q_{0,T}(\bar{\theta}, \bar{v}) \right| > \eta \right) < \varepsilon. \]

Thus, the sequence \(Q_{0,T} (\cdot, \cdot)\) is stochastically equicontinuous. Noting that the same proof can be repeated for \(Q_{S,T} (\cdot, \cdot)\), we conclude that the sequence of processes \(Q_{T} (\theta^{*}) , T \geq 1\) in (S.86) is stochastically equicontinuous. Furthermore, by (S.87) and (S.88) we obtain,

\[\left( \delta_{p}^{0} \right)^{T} R_{1,h} (\theta, v) \delta_{p}^{0} + \left( \delta_{Z}^{0} \right)^{T} (R_{2,h} ((\theta, v))) \delta_{Z}^{0} \overset{\text{u.c.p.}}{\Rightarrow} \left( \delta_{p}^{0} \right)^{T} \Lambda (v) \delta_{0}. \]

This suffices to guarantee the \(\mathcal{G}\)-stable convergence in law of the process \(Q_{T} (\cdot, \cdot) , T \geq 1\) towards a process \(\mathcal{W} (\cdot)\) with drift \(\Lambda (\cdot)\) which, conditional on \(\mathcal{G}\), is a two-sided Gaussian martingale process with covariance matrix given in (S.7). By definition, \(D^{*} (C)\) is compact and \(Th (\hat{\lambda}_{b} - \lambda_{0}) = O_{p} (1)\), which together with the fact that the limit process is a continuous Gaussian process enable one to deduce the main assertion from the continuous mapping theorem for the argmax functional. \(\square\)
S.D.5.3 Proof of Proposition S.C.2

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

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

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

$$\tilde{Y}_t - \tilde{Y}^0_t = \int_0^t \alpha^0_1 \left(\tilde{Y}_s - \tilde{Y}^0_s\right) ds + B\gamma_h \left(\nu^0\right)' D_t + B\gamma_h \left(\delta^0_{Z,1}\right)' \int_0^t dZ_s + B\gamma_h \int_0^t \sigma, dW_{e,s}.$$

We can use Cauchy-Schwarz's inequality,

$$|\tilde{Y}_t - \tilde{Y}^0_t|^2 \leq 2K \left[ \int_0^t |\alpha^0_1 \left(\tilde{Y}_s - \tilde{Y}^0_s\right)| ds \right]^2 + \left| \nu^0 D_t \right|^2 + \left| \delta^0_{Z,1} \int_0^t dZ_s \right|^2 + \left| \int_0^t \nu^0 D_t \right|^2 \right] (B\gamma_h)^2 \leq 2K \left[ \alpha^0_1 \left(\tilde{Y}_s - \tilde{Y}^0_s\right) \right]^2 ds + \left( \sup_{0 \leq s \leq t} |\nu^0 D_s| \right) \leq 2K \left[ \alpha^0_1 \int_0^t \tilde{Y}_s - \tilde{Y}^0_s \right]^2 ds + \left( \sup_{0 \leq s \leq t} |\nu^0 D_s| \right) \leq 2K \left( \alpha_1 \int_0^t 2K ds \right) \leq 2K \left( \alpha_1 \int_0^t 2K ds \right)$$

By Gronwall's inequality,

$$|\tilde{Y}_t - \tilde{Y}^0_t|^2 \leq 2 (B\gamma_h)^2 C \exp \left( \int_0^t 2K^2 ds \right) \leq 2 (B\gamma_h)^2 C \exp \left( 2K^2 t^2 \right),$$

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

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

Proof. Note that

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

$$= \sup_{\mu_1, \alpha_1} \left| \int_0^{N_b^0} \left(\mu_1 + \alpha_1 \tilde{Y}^*_s\right) ds - \int_0^{N_b^0} \left(\mu_1 + \alpha_1 \tilde{Y}^0_s\right) ds \right| \leq \sup_{\alpha_1} \int_0^{N_b^0} \left|\alpha_1\right| \left|\tilde{Y}^*_s - \tilde{Y}^0_s\right| ds \leq KO_p(\gamma_h) \sup_{\alpha_1} \left|\alpha_1\right|,$$
Lemma S.D.15. For each $3 \leq j \leq p + 2$ and each $\theta$, as $h \downarrow 0$,

$$
\sum_{k=1}^{[N_0^h/h]} \left( \mu_1 + \alpha_1 \tilde{Y}_{(k-1)h} \right) \delta_{Z,1}^{(j)} Z_k^{(j)} \overset{P}{\to} \int_0^{N_0^h} \left( \mu_1 + \alpha_1 \tilde{Y}_{(k-1)h} \right) dZ_s^{(j)}.
$$

Proof. Note that

$$
\sum_{k=1}^{[N_0^h/h]} \left( \mu_1 + \alpha_1 \tilde{Y}_{(k-1)h} \right) \delta_{Z,1}^{(j)} Z_k^{(j)} = \int_0^{N_0^h} \left( \mu_1 + \alpha_1 \tilde{Y}_s^* \right) dZ_s^{(j)}.
$$

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

$$
P \left( \int_0^{N_0^h} \alpha_1 \left( \tilde{Y}_s^* - \tilde{Y}_s^0 \right) \delta_{Z,1}^{(j)} dZ_s^{(j)} \right) > \eta
$$

$$
\leq \left( \sup_{0 \leq s \leq N} \sum_{r=1}^{p} \left( \sigma_{Z,s}^{(j,r)} \right)^2 \right)^{1/2} |\alpha_1| \left| \delta_{Z,1}^{(j)} \right| \left( \int_0^{N_0^h} \mathbb{E} \left[ \left( \tilde{Y}_s^* - \tilde{Y}_s^0 \right)^2 \right] ds \right)^{1/2},
$$

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

Lemma S.D.16. As $h \downarrow 0$, uniformly in $\mu_1$, $\alpha_1$,

$$
\sum_{k=1}^{T_0^h} \left( \mu_1 + \alpha_1 \tilde{Y}_{(k-1)h} \right) \left( \tilde{Y}_{kh} - \tilde{Y}_{(k-1)h} \right) - \left( \mu_1^0 + \alpha_1^0 \tilde{Y}_{(k-1)h} \right) h \overset{P}{\to} 0.
$$

