Essex Finance Centre
Working Paper Series

Working Paper No 52: 12-2019

"Level Shift Estimation in the Presence of Non-stationary Volatility with an Application to the Unit Root Testing Problem"

"David Harris, Hsein Kew, A.M. Robert Taylor"

Essex Business School, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ
Web site: http://www.essex.ac.uk/ebs/

Level Shift Estimation in the Presence of Non-stationary Volatility with an Application to the Unit Root Testing Problem*

David Harris[a], Hsein Kew[b] and A.M. Robert Taylor[c]

[a] Department of Economics, University of Melbourne

[b] Department of Econometrics and Business Statistics, Monash University

[c] Essex Business School, University of Essex

Abstract

This paper focuses on the estimation of the location of level breaks in time series whose shocks display non-stationary volatility (permanent changes in unconditional volatility). We propose a new feasible weighted least squares (WLS) estimator, based on an adaptive estimate of the volatility path of the shocks. We show that this estimator belongs to a generic class of weighted residual sum of squares which also contains the ordinary least squares (OLS) and WLS estimators, the latter based on the true volatility process. For fixed magnitude breaks we show that the consistency rate of the generic estimator is unaffected by non-stationary volatility. We also provide local limiting distribution theory for cases where the break magnitude is either local-to-zero at some polynomial rate in the sample size or is exactly zero. The former includes the Pitman drift rate which is shown via Monte Carlo experiments to predict well the key features of the finite sample behaviour of both the OLS and our feasible WLS estimators. The simulations highlight the importance of the break location, break magnitude, and the form of non-stationary volatility for the finite sample performance of these estimators, and show that our proposed feasible WLS estimator can deliver significant improvements over the OLS estimator under heteroskedasticity. We discuss how these results can be applied, by using level break fraction estimators on the first differences of the data, when testing for a unit root in the presence of trend breaks and/or non-stationary volatility. Methods to select between the break and no break cases, using standard information criteria and feasible weighted information criteria based on our adaptive volatility estimator, are also discussed. Simulation evidence suggests that unit root tests based on these weighted quantities can display significantly improved finite sample behaviour under heteroskedasticity relative to their unweighted counterparts. An empirical illustration to U.S. and U.K. real GDP is also considered.

Keywords: Level break fraction, non-stationary volatility, adaptive estimation, feasible weighted estimator, information criteria, unit root tests and trend breaks.

JEL: C12, C22.

*We thank the Guest Editor, Tong Li, two anonymous referees, Giuseppe Cavaliere, and participants at the International Conference in Celebration of the 65th Birthday of Maxwell King for their helpful and constructive comments on earlier versions of this paper. Taylor gratefully acknowledges financial support provided by the Economic and Social Research Council of the United Kingdom under research grant ES/M01147X/1. Correspondence to: Robert Taylor, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, United Kingdom. Email: rtaylor@essex.ac.uk.

1 Introduction

Breaks in the deterministic trend function appear prevalent in macroeconomic series; see, *inter alia*, Stock and Watson (1996,1999,2005) and Perron and Zhu (2005). The impact of these on standard unit root tests has been well known since Perron (1989) who treated the location of potential breaks as known. Subsequent approaches have focused on the case where the break date is unknown and is replaced by a break fraction estimator; see, *inter alia*, Perron (1997) and Perron and Rodríguez (2003). Harris *et al.* (2009) and Carrion-i-Silvestre *et al.* (2009), among others, extend these approaches to incorporate pre-test procedures for the presence of a trend break. Under a fixed magnitude trend break the break fraction estimator used in these procedures needs to be consistent at a rate *faster* than $T^{-1/2}$, T denoting the sample size, for the resulting unit root test to be asymptotically valid where a trend break occurs. As a result, the ordinary least squares [OLS] level break estimator of Bai (1994) has tended to be applied to the first differences of the series because it is consistent at rate T^{-1} for the true break fraction where a break occurs, while the corresponding OLS-based estimator based on the levels is only consistent at rate $T^{-1/2}$.