Proof. By definition [recall the notation in (S.4)],

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

Then,

$$
\sum_{k=1}^{T_0^h} \left( \mu_1 + \alpha_1 \tilde{Y}_{(k-1)h} \right) \left( \tilde{Y}_{kh} - \tilde{Y}_{(k-1)h} \right) - \left( \mu_1^0 + \alpha_1^0 \tilde{Y}_{(k-1)h} \right) h
$$

$$
= \sum_{k=1}^{T_0^h} \int_{(k-1)h}^{kh} \left( \mu_1 + \alpha_1 \tilde{Y}_{(k-1)h} \right) \left( \mu_1^0 + \alpha_1^0 \tilde{Y}_s - \left( \mu_1^0 + \alpha_1^0 \tilde{Y}_{(k-1)h} \right) \right) dt
$$

$$
+ \sum_{k=1}^{T_0^h} \int_{(k-1)h}^{kh} \left( \mu_1 + \alpha_1 \tilde{Y}_{(k-1)h} \right) \Delta_h \tilde{V}_k \left( \nu^0, \delta_{Z,1}^0, \delta_{Z,2}^0 \right)
$$

$$
= \int_0^{N_0^h} \left( \mu_1 + \alpha_1 \tilde{Y}_{(k-1)h} \right) \left( \alpha_1^0 \left( \tilde{Y}_s - \tilde{Y}_{(k-1)h} \right) \right) ds
$$

$$
+ B \gamma_h \int_0^{N_0^h} \left( \mu_1 + \alpha_1 \tilde{Y}_s^* \right) dV_s.
$$
For the first term on the right-hand side,

\[
\sup_{\mu_1, \alpha_1} \left| \int_0^{N_0^h} \left( \mu_1 + \alpha_1 \bar{Y}_s^* \right) \left( \alpha_1 \left( \bar{Y}_s^* - \bar{Y}_s^\ast \right) \right) ds \right|
\leq \left| \alpha_1 \right| \sup_{\mu_1, \alpha_1} \left| \int_0^{N_0^h} \left( \mu_1 + \alpha_1 \bar{Y}_s^* \right) \left( \bar{Y}_s^* - \bar{Y}_s^\ast \right) ds \right|
\leq \left| \alpha_1 \right| K \left( \int_0^{N_0^h} \sup_{0 \leq s \leq N_0^h} \left| \bar{Y}_s^* - \bar{Y}_s^\ast \right| + \sup_{0 \leq s \leq N_0^h} \left| \bar{Y}_s^0 - \bar{Y}_s^* \right| ds \right),
\]

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

\[
B_{\gamma h} \sup_{\mu_1, \alpha_1} \left| \int_0^{N_0^h} \left( \mu_1 + \alpha_1 \bar{Y}_s^* \right) dZ_s^{(j)} \right|
\leq B_{\gamma h} \sup_{\mu_1, \alpha_1} \left| \int_0^{N_0^h} \left( \mu_1 + \alpha_1 \bar{Y}_s^* \right) \sum_{r=1}^p \sigma_{Z,s}^{(j,r)} dW_s^{(r)} \right|.
\]

Let \( R_{j,h} = R_{j,h} \left( \mu_1, \alpha_1 \right) \triangleq \int_0^{N_0^h} B_{\gamma h} \left( \mu_1 + \alpha_1 \bar{Y}_s^* \right) \sum_{r=1}^p \sigma_{Z,s}^{(j,r)} dW_s^{(r)} \) (we index \( R_j \) by \( h \) because \( \bar{Y}_s^* \) depends on \( h \)). Then, we want to show that, for every \( \varepsilon > 0 \) and \( K > 0 \),

\[
P \left( \sup_{\mu_1, \alpha_1} \left| R_{j,h} \left( \mu_1, \alpha_1 \right) \right| > K \right) \leq \varepsilon.
\] (S.95)

In view of Chebyshev’s inequality and the Itō’s isometry,

\[
P \left( \left| R_{j,h} \right| > K \right) \leq \left( \frac{B_{\gamma h}}{K} \right)^2 \mathbb{E} \left[ \left( \int_0^{N_0^h} \left( R_{j,h} / (B_{\gamma h}) \right)^2 \right)^2 \right],
\]

\[
\leq \left[ \sup_{0 \leq s \leq N_r^h} \sum_{r=1}^p \left( \sigma_{Z,s}^{(j,r)} \right)^2 \right] \left( \frac{B_{\gamma h}}{K} \right)^2 \mathbb{E} \left[ \left( \mu_1 + \alpha_1 \bar{Y}_s^* \right)^2 ds \right],
\]

so that by the boundness of the processes (cf. Assumption S.D.2) and the compactness of \( \Theta_0 \), we have for some \( A < \infty \),

\[
P \left( \left| R_{j,h} \right| > K \right) \leq A \left[ \sup_{0 \leq s \leq T} \sum_{r=1}^p \left( \sigma_{Z,s}^{(j,r)} \right)^2 \right] \left( \frac{B_{\gamma h}}{K} \right)^2 \rightarrow 0,
\] (S.96)

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 Itō 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.96) with the Burkholder-Davis-Gundy inequality in place of the Itō’s isometry. Let \( g \left( \bar{Y}_s^*, \bar{\theta} \right) \triangleq \mu_{1,1} + \alpha_{1,1} \bar{Y}_s^* \), \( \bar{\theta}_1 \triangleq (\mu_{1,1}, \alpha_{1,1})^t \) and \( \bar{\theta}_2 \triangleq (\mu_{2,1}, \alpha_{2,1})^t \). For any \( \bar{\theta}_1, \bar{\theta}_2 \), first use the Burkholder-Davis-Gundy inequality to yield,

\[
\mathbb{E} \left[ \left| R_{j,h}(\bar{\theta}_2) - R_{j,h}(\bar{\theta}_1) \right|^{2m} \right]
\]
\[ \leq (B\gamma h)^{2m} K_m \left[ \sup_{0 \leq s \leq N} \sum_{r=1}^{p} \left( \sigma_{Z,s}^{(j,r)} \right)^2 \right]^{m} \]

\[ \times E \left[ \left( \int_{0}^{N_0} \left( g \left( \bar{Y}_s^* , \bar{\theta}_2 \right) - g \left( \bar{Y}_s^* , \bar{\theta}_1 \right) \right)^2 \, ds \right)^{m} \right] \]

\[ \leq (B\gamma h)^{2m} K_m \left[ \sup_{0 \leq s \leq N} \sum_{r=1}^{p} \left( \sigma_{Z,s}^{(j,r)} \right)^2 \right]^{m} \]

\[ \times E \left[ \left( \int_{0}^{N_0} \left( (\mu_{1,2} - \mu_{1,1}) + (\alpha_{1,2} - \alpha_{1,1}) \bar{Y}_s^* \right)^2 \, ds \right)^{m} \right] \]

\[ \leq (B\gamma h)^{2m} K_m \left[ \sup_{0 \leq s \leq N} \sum_{r=1}^{p} \left( \sigma_{Z,s}^{(j,r)} \right)^2 \right]^{m} \]

\[ \times E \left[ \left( \int_{0}^{N_0} \left( (\mu_{1,2} - \mu_{1,1}) + (\alpha_{1,2} - \alpha_{1,1}) \bar{Y}_s^* \right)^2 \, ds \right)^{m} \right] \]

\[ \leq (B\gamma h)^{2m} K_m \mathbb{E} \left[ \left( \int_{0}^{N_0} \left( 2(\mu_{1,2} - \mu_{1,1})^2 + 2C(\alpha_{1,2} - \alpha_{1,1})^2 \right) \, ds \right)^{m} \right] \]

\[ \leq 2^m (B\gamma h)^{2m} K_m \left\| 2 \left( \bar{\theta}_2 - \bar{\theta}_1 \right) \right\|^{2m} \left( \int_{0}^{N_0} \, ds \right)^{m} \]  \hspace{1cm} (S.97)

\[ + 2^m (B\gamma h)^{2m} K \left( \bar{\theta}_1, \bar{\theta}_2, m, C \right) \]

where \( C = \sup_{s \geq 0} |\bar{Y}_s^*| \), \( K \left( \bar{\theta}_1, \bar{\theta}_2, m, C \right) \) is some constant that depends on its arguments and we have used that \((a + b)^2 \leq 2a^2 + 2b^2\). Thus, since \( \gamma_n \downarrow 0 \), the mapping \( R_{j,h}(\cdot) \) satisfies a Lipschitz-type condition [cf. Section 2 in Andrews (1992)]. This is sufficient for the asymptotic stochastic equicontinuity of \( \{R_{j,h}(\cdot)\} \). Therefore, using Theorem 20 in Appendix I of Ibragimov and Has’minskii (1981), (S.96) and (S.97) yield (S.95). Since the same result can be shown to remain valid for each term in the stochastic element \( \Delta h V_k(\nu, \delta_{Z,1}, \delta_{Z,2}) \), this establishes the claim. \( \square \)