The aforementioned procedures do not allow for time-varying behaviour in the unconditional volatility (often referred to as non-stationary volatility) of the driving shocks. This is an important practical drawback given that a large number of empirical studies have reported a substantial decline, often referred to as the Great Moderation, in the unconditional volatility of the shocks driving macroeconomic series in the twenty years or so leading up to the Great Recession that started in late 2007, with a subsequent sharp increase again in volatility observed after 2007; see, *inter alia*, McConnell and Perez-Quiros (2000), Clark (2009), Stock and Watson (2012), and the references therein. Cavaliere *et al.* (2011) refine the approach of Harris *et al.* (2009) to use wild bootstrap unit root tests. However, their procedures are still based around applying the OLS level break fraction estimator of Bai (1994) to the first differences and trend break pre-test, each of which were developed for homoskedastic innovations. While they show that both of these are asymptotically robust to time-varying volatility, their finite sample efficacy will clearly have important forward implications for the behaviour of the resulting unit root tests.

Our principal aim here is to explore the properties of the OLS level break estimator of Bai (1994) in cases where the shocks can display non-stationary volatility and to develop and explore the properties of a corresponding feasible weighted least squares [WLS] level break estimator based around the use of data which have been weighted by a non-parametric estimate of the volatility path. We will consider a very general set-up for the volatility process which allows, for example, single and multiple abrupt variance breaks, smooth transition variance breaks, and trending variances. The feasible WLS estimator we propose uses adaptive methods to estimate the volatility path of the shocks. Adaptive methods have been successfully employed in a number of areas of the literature including inference on the parameters of finite-order unconditionally heteroskedastic autoregressive models by Phillips and Xu (2006) and Xu and Phillips (2008), testing for ARCH effects in unconditionally heteroskedastic autoregressive models by Patilea and Raïssi (2014), testing for long memory in unconditionally heteroskedastic ARFIMA models by Harris and Kew (2017) and adaptive testing for autocorrelation in Harris and Kew (2014), and for adaptive estimation of VAR models in Patilea and Raïssi (2012, 2013). Of most relevance to this paper, Boswijk and Zu (2018) propose an adaptive estimator of the unconditional variance process in the context of testing for a unit root in an autoregressive model driven by heteroskedastic errors, although no allowance is made for the possibility of a trend break.

We establish the large sample properties of the OLS and feasible WLS break fraction estimators under a variety of assumptions on the magnitude of the level shift. For level shifts of either fixed (non-zero) magnitude or where the magnitude is local-to-zero at a rate slower than the Pitman rate of $T^{-1/2}$, we demonstrate the consistency of these estimators, and indeed those from a generic class of residual sums of squares [RSS] based estimators. The consistency rate for the OLS and feasible WLS estimators coincides and is unaffected by the location of the break or by time variation in the volatility process. We also derive the asymptotic distributions of these estimators where the magnitude of the level shift lies within a Pitman neighbourhood of zero. Elliott and Müller (2007) argue that the finite sample behaviour of break fraction estimators such as those considered in this paper is likely to be far better approximated for the sort of break magnitudes typically encountered in practice by asymptotic theory based on the Pitman rate rather than a fixed magnitude break. Our results accord with this view. Under Pitman drift the limiting distributions of the OLS and feasible WLS estimators are shown to differ and to depend on the location and (local drift) magnitude of the level break and, to differing extents, on the time path of the volatility process.

We investigate and compare the finite sample behaviour of the estimators using Monte Carlo simulation. These agree closely with the qualitative predictions drawn from the limiting distributions under Pitman drift. In particular, they show that a break fraction estimator can be erroneously drawn towards the most volatile parts of a time series, potentially away from a genuine level break. They highlight that unmodelled heteroskedasticity can result in large *bias* and other distributional distortions in break fraction estimation for various configurations of the break location and time path of volatility, and that the consequences may be more severe than just loss of estimator efficiency as occurs in more standard statistical settings. The results also show that the feasible WLS estimator can deliver substantial improvements over the OLS estimator in certain heteroskedastic environments, most notably where the level break occurs in a low volatility regime.

It would be unusual that a break fraction is the final quantity of interest in a time series analysis rather than an input into subsequent inference. As an application, we also investigate to what extent the improved behaviour of the feasible WLS estimator relative to the standard OLS estimator of Bai (1994) when non-stationary volatility is present can effect improvements in the finite sample behaviour of the unit root tests discussed above. Here we also propose model selection methods based on the familiar Schwarz (1978) criterion to choose between the trend break and no trend break models in the practically relevant case where it is unknown if a trend break is present. We discuss such information criteria based on both OLS and feasible WLS model estimation, the latter using the same adaptive estimator of the unconditional variance process as for the feasible WLS break fraction estimator. Simulation evidence suggests that the use of these feasible weighted quantities can deliver unit root tests with significantly improved finite sample behaviour in the presence of non-stationary volatility relative to using their unweighted counterparts.

The paper is organised as follows. Our reference heteroskedastic level break model is outlined in section 2. Section 3 details a generic RSS level break fraction estimator which contains the standard OLS estimator of Bai (1994) and the infeasible WLS estimator as special cases. Here we also show how a feasible version of the WLS estimator can be constructed, using an adaptive estimator of the volatility path of the innovations. The large sample properties of these estimators are compared for both fixed, local and zero magnitude level shifts. The relative finite sample properties of these estimators in both homoskedastic and a variety of heteroskedastic environments are explored. Section 4 discusses the application of level break estimation methods to the problem of unit root testing when breaks in trend and/or volatility may be present. In section 5 we illustrate the methods discussed in the paper with an application to U.S. and U.K. real GDP. For both series, OLS estimation estimates a break early in the data in a high volatility period whereas the feasible WLS estimator estimates a much later breakdate in a relatively low volatility regime. For the case of the U.K. data, this alters the outcome of conventional unit root tests allowing the unit root null hypothesis to be rejected when based on the trend break date estimate by the feasible WLS estimator. Section 6 concludes. Supporting material, including mathematical proofs, is provided in an on-line supplementary appendix.

In what follows, '$\lfloor \cdot \rfloor$' denotes the integer part and '$1_{(.)}$' denotes the indicator function. The symbols '$\xrightarrow{d}$' and '$\xrightarrow{p}$' are used to denote convergence in distribution and probability respectively as $T \to \infty$. The maximum and minimum of a and b are denoted $a \vee b$ and $a \wedge b$, respectively. Finally, $\mathcal{D} := D[0,1]$ denotes the space of right continuous with left limit (càdlàg) processes on $[0,1]$.

2 The Heteroskedastic Level Break Model

We consider the time series process $\{y_t\}$ generated according to the following level break model,

$$y_t = \mu + \delta_T \cdot 1_{t > \lfloor \tau_0 T \rfloor} + e_t, \quad t = 1, ..., T \tag{2.1}$$

$$e_t = \sigma_t \varepsilon_t. \tag{2.2}$$

Equation (2.1) comprises a constant, a level shift at time $\lfloor \tau_0 T \rfloor$, and a stochastic component e_t. As is standard, for the purposes of the large sample results which follow, the level shift is taken to occur at a fixed fraction of the sample size, τ_0, with $0 < \tau_L \leq \tau_0 \leq \tau_U < 1$.