**Proof of Proposition S.C.2.** To avoid clutter, we prove the case for which the true parameters are \((\mu_0^1, \alpha_0^1)'\). The extension to parameters being local-to-zero is straightforward. The least-squares estimates of \((\mu_1^0, \alpha_1^0)'\) are given by,

\[ \hat{\mu}_1 \hat{N}_b = \bar{Y}_{\hat{N}_b} - \hat{\theta}_0 - \hat{x}_1 h \sum_{k=1}^{\hat{N}_b} \bar{Y}_{(k-1)h} \]  \hspace{1cm} (S.98)

\[ \hat{\alpha}_1 = \frac{\sum_{k=1}^{\hat{N}_b} (\bar{Y}_{kh} - \bar{Y}_{(k-1)h}) \bar{Y}_{(k-1)h}}{h \sum_{k=1}^{\hat{N}_b} \bar{Y}_{(k-1)h}^2 - \hat{N}_b^{-1} \left( h \sum_{k=1}^{\hat{N}_b} \bar{Y}_{(k-1)h} \right)^2} \]  \hspace{1cm} (S.99)
Then, assuming $T_b < T_b^0$,

\[
\hat{\alpha}_1 = \frac{\sum_{k=1}^{T_b} \left( \mu_1^0h + \alpha_1^0 \bar{Y}_{(k-1)h}h + \Delta_h \bar{V}_{h,k} \right) \bar{Y}_{(k-1)h}}{h \sum_{k=1}^{T_b} \bar{Y}_{(k-1)h}^2 - \hat{N}_b^{-1} \left( h \sum_{k=1}^{T_b} \bar{Y}_{(k-1)h} \right)^2}
- \frac{\left( \mu_1^0 + \alpha_1^0 \hat{N}_b^{-1} \sum_{k=1}^{T_b} \bar{Y}_{(k-1)h}h + \hat{N}_b^{-1} B \gamma_h \left( V_{N_b} - V_0 \right) \right) h \sum_{k=1}^{T_b} \bar{Y}_{(k-1)h}}{h \sum_{k=1}^{T_b} \bar{Y}_{(k-1)h}^2 - \hat{N}_b^{-1} \left( h \sum_{k=1}^{T_b} \bar{Y}_{(k-1)h} \right)^2}
+ o_p(1),
\]

and thus

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

By part (ii) of Theorem S.C.1, $N_b^0 - \hat{N}_b = O_p(h^{1-\kappa})$, 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 $\Delta_h \bar{V}_{h,k} = h^{1/2} \Delta \bar{V}_{h,k}$ from (S.4), we have by ordinary convergence of approximations to Riemann sums, Lemma S.D.14 and the continuity of probability limits,

\[
\alpha_1^0 \sum_{k=1}^{T_b} \bar{Y}_{(k-1)h} \overset{P}{\to} \alpha_1^0 \int_0^{N_b^0} \bar{Y}_s ds,
\sum_{k=1}^{T_b^0} \mu_1^0h \overset{P}{\to} \mu_1^0 \int_0^{N_b^0} ds,
\]

and by Lemma S.D.15, $\sum_{k=1}^{T_b^0} \bar{Y}_{(k-1)h} \Delta_h \bar{V}_k \overset{P}{\to} 0$. Thus, we deduce that

\[
\hat{\alpha}_1 = \alpha_1^0 + O_p(B \gamma_h).
\]
Using (S.100) into (S.98),

\[ \hat{\mu}_1 \tilde{N}_b = \bar{Y}_{\tilde{N}_b} - \bar{Y}_0 - \alpha_1^0 h \sum_{k=1}^{\tilde{T}_b} \bar{Y}_{(k-1)h} - O_p (B\gamma_h), \]

\[ = \bar{Y}_{\tilde{N}_b} - \bar{Y}_0 - \alpha_1^0 h \sum_{k=1}^{T_b^0} \bar{Y}_{(k-1)h} - \alpha_1^0 h \sum_{k=\tilde{T}_b+1}^{T_b^0} \bar{Y}_{(k-1)h} - o_p (1). \]

By part (ii) of Theorem S.C.1, the number of terms in the second sum above increases at rate \( T^\kappa \) and thus, \( \alpha_1^0 h \sum_{k=\tilde{T}_b+1}^{T_b^0} \bar{Y}_{(k-1)h} = KO_p (h^{1-\kappa}) \), where we have also used standard estimates for the drift arising from the Burkholder-Davis-Gundy inequality. This gives

\[ \hat{\mu}_1 \tilde{N}_b = \bar{Y}_{N_b^0} - \bar{Y}_0 - \alpha_1^0 \int_0^{N_b^0} \bar{Y}_s ds - \alpha_1^0 O_p (h^{1-\kappa}) - o_p (1). \]

Noting that

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

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

\[ \hat{\mu}_1 = \mu_1^0 + o_p (B\gamma_h). \quad \text{(S.101)} \]

Thus, as \( h \downarrow 0 \), \( \hat{\mu}_1 \) is consistent for \( \mu_1^0 \). The case where \( \hat{T}_b > T_b^0 \) can be treated in the same fashion and is omitted. Further, the consistency proof for \( (\hat{\mu}_2, \hat{\alpha}_2)' \) 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,

\[ \hat{u}_{kh} = \begin{cases} 
 h^{-1/2} \left( \Delta_h \bar{Y}_k - \hat{\mu}_1 \hat{x}_{kh}^{(1)} - \hat{\alpha}_1 \hat{x}_{kh}^{(2)} \right), & k \leq \hat{T}_b \\
 h^{-1/2} \left( \Delta_h \bar{Y}_k - \hat{\mu}_2 \hat{x}_{kh}^{(1)} - \hat{\alpha}_2 \hat{x}_{kh}^{(2)} \right), & k > \hat{T}_b,
\end{cases} \]

where \( \hat{x}_{kh}^{(1)} = h \) and \( \hat{x}_{kh}^{(2)} = \bar{Y}_{(k-1)h} h \). This yields, for \( k \leq T_b^0 \leq \hat{T}_b \),

\[ \hat{u}_{kh} = h^{-1/2} \left( \mu_1^0 h + \alpha_1^0 \bar{Y}_{(k-1)h} h + B\gamma_h \Delta_h V_k - \hat{\mu}_1 h - \hat{\alpha}_1 \bar{Y}_{(k-1)h} h \right), \]

and using (S.100) and (S.101),

\[ \hat{u}_{kh} = h^{-1/2} \left( \mu_1^0 h + \alpha_1^0 \bar{Y}_{(k-1)h} h + B\gamma_h \Delta_h V_k - \mu_1^0 h \\
- O_p \left( h^{3/2} \right) - \alpha_1^0 \bar{Y}_{(k-1)h} h - o_p \left( h^{3/2} \right) \right) = h^{-1/2} B\gamma_h \Delta_h V_k - O_p (h). \quad \text{(S.102)} \]