We follow Elliott and Müller (2007) and parameterise the break magnitude parameter as $\delta_T := \delta T^{-d}$ with δ a fixed constant and $d \geq 0$. For a given value of T a level break exists in y_t only if $\delta \neq 0$. No break occurs when $\delta = 0$, regardless of d, while a level break of fixed magnitude δ occurs when $d = 0$ and $\delta \neq 0$. In the unconditionally homoskedastic case, where $\sigma_t = \sigma$ for all t, Bai (1997), shows that when $\delta \neq 0$, then τ_0 is consistently estimated by OLS for any $0 \leq d < 1/2$.[1] In particular, although the magnitude of the level break shrinks here as the sample size increases, the level break is still sufficiently large for the location of the break, τ_0, to be consistently estimated and for consistent tests for a level break to exist. In contrast, $d = 1/2$ gives the Pitman drift rate for this problem such that τ_0 cannot be consistently estimated nor can a consistent test for the presence of a level break be obtained. We will show that these consistency rates in d also hold in the heteroskedastic case we focus on here.

[1]The consistency results given in Bai (1997) also hold in the case where σ_t displays a one-time break, provided it occurs at the same break fraction, τ_0, as the level break.

To complete the specification of (2.1)-(2.2) the following conditions, collectively labelled Assumption $\mathcal{A}$, will be assumed to hold on e_t.

Assumption $\mathcal{A}$.

$\mathcal{A}_1$. *The innovations* $\{\varepsilon_t\}$ *form a martingale difference sequence with respect to the filtration* $\mathcal{F}_t$, *where* $\mathcal{F}_{t-1} \subseteq \mathcal{F}_t$ *for* $t = ..., -1, 0, 1, 2, ...$, *satisfying: (i) the global homoskedasticity condition:* $\frac{1}{T}\sum_{t=1}^{T} E\left(\varepsilon_t^2 | \mathcal{F}_{t-1}\right) \xrightarrow{p} 1$, *and* (ii) $E|\varepsilon_t|^r < K < \infty$ *for some* $r \geq 4$;

$\mathcal{A}_2$. *The volatility term* σ_t *satisfies* $\sigma_t = \sigma(t/T)$, *where* $\sigma(\cdot) \in \mathcal{D}$ *is non-stochastic, bounded above and below as* $0 < \underline{\sigma} \leq \sigma(s) \leq \bar{\sigma} < \infty$ *for all* s, *and satisfies a Lipschitz condition except at a finite number of points of discontinuity.*

Remark 2.1. The process $\{e_t\}$ in (2.2) is formed as the product of two components, $\{\varepsilon_t\}$ and $\{\sigma_t\}$. The former is assumed to satisfy a relatively weak globally stationary martingale difference assumption which allows for certain forms of conditional heteroskedasticity, such as that arising from stationary GARCH models, in the errors; see Davidson (1994, pp.454-455) for further discussion. It should be noted, however, that we will later require the additional assumption of conditional homoskedasticity for the feasible WLS break fraction estimator considered in section 3.4. Notice that, under Assumption $\mathcal{A}$, e_t has mean zero and time-varying unconditional variance, σ_t^2. □

Remark 2.2. Assumption $\mathcal{A}_2$ casts the dynamics of the disturbance variance in a quite general framework, requiring it only to be non-stochastic, bounded and to be smooth in between a countable number of jumps. A detailed discussion of the class of variance processes allowed under $\mathcal{A}_2$ is given in Cavaliere and Taylor (2007). They show that this includes variance processes displaying multiple volatility shifts, polynomially (possibly piecewise) trending volatility and smooth transition variance breaks, among other things. In the case where volatility displays jumps, these are not constrained to be located at the same point in the sample as the putative level shift, nor indeed are they required to lie within the search set, Λ. The conventional unconditionally homoskedastic assumption that $\sigma_t = \sigma$ for all t, also satisfies Assumption $\mathcal{A}_2$, since here $\sigma(s) = \sigma$ for all s. □

Remark 2.3. To focus our attention on the impact of non-stationary volatility on level break estimation, we have omitted autocorrelation in the model for the disturbance e_t. We will, however, discuss generalisations to allow for this at relevant points in the text. □

3 Level Break Fraction Estimation

3.1 Residual Sum of Squares Break Fraction Estimator

In what follows we define a generic RSS-based level break fraction estimator which contains weighted and unweighted break fraction estimators as special cases. To that end, define the weights x_t, $t = 1, ..., T$, and a generic RSS-based estimator

$$\hat{\tau} := \arg \min_{\tau \in [\tau_L, \tau_U]} \sum_{t=1}^{T} \hat{e}_{\tau,t}^{*2} \tag{3.1}$$

where, for any $\tau \in [\tau_L, \tau_U] \subset [0, 1]$, the residuals $\hat{e}_{\tau,t}^*$ are obtained from the OLS regression

$$y_t^* = \hat{\mu}_\tau x_t + \hat{\delta}_\tau (1_{t > \lfloor \tau T \rfloor} \cdot x_t) + \hat{e}_{\tau,t}^* \tag{3.2}$$

where $y_t^* := x_t y_t$.[2] Setting $x_t := 1$ for $t = 1, ..., T$, in (3.2) yields the usual OLS estimator of τ_0 considered in Bai (1994), while setting $x_t := 1/\sigma_t$, $t = 1, ..., T$, yields the infeasible WLS estimator that obtains with knowledge of σ_t. In what follows, where we wish to make reference to the OLS and WLS estimators specifically, rather than the generic RSS-based estimator in (3.1), we will use the notation $\hat{\tau}_{OLS}$ and $\hat{\tau}_{WLS}$, respectively. The assumption of non-stochastic weights will be relaxed in section 3.4 when we detail our feasible WLS estimator of τ_0 based on adaptive estimation of σ_t.

3.2 Asymptotic Behaviour of $\hat{\tau}$ under Large Breaks

We first analyse the large sample behaviour of $\hat{\tau}$ in the case where the trend break magnitude is "large" in that it can be either non-zero and fixed or is such that it is local-to-zero but at a rate slower than the Pitman drift rate of $T^{-1/2}$. We will show that the standard OLS estimator of τ_0 retains the consistency property established under unconditional homoskedasticity in Bai (1997) and that the rate also holds for the corresponding WLS estimator, and indeed for any of a wide class of weights. These results are now formally stated in Theorem 1.

Theorem 1. *Let y_t be generated according to* (2.1) *with $\delta_T := \delta T^{-d}$ and let Assumption $\mathcal{A}$ hold. Moreover let the non-stochastic weights, $x_t = x(t/T)$, $t = 1, ..., T$, used in constructing $\hat{\tau}$ of* (3.1) *be such that $x(.)$ satisfies the same conditions as $\sigma(.)$ given in Assumption $\mathcal{A}_2$. Then if $\delta \neq 0$ and $0 \leq d < 1/2$, it holds that $\hat{\tau} \xrightarrow{p} \tau_0$. Moreover, if $\delta \neq 0$ and $0 < d < 1/2$ then*

$$\frac{T\delta_T^2}{\underline{\sigma}(\tau_0)^2}(\hat{\tau} - \tau_0) \xrightarrow{d} \arg\max_{s \in (-\infty, \infty)} Z(s), \tag{3.3}$$

where

$$Z(s) := \begin{cases} W_1(-s) - \frac{|s|}{2}, & s \leq 0 \\ \sqrt{\phi} W_2(s) - \xi \frac{|s|}{2}, & s > 0 \end{cases}$$

in which W_1 and W_2 are independent standard Brownian motions each on $[0, \infty)$, and

$$\phi := \frac{\bar{\sigma}(\tau_0)^2 \bar{x}(\tau_0)^4}{\underline{\sigma}(\tau_0)^2 \underline{x}(\tau_0)^4}, \quad \xi := \left(\frac{\bar{x}(\tau_0)}{\underline{x}(\tau_0)}\right)^2,$$

where $\bar{\sigma}(\tau_0) := \lim_{\tau \downarrow \tau_0} \sigma(\tau)$, $\underline{\sigma}(\tau_0) := \lim_{\tau \uparrow \tau_0} \sigma(\tau)$, $\bar{x}(\tau_0) := \lim_{\tau \downarrow \tau_0} x(\tau)$ and $\underline{x}(\tau_0) := \lim_{\tau \uparrow \tau_0} x(\tau)$.