Similarly, for \( T_b^0 \leq \hat{T}_b \leq k, \)

\[ \hat{u}_{kh} = h^{-1/2} B\gamma_h \Delta_h V_k - O_p (h), \quad \text{(S.103)} \]

whereas for \( \hat{T}_b < k \leq T_b^0 \),

\[ \hat{u}_{kh} = h^{-1/2} \left( \mu_1^0 h + \alpha_1^0 \bar{Y}_{(k-1)h} h + B\gamma_h \Delta_h V_k - \mu_1^0 h \right). \]
\[- O_p \left( h^{3/2} \right) - \alpha_0^0 \bar{Y}_{(k-1)h} h - O_p \left( h^{3/2} \right) \]
\[= h^{-1/2} \left( -\mu_0^0 h - \alpha_0^0 \bar{Y}_{(k-1)h} h + B \gamma_h \Delta_h V_k - O_p \left( h^{3/2} \right) \right) \]
\[= -\mu_0^0 h^{1/2} - \alpha_0^0 \bar{Y}_{(k-1)h} h^{1/2} + h^{-1/2} B \gamma_h \Delta_h V_k - O_p \left( h \right). \quad (S.104)\]

Next, note that \( \sum_{k=\tilde{T}_b+1}^{T_b} \mu_0^0 h^{1/2} \leq K h^{1/2-\kappa} \) and \( \sum_{k=\tilde{T}_b+1}^{T_b} \alpha_0^0 \bar{Y}_{(k-1)h} h^{1/2} \leq K h^{1/2-\kappa} \) since by Theorem S.C.1-(ii) there are \( T^\kappa \) terms in each sum. Moreover, recall that \( \epsilon_{kh} = \Delta_h \epsilon_k^* \sim N \left( 0, \sigma_{e,k-1}^2 h \right) \) and thus
\[\sum_{k=\tilde{T}_b+1}^{T_b} \epsilon_{kh} = \sqrt{h} \sum_{k=\tilde{T}_b+1}^{T_b} \epsilon_{kh} = h^{1/2-\kappa} o_p \left( 1 \right). \]
Therefore, \( \sum_{k=\tilde{T}_b+1}^{T_b} \hat{\epsilon}_{kh} = K o_p \left( h^{1/2-\kappa} \right). \)

Since \( \kappa \in (0, 1/2) \), this shows that the residuals \( \hat{\epsilon}_{kh} \) from equation (S.104) are asymptotically negligible. That is, asymptotically the estimator of \( \left( (\beta_0^0)' \right)' \) minimizes (assuming \( \tilde{T}_b \leq T_b \)),

\[\sum_{k=1}^{\tilde{T}_b} (\hat{u}_{kh} - \bar{z}_{kh}^J \beta_S)^2 + \sum_{k=T_b^0+1}^{T} (\hat{u}_{kh} - \bar{z}_{kh}^J \beta_S - \bar{z}_{0,kh}^J \delta_S)^2 + o_p \left( 1 \right),\]

where \( X = \left[ \bar{X}(1) \quad \bar{X}(2) \right] \), \( \beta_0^0 = \left[ \mu_1^0 \quad \alpha_1^0 \right] (\beta_0^2)' \), and \( Z_0 \) and \( \delta_0^S \) 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 \( \beta_0^S \) and \( \delta_0^S \):

\[\hat{U} = \bar{X} \hat{\beta}_S + \hat{Z}_0 \delta_S + \text{residuals},\]

where \( \hat{Z}_0 = \begin{bmatrix} \hat{z}(1) & \cdots & \hat{z}(p) \\ \vdots & \ddots & \vdots \\ \hat{z}(T_b^0+1)^h & \cdots & \hat{z}(T_b^0+1)^h \end{bmatrix} \), and \( \hat{U} = \left( \hat{u}_{kh}; k = 1, \ldots, \tilde{T}_b, T_b^0 + 1, \ldots, N \right) \). Therefore, using (S.102) and (S.103), we have

\[h^{-1/2} \begin{bmatrix} \hat{\beta}_S - \beta_0^0 \\ \hat{\delta}_S - \delta_0 \end{bmatrix} = \begin{bmatrix} \bar{X}' \bar{X} & \bar{X}' \hat{Z}_0 \\ \hat{Z}_0' \bar{X} & \hat{Z}_0' \hat{Z}_0 \end{bmatrix}^{-1} \]
\[\times h^{-1/2} \begin{bmatrix} \bar{X}' e \quad \hat{X}' \left( Z_0 - \hat{Z}_0 \right) \delta_0^0 + \bar{X}' AO_p \left( h \right) \\ \hat{Z}_0' e \quad \hat{Z}_0' \left( Z_0 - \hat{Z}_0 \right) \delta_0^0 + \hat{Z}_0' AO_p \left( h \right) \end{bmatrix},\]

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

\[\begin{bmatrix} \bar{X}' \bar{X} & \bar{X}' \hat{Z}_0 \\ \hat{Z}_0' \bar{X} & \hat{Z}_0' \hat{Z}_0 \end{bmatrix}^{-1} \bar{X}' AO_p \left( h^{1/2} \right) = o_p \left( 1 \right), \quad (S.105)\]

---

4The same bound holds for the corresponding sum involving the other terms in \( \Delta_h V_k \).
and
\[
\begin{bmatrix}
\tilde{X}' \tilde{X} & \tilde{X}' \tilde{Z}_0 \\
\tilde{Z}_0' \tilde{X} & \tilde{Z}_0' \tilde{Z}_0
\end{bmatrix}^{-1} \frac{1}{h^{1/2}} \tilde{X}' \left( Z_0 - \tilde{Z}_0 \right) \delta^0 = O_p(1) o_p(1) = o_p(1). \tag{S.106}
\]

The same arguments can be used for \( \tilde{Z}_0' \left( Z_0 - \tilde{Z}_0 \right) \delta^0 \) and \( \tilde{Z}_0' A O_p(h) \). Therefore, in view of (S.100) and (S.101), we obtain \( \hat{\mu}_1 = \mu_1^0 + o_p(1) \) and \( \hat{\alpha}_1 = \alpha_1^0 + o_p(1) \), respectively, whereas (S.105) and (S.106) imply \( \hat{\beta}_S = \beta_S^0 + o_p(1) \) and \( \hat{\delta}_S = \delta_S^0 + o_p(1) \), respectively. Under the setting where the magnitude of the shifts is local to zero, we observe that by Proposition 4.1, \( \tilde{N}_b - \tilde{N}_b^0 = O_p(h^{1-k}) \) and one can follow the same steps that led to (S.100) and (S.101) and proceed as above. The final result is \( \hat{\theta} = \theta^0 + o_p(1) \), which is what we wanted to show. \( \Box \)

**S.D.5.4 Negligibility of the Drift Term**

Recall Lemma S.D.10 and apply the same proof as in Section S.D.4.9. Of course, the negligibility only applies to the drift processes \( \mu_t \) from (2.3) (i.e., only the drift processes of the semimartingale regressors) and not to \( \mu_1^0, \mu_2^0, \alpha_1^0 \) or \( \alpha_2^0 \). The steps are omitted since they are the same.