Remark 3.1. Theorem 1 implies that $\hat{\tau}$ is a consistent estimator for τ_0 at rate $O_p(T^{-1}\delta_T^{-2})$ for any $0 < d < 1/2$, irrespective of any conditional or unconditional heteroskedasticity present in σ_t satisfying Assumption $\mathcal{A}$, or the form of the weights, x_t, used in its construction. Moreover, the non-stationary volatility process $\sigma(\cdot)$ has no effect (other than scaling) on the asymptotic distribution of $\hat{\tau}$, with the exception of the situation where a jump (from $\underline{\sigma}(\tau_0)$ to $\bar{\sigma}(\tau_0)$) occurs at τ_0 and affects the term ϕ in $Z(s)$. When the break magnitude is "large", *only the single value of the variance process* $\sigma(\tau_0)$ (or the two values $\underline{\sigma}(\tau_0)$ and $\bar{\sigma}(\tau_0)$, if they differ) features in this asymptotic approximation. The intuition behind this is that the limit in (3.3) is derived from a functional central limit theorem [FCLT] applied only to observations within a shrinking neighbourhood of τ_0. The càdlàg assumption

[2]The form of estimated coefficients $\hat{\mu}_\tau$ and $\hat{\delta}_\tau$ obviously depend on the choice of x_t but this is omitted from the notation for brevity.

on $\sigma(\cdot)$ therefore implies that all such observations will have (asymptotically) the same variance. As we will see in section 3.3, this contrasts with the "small" breaks asymptotic approximation, in which the asymptotic distribution of $\hat{\tau}$ depends on the *entire sample path* of the volatility function. This turns out to be an important point of difference when evaluating the finite sample relevance of the two asymptotic approximations. □

Remark 3.2. Theorem 1 extends the results of Bai (1994) to cover both weighted and unweighted level break estimators and to allow for the general form of unconditional heteroskedasticity permitted in σ_t under Assumption $\mathcal{A}_2$. Bai (1997) establishes the same $O_p(T^{-1}\delta_T^{-2})$ rate in regression models (including (3.2)) allowing for weak dependence and conditional heteroskedasticity in the shocks, the latter of a similar form to that allowed under Assumption $\mathcal{A}_1$. We specify martingale difference disturbances here in to order to focus attention on the role of non-stationary volatility in this model, but it can also be shown that the $O_p(T^{-1}\delta_T^{-2})$ consistency rate given in Theorem 1 continues to hold when e_t is autocorrelated. For example, if $e_t = \sigma_t u_t$ where, as in equation (2) of Bai (1994), u_t is generated by a linear process $u_t = C(L)\varepsilon_t$, where $C(L) := \sum_{j=0}^{\infty} c_j L^j$ satisfies the standard summability condition (assumption B of Bai, 1994) $\sum_{j=0}^{\infty} j|c_j| < \infty$, and ε_t and σ_t again satisfy the conditions in Assumption $\mathcal{A}$, then the short run variance, $\underline{\sigma}(\tau_0)^2$, in (3.3) would simply need to be replaced by the corresponding long run variance, $\underline{\sigma}(\tau_0)^2 C(1)^2$. □

Remark 3.3. The role of the weighting factor x_t in Theorem 1 is qualitatively the same as that of σ_t. The weights make no difference to the asymptotic distribution of $\hat{\tau}$, but again with the one exception where a break in $x(s)$ occurs at τ_0 and hence influences the parameters ϕ and ξ. □