**S.E Additional Simulations Results on HDR Confidence Sets**

We continue with the analysis of finite-sample from Section 7. We consider discrete-time DGPs of the form
\[
y_t = D'_t \nu^0 + Z'_t \delta^0 + Z'_t \delta^0 \mathbb{1}_{\{t>T\}} + e_t, \quad t = 1, \ldots, T, \tag{S.1}
\]
with \( T = 100 \) and, without loss of generality, \( \nu^0 = 0 \) (except for M5-M6, M8-M9). We consider eight versions of (S.1): M3 involves a break in the simultaneous mean and variance of an i.i.d. series with \( Z_t = 1 \) for all \( t \), \( D_t \) absent, and \( e_t = \left( 1 + \mathbb{1}_{\{t>T\}} \right) u_t \) with \( u_t \sim i.i.d. \mathcal{N}(0, 1) \); M4 is the same as M1 but with stationary Gaussian AR(1) disturbances \( e_t = 0.3 e_{t-1} + u_t, u_t \sim i.i.d. \mathcal{N}(0, 0.49) \); M5 is a partial structural change model with \( D_t = 1 \) for all \( t, \nu^0 = 1 \) and \( Z_t = 0.5 Z_t + u_t \) with \( u_t \sim i.i.d. \mathcal{N}(0, 0.75) \) independent of \( e_t \sim i.i.d. \mathcal{N}(0, 1) \); M6 is similar to M5 but with \( u_t \sim i.i.d. \mathcal{N}(0, 1) \) and heteroskedastic disturbances given by \( e_t = v_t \mid Z_t \) where \( v_t \) is a sequence of i.i.d. \( \mathcal{N}(0, 1) \) random variables independent of \( \{Z_t\} \); M7 is the same as M4 but with \( u_t \) drawn from a t\(_v\) distribution with \( v = 5 \) degrees of freedom; M8 is a model with a lagged dependent variable with \( D_t = y_{t-1}, Z_t = 1, e_t \sim i.i.d. \mathcal{N}(0, 0.49), \nu^0 = 0.3 \) and \( Z'_t \delta^0 \mathbb{1}_{\{t>T\}} \) is replaced by \( Z'_t \left( 1 - \nu^0 \right) \delta^0 \mathbb{1}_{\{t>T\}} \); M9 has FIGARCH(1,d,1) errors given by \( e_t = \sigma_t u_t, u_t \sim i.i.d. \mathcal{N}(0, 1) \) and \( \sigma_t = 0.1 + (1 - 0.2 L (1 - L)) e_t^2 \) where \( d = 0.6 \) is the order of differencing and \( L \) the lag operator, \( D_t = 1, \nu^0 = 1 \) and \( Z_t \sim i.i.d. \mathcal{N}(1, 1.44) \) independent of \( e_t \). M10 is similar to M6 but with an ARFIMA(0.3, d, 0) regressor \( Z_t \) with order of differencing \( d = 0.5 \), \( \text{Var} (Z_t) = 1 \) and \( e_t \sim \mathcal{N}(0, 1) \) independent of \( \{Z_t\} \). We set \( \beta^0 = 1 \) in all models, except in M8 where \( \beta^0 = 0 \). The Results are reported in Table 4-11.
Table 4: Small-sample coverage rate and length of the confidence set for model M3

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_0 = 0.5$</td>
<td>HDR</td>
<td>0.970</td>
<td>86.65</td>
<td>0.937</td>
<td>76.29</td>
<td>0.901</td>
<td>55.59</td>
<td>0.934</td>
</tr>
<tr>
<td>Bai (1997)</td>
<td>HDR</td>
<td>0.854</td>
<td>70.60</td>
<td>0.843</td>
<td>58.27</td>
<td>0.857</td>
<td>40.70</td>
<td>0.923</td>
</tr>
<tr>
<td>$\widehat{U}_T$.neq</td>
<td>HDR</td>
<td>0.961</td>
<td>88.95</td>
<td>0.961</td>
<td>80.33</td>
<td>0.961</td>
<td>61.15</td>
<td>0.964</td>
</tr>
<tr>
<td>ILR</td>
<td>HDR</td>
<td>0.989</td>
<td>92.53</td>
<td>0.985</td>
<td>84.06</td>
<td>0.977</td>
<td>58.05</td>
<td>0.958</td>
</tr>
<tr>
<td>$\lambda_0 = 0.35$</td>
<td>HDR</td>
<td>0.976</td>
<td>89.81</td>
<td>0.961</td>
<td>83.26</td>
<td>0.935</td>
<td>64.87</td>
<td>0.934</td>
</tr>
<tr>
<td>Bai (1997)</td>
<td>HDR</td>
<td>0.823</td>
<td>69.86</td>
<td>0.822</td>
<td>55.87</td>
<td>0.844</td>
<td>38.91</td>
<td>0.932</td>
</tr>
<tr>
<td>$\widehat{U}_T$.neq</td>
<td>HDR</td>
<td>0.963</td>
<td>89.84</td>
<td>0.963</td>
<td>82.26</td>
<td>0.961</td>
<td>65.87</td>
<td>0.964</td>
</tr>
<tr>
<td>ILR</td>
<td>HDR</td>
<td>0.990</td>
<td>93.48</td>
<td>0.985</td>
<td>88.69</td>
<td>0.982</td>
<td>68.23</td>
<td>0.977</td>
</tr>
<tr>
<td>$\lambda_0 = 0.2$</td>
<td>HDR</td>
<td>0.978</td>
<td>90.39</td>
<td>0.975</td>
<td>85.89</td>
<td>0.934</td>
<td>70.05</td>
<td>0.957</td>
</tr>
<tr>
<td>Bai (1997)</td>
<td>HDR</td>
<td>0.782</td>
<td>70.24</td>
<td>0.805</td>
<td>56.37</td>
<td>0.831</td>
<td>37.66</td>
<td>0.928</td>
</tr>
<tr>
<td>$\widehat{U}_T$.neq</td>
<td>HDR</td>
<td>0.968</td>
<td>91.11</td>
<td>0.968</td>
<td>87.62</td>
<td>0.972</td>
<td>78.17</td>
<td>0.967</td>
</tr>
<tr>
<td>ILR</td>
<td>HDR</td>
<td>0.980</td>
<td>93.32</td>
<td>0.981</td>
<td>91.60</td>
<td>0.978</td>
<td>81.60</td>
<td>0.981</td>
</tr>
</tbody>
</table>

The model is $y_t = \beta^0 + \delta^0 1_{\{t > \lfloor T \lambda_0 \rfloor \}} + e_t$, $e_t = (1 + 1_{\{t > \lfloor T \lambda_0 \rfloor \}}) u_t$, $u_t \sim i.i.d. \mathcal{N}(0, 1)$, $T = 100$. The notes of Table 2 apply.
Table 5: Small-sample coverage rate and length of the confidence set for model M4

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_0 = 0.5$</td>
<td>HDR</td>
<td>0.904</td>
<td>72.44</td>
<td>0.901</td>
<td>57.37</td>
<td>0.919</td>
<td>29.70</td>
<td>0.971</td>
</tr>
<tr>
<td>Bai (1997)</td>
<td>0.833</td>
<td>66.34</td>
<td>0.834</td>
<td>41.32</td>
<td>0.895</td>
<td>18.63</td>
<td>0.969</td>
<td>5.49</td>
</tr>
<tr>
<td>$\widehat{U}_{T, eq}$</td>
<td>0.958</td>
<td>87.16</td>
<td>0.968</td>
<td>71.47</td>
<td>0.958</td>
<td>45.82</td>
<td>0.957</td>
<td>28.01</td>
</tr>
<tr>
<td>ILR</td>
<td>0.932</td>
<td>79.38</td>
<td>0.944</td>
<td>53.48</td>
<td>0.966</td>
<td>21.98</td>
<td>0.993</td>
<td>4.87</td>
</tr>
<tr>
<td>$\lambda_0 = 0.35$</td>
<td>HDR</td>
<td>0.910</td>
<td>70.98</td>
<td>0.902</td>
<td>53.88</td>
<td>0.917</td>
<td>28.07</td>
<td>0.973</td>
</tr>
<tr>
<td>Bai (1997)</td>
<td>0.849</td>
<td>65.13</td>
<td>0.840</td>
<td>40.43</td>
<td>0.900</td>
<td>18.69</td>
<td>0.974</td>
<td>5.49</td>
</tr>
<tr>
<td>$\widehat{U}_{T, eq}$</td>
<td>0.960</td>
<td>87.46</td>
<td>0.961</td>
<td>72.79</td>
<td>0.962</td>
<td>46.44</td>
<td>0.961</td>
<td>28.03</td>
</tr>
<tr>
<td>ILR</td>
<td>0.942</td>
<td>80.94</td>
<td>0.946</td>
<td>55.20</td>
<td>0.965</td>
<td>23.55</td>
<td>0.993</td>
<td>4.93</td>
</tr>
<tr>
<td>$\lambda_0 = 0.2$</td>
<td>HDR</td>
<td>0.905</td>
<td>72.26</td>
<td>0.913</td>
<td>50.61</td>
<td>0.933</td>
<td>25.07</td>
<td>0.973</td>
</tr>
<tr>
<td>Bai (1997)</td>
<td>0.829</td>
<td>65.56</td>
<td>0.899</td>
<td>41.42</td>
<td>0.932</td>
<td>19.62</td>
<td>0.966</td>
<td>5.55</td>
</tr>
<tr>
<td>$\widehat{U}_{T, eq}$</td>
<td>0.962</td>
<td>88.77</td>
<td>0.968</td>
<td>78.61</td>
<td>0.963</td>
<td>57.87</td>
<td>0.965</td>
<td>29.88</td>
</tr>
<tr>
<td>ILR</td>
<td>0.938</td>
<td>83.24</td>
<td>0.951</td>
<td>63.66</td>
<td>0.972</td>
<td>28.94</td>
<td>0.994</td>
<td>5.16</td>
</tr>
</tbody>
</table>

The model is $y_t = \beta^0 + \delta^0 1_{\{t>T\lambda_0\}} + \epsilon_t, \epsilon_t = 0.3\epsilon_{t-1} + u_t, u_t \sim i.i.d. \mathcal{N}(0, 0.49), T = 100$. The notes of Table 2 apply.
The model is \( y_t = \nu^0 + Z_t \beta^0 + Z_t \delta^0 1_{\{t > \lfloor T \lambda_0 \rfloor\}} + \epsilon_t, \quad X_t = 0.5 X_{t-1} + \epsilon_t, \quad u_t \sim i.i.d. \mathcal{N}(0, 0.75), \quad \epsilon_t \sim i.i.d. \mathcal{N}(0, 1), \quad T = 100. \) The notes of Table 2 apply.
Table 7: Small-sample coverage rate and length of the confidence set for model M6

<table>
<thead>
<tr>
<th></th>
<th>$\delta^0 = 0.3$</th>
<th>$\delta^0 = 0.6$</th>
<th>$\delta^0 = 1$</th>
<th>$\delta^0 = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_0 = 0.5$</td>
<td>HDR</td>
<td>0.920</td>
<td>77.03</td>
<td>0.923</td>
</tr>
<tr>
<td>Bai (1997)</td>
<td>0.690</td>
<td>56.73</td>
<td>0.716</td>
<td>41.63</td>
</tr>
<tr>
<td>$\delta^0 = 0.6$</td>
<td>HDR</td>
<td>0.962</td>
<td>87.76</td>
<td>0.962</td>
</tr>
<tr>
<td>$\delta^0 = 0.6$</td>
<td>HDR</td>
<td>0.790</td>
<td>71.07</td>
<td>0.805</td>
</tr>
<tr>
<td>$\lambda_0 = 0.35$</td>
<td>HDR</td>
<td>0.928</td>
<td>76.41</td>
<td>0.925</td>
</tr>
<tr>
<td>Bai (1997)</td>
<td>0.691</td>
<td>55.18</td>
<td>0.720</td>
<td>40.25</td>
</tr>
<tr>
<td>$\delta^0 = 0.35$</td>
<td>HDR</td>
<td>0.953</td>
<td>87.76</td>
<td>0.953</td>
</tr>
<tr>
<td>$\delta^0 = 0.35$</td>
<td>HDR</td>
<td>0.795</td>
<td>71.34</td>
<td>0.804</td>
</tr>
<tr>
<td>$\lambda_0 = 0.2$</td>
<td>HDR</td>
<td>0.915</td>
<td>75.86</td>
<td>0.919</td>
</tr>
<tr>
<td>Bai (1997)</td>
<td>0.707</td>
<td>55.03</td>
<td>0.770</td>
<td>39.77</td>
</tr>
<tr>
<td>$\delta^0 = 0.2$</td>
<td>HDR</td>
<td>0.951</td>
<td>88.48</td>
<td>0.952</td>
</tr>
<tr>
<td>$\delta^0 = 0.2$</td>
<td>HDR</td>
<td>0.795</td>
<td>72.01</td>
<td>0.809</td>
</tr>
</tbody>
</table>

The model is $y_t = \nu_0 + Z_t \beta_0 + Z_t \delta^0 1_{\{t > \lfloor T\lambda_0 \rfloor \}} + e_t, e_t = v_t |Z_t|, v_2 \sim i.i.d. N(0, 1), Z_t = 0.5Z_{t-1} + u_t, u_t \sim i.i.d. N(0, 1)$ $T = 100$. The notes of Table 2 apply.
Table 8: Small-sample coverage rate and length of the confidence set for model M7

<table>
<thead>
<tr>
<th>$\lambda_0$</th>
<th>HDR</th>
<th>$\delta^0 = 0.3$</th>
<th>Cov.</th>
<th>Lgth.</th>
<th>$\delta^0 = 0.6$</th>
<th>Cov.</th>
<th>Lgth.</th>
<th>$\delta^0 = 1$</th>
<th>Cov.</th>
<th>Lgth.</th>
<th>$\delta^0 = 2$</th>
<th>Cov.</th>
<th>Lgth.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>HDR</td>
<td>0.918</td>
<td>75.64</td>
<td>0.910</td>
<td>67.46</td>
<td>0.931</td>
<td>48.54</td>
<td>0.957</td>
<td>12.50</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bai (1997)</td>
<td>0.834</td>
<td>70.13</td>
<td>0.824</td>
<td>52.16</td>
<td>0.861</td>
<td>28.69</td>
<td>0.948</td>
<td>8.45</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\hat{U}_{T,T}$</td>
<td>0.959</td>
<td>88.62</td>
<td>0.959</td>
<td>78.87</td>
<td>0.959</td>
<td>58.60</td>
<td>0.952</td>
<td>30.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ILR</td>
<td>0.969</td>
<td>86.75</td>
<td>0.959</td>
<td>67.91</td>
<td>0.967</td>
<td>34.13</td>
<td>0.995</td>
<td>9.17</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.35</td>
<td>HDR</td>
<td>0.926</td>
<td>74.78</td>
<td>0.914</td>
<td>64.86</td>
<td>0.924</td>
<td>45.69</td>
<td>0.956</td>
<td>12.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bai (1997)</td>
<td>0.851</td>
<td>69.35</td>
<td>0.847</td>
<td>51.17</td>
<td>0.878</td>
<td>28.59</td>
<td>0.944</td>
<td>8.47</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\hat{U}_{T,T}$</td>
<td>0.964</td>
<td>88.82</td>
<td>0.960</td>
<td>79.74</td>
<td>0.964</td>
<td>60.26</td>
<td>0.964</td>
<td>30.64</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ILR</td>
<td>0.972</td>
<td>88.69</td>
<td>0.975</td>
<td>73.95</td>
<td>0.981</td>
<td>39.08</td>
<td>0.992</td>
<td>9.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td>HDR</td>
<td>0.909</td>
<td>78.12</td>
<td>0.921</td>
<td>61.87</td>
<td>0.933</td>
<td>40.66</td>
<td>0.961</td>
<td>11.70</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bai (1997)</td>
<td>0.824</td>
<td>65.23</td>
<td>0.867</td>
<td>51.35</td>
<td>0.915</td>
<td>29.83</td>
<td>0.955</td>
<td>8.70</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\hat{U}_{T,T}$</td>
<td>0.961</td>
<td>89.71</td>
<td>0.960</td>
<td>83.68</td>
<td>0.961</td>
<td>69.25</td>
<td>0.960</td>
<td>35.78</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ILR</td>
<td>0.966</td>
<td>91.48</td>
<td>0.971</td>
<td>82.78</td>
<td>0.984</td>
<td>51.93</td>
<td>0.995</td>
<td>10.87</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The model is $y_t = \beta^0 + \delta^0 1_{\{t > \lfloor T\lambda_0\rfloor\}} + e_t$, $e_t = 0.3e_{t-1} + u_t$, $u_t \sim i.i.d. t_\nu$, $\nu = 5$, $T = 100$. The notes of Table 2 apply.
Table 9: Small-sample coverage rate and length of the confidence set for model M8

<table>
<thead>
<tr>
<th></th>
<th>$\delta^0 = 0.3$</th>
<th>$\delta^0 = 0.6$</th>
<th>$\delta^0 = 1$</th>
<th>$\delta^0 = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_0 = 0.5$ HDR</td>
<td>0.918</td>
<td>75.08</td>
<td>0.913</td>
<td>60.44</td>
</tr>
<tr>
<td>Bai (1997) $\hat{U}_{T,\text{eq}}$</td>
<td>0.949</td>
<td>84.56</td>
<td>0.950</td>
<td>67.64</td>
</tr>
<tr>
<td>ILR</td>
<td>0.943</td>
<td>83.69</td>
<td>0.946</td>
<td>63.24</td>
</tr>
<tr>
<td>$\lambda_0 = 0.35$ HDR</td>
<td>0.919</td>
<td>74.16</td>
<td>0.916</td>
<td>58.53</td>
</tr>
<tr>
<td>Bai (1997) $\hat{U}_{T,\text{eq}}$</td>
<td>0.951</td>
<td>85.01</td>
<td>0.948</td>
<td>69.14</td>
</tr>
<tr>
<td>ILR</td>
<td>0.946</td>
<td>84.12</td>
<td>0.944</td>
<td>63.99</td>
</tr>
<tr>
<td>$\lambda_0 = 0.2$ HDR</td>
<td>0.912</td>
<td>73.43</td>
<td>0.929</td>
<td>56.18</td>
</tr>
<tr>
<td>Bai (1997) $\hat{U}_{T,\text{eq}}$</td>
<td>0.950</td>
<td>86.94</td>
<td>0.951</td>
<td>76.52</td>
</tr>
<tr>
<td>ILR</td>
<td>0.945</td>
<td>83.94</td>
<td>0.953</td>
<td>63.55</td>
</tr>
</tbody>
</table>

The model is $y_t = \delta^0 (1 - \nu^0) \mathbb{1}_{\{t > |T\lambda_0|\}} + \nu^0 y_{t-1} + \epsilon_t, \epsilon_t \sim \text{i.i.d. } \mathcal{N}(0, 0.49), \nu^0 = 0.3, T = 100$. The notes of Table 2 apply.
Table 10: Small-sample coverage rate and length of the confidence sets for model M9

<table>
<thead>
<tr>
<th>( \lambda_0 ) = 0.5</th>
<th>HDR</th>
<th>Cov.</th>
<th>Lgth.</th>
<th>Cov.</th>
<th>Lgth.</th>
<th>Cov.</th>
<th>Lgth.</th>
<th>Cov.</th>
<th>Lgth.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.903</td>
<td>61.09</td>
<td>0.927</td>
<td>31.14</td>
<td>0.930</td>
<td>18.33</td>
<td>0.930</td>
<td>9.10</td>
<td></td>
</tr>
<tr>
<td>Bai (1997)</td>
<td>0.791</td>
<td>37.86</td>
<td>0.831</td>
<td>17.73</td>
<td>0.855</td>
<td>10.43</td>
<td>0.868</td>
<td>5.30</td>
<td></td>
</tr>
<tr>
<td>( \hat{U}_T eq )</td>
<td>0.947</td>
<td>65.23</td>
<td>0.947</td>
<td>39.76</td>
<td>0.947</td>
<td>28.82</td>
<td>0.947</td>
<td>20.36</td>
<td></td>
</tr>
<tr>
<td>ILR</td>
<td>0.909</td>
<td>72.62</td>
<td>0.946</td>
<td>45.06</td>
<td>0.962</td>
<td>23.97</td>
<td>0.978</td>
<td>9.34</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( \lambda_0 ) = 0.35</th>
<th>HDR</th>
<th>Cov.</th>
<th>Lgth.</th>
<th>Cov.</th>
<th>Lgth.</th>
<th>Cov.</th>
<th>Lgth.</th>
<th>Cov.</th>
<th>Lgth.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.904</td>
<td>60.58</td>
<td>0.918</td>
<td>30.96</td>
<td>0.904</td>
<td>18.16</td>
<td>0.928</td>
<td>0.34</td>
<td></td>
</tr>
<tr>
<td>Bai (1997)</td>
<td>0.791</td>
<td>37.70</td>
<td>0.829</td>
<td>18.04</td>
<td>0.852</td>
<td>10.61</td>
<td>0.870</td>
<td>5.34</td>
<td></td>
</tr>
<tr>
<td>( \hat{U}_T eq )</td>
<td>0.942</td>
<td>66.27</td>
<td>0.942</td>
<td>40.63</td>
<td>0.942</td>
<td>29.39</td>
<td>0.942</td>
<td>20.67</td>
<td></td>
</tr>
<tr>
<td>ILR</td>
<td>0.922</td>
<td>72.20</td>
<td>0.947</td>
<td>45.27</td>
<td>0.959</td>
<td>24.93</td>
<td>0.973</td>
<td>8.55</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( \lambda_0 ) = 0.2</th>
<th>HDR</th>
<th>Cov.</th>
<th>Lgth.</th>
<th>Cov.</th>
<th>Lgth.</th>
<th>Cov.</th>
<th>Lgth.</th>
<th>Cov.</th>
<th>Lgth.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.920</td>
<td>61.37</td>
<td>0.946</td>
<td>31.00</td>
<td>0.942</td>
<td>20.44</td>
<td>0.944</td>
<td>9.04</td>
<td></td>
</tr>
<tr>
<td>Bai (1997)</td>
<td>0.791</td>
<td>39.23</td>
<td>0.841</td>
<td>19.28</td>
<td>0.876</td>
<td>11.99</td>
<td>0.886</td>
<td>6.16</td>
<td></td>
</tr>
<tr>
<td>( \hat{U}_T eq )</td>
<td>0.934</td>
<td>71.42</td>
<td>0.931</td>
<td>47.53</td>
<td>0.934</td>
<td>34.12</td>
<td>0.934</td>
<td>24.06</td>
<td></td>
</tr>
<tr>
<td>ILR</td>
<td>0.920</td>
<td>72.68</td>
<td>0.935</td>
<td>49.61</td>
<td>0.959</td>
<td>27.90</td>
<td>0.972</td>
<td>10.01</td>
<td></td>
</tr>
</tbody>
</table>

The model is \( y_t = \nu + Z_t \beta_0 + Z_t \delta_0 1_{\{t > \lceil T \lambda_0 \rceil \}} + e_t \cdot \left. \right. \sim \text{i.i.d. } \mathcal{N}(1, 1.44), \{e_t\} \) 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

<table>
<thead>
<tr>
<th>( \lambda_0 = 0.5 )</th>
<th>HDR</th>
<th>Cov.</th>
<th>Lgth.</th>
<th>Cov.</th>
<th>Lgth.</th>
<th>Cov.</th>
<th>Lgth.</th>
<th>Cov.</th>
<th>Lgth.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bai (1997)</td>
<td></td>
<td>0.809</td>
<td>45.33</td>
<td>0.844</td>
<td>17.11</td>
<td>0.864</td>
<td>8.27</td>
<td>0.883</td>
<td>3.61</td>
</tr>
<tr>
<td>( \hat{U}_T.\text{eq} )</td>
<td>0.959</td>
<td>72.69</td>
<td>0.959</td>
<td>39.81</td>
<td>0.959</td>
<td>24.25</td>
<td>0.959</td>
<td>14.79</td>
<td></td>
</tr>
<tr>
<td>ILR</td>
<td>0.929</td>
<td>83.23</td>
<td>0.951</td>
<td>69.67</td>
<td>0.971</td>
<td>44.40</td>
<td>0.987</td>
<td>10.44</td>
<td></td>
</tr>
<tr>
<td>( \lambda_0 = 0.35 )</td>
<td>HDR</td>
<td>0.934</td>
<td>73.08</td>
<td>0.937</td>
<td>35.37</td>
<td>0.923</td>
<td>13.68</td>
<td>0.920</td>
<td>4.55</td>
</tr>
<tr>
<td>Bai (1997)</td>
<td>0.821</td>
<td>45.70</td>
<td>0.838</td>
<td>17.78</td>
<td>0.867</td>
<td>8.53</td>
<td>0.889</td>
<td>3.71</td>
<td></td>
</tr>
<tr>
<td>( \hat{U}_T.\text{eq} )</td>
<td>0.964</td>
<td>76.14</td>
<td>0.964</td>
<td>44.61</td>
<td>0.965</td>
<td>27.33</td>
<td>0.964</td>
<td>15.84</td>
<td></td>
</tr>
<tr>
<td>ILR</td>
<td>0.934</td>
<td>81.32</td>
<td>0.959</td>
<td>62.98</td>
<td>0.977</td>
<td>34.38</td>
<td>0.984</td>
<td>9.12</td>
<td></td>
</tr>
<tr>
<td>( \lambda_0 = 0.2 )</td>
<td>HDR</td>
<td>0.941</td>
<td>71.46</td>
<td>0.959</td>
<td>59.03</td>
<td>0.950</td>
<td>15.39</td>
<td>0.919</td>
<td>5.03</td>
</tr>
<tr>
<td>Bai (1997)</td>
<td>0.818</td>
<td>47.82</td>
<td>0.872</td>
<td>20.44</td>
<td>0.878</td>
<td>9.60</td>
<td>0.873</td>
<td>3.92</td>
<td></td>
</tr>
<tr>
<td>( \hat{U}_T.\text{eq} )</td>
<td>0.971</td>
<td>82.40</td>
<td>0.971</td>
<td>59.03</td>
<td>0.971</td>
<td>39.02</td>
<td>0.972</td>
<td>20.42</td>
<td></td>
</tr>
<tr>
<td>ILR</td>
<td>0.928</td>
<td>83.26</td>
<td>0.952</td>
<td>70.03</td>
<td>0.964</td>
<td>42.65</td>
<td>0.982</td>
<td>10.30</td>
<td></td>
</tr>
</tbody>
</table>

The model is \( y_t = \nu^0 + Z_t\beta^0 + Z_t\delta^01_{\{t > \lfloor T\lambda_0 \rfloor \}} + \epsilon_t, \epsilon_t \sim i.i.d. \mathcal{N}(0, 1), Z_t \sim \text{ARFIMA}(0.3, 0.6, 0), T = 100. \) The notes of Table 2 apply.