To illustrate the single special case of Theorem 1 for which the form of heteroskedasticity and weighting influence the asymptotic approximation, consider a volatility process of the form $\sigma_t = 1 + 1_{t > \lfloor T\tau_0 \rfloor}$, which has a break at the same time $\lfloor T\tau_0 \rfloor$ as the level break, implying $\overline{\sigma}(\tau_0) = 2$ and $\underline{\sigma}(\tau_0) = 1$. The OLS estimator $\hat{\tau}_{OLS}$ of the level break fraction τ_0 is defined by a continuous weighting function $x_t = 1$, and hence $\bar{x}(\tau_0) = \underline{x}(\tau_0) = 1$, which produces $\phi = 4$ and $\xi = 1$ in Theorem 1. The (infeasible) WLS estimator $\hat{\tau}_{WLS}$ has weighting function using $x_t = 1/\sigma_t$ which is discontinuous at τ_0, producing $\bar{x}(\tau_0) = 1/2, \underline{x}(\tau_0) = 1$ and hence $\phi = \xi = 1/4$ in Theorem 1. Bai (1997, Appendix B) shows that the density function $g(x)$ of $\arg\max_s Z(s)$ in (3.3) is

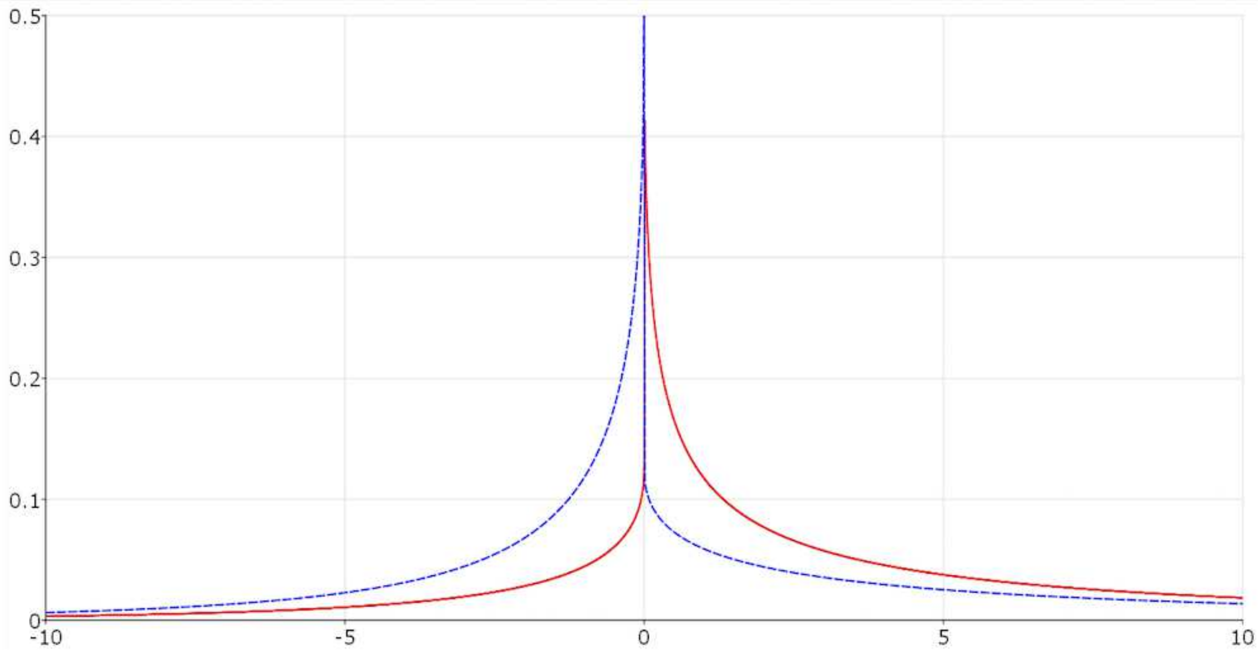
$$g(x) = \begin{cases} -\frac{1}{2}\Phi\left(-\frac{1}{2}\sqrt{|x|}\right) + \frac{1}{2}(1+2\alpha)\exp\left(\frac{1}{2}\alpha(1+\alpha)|x|\right)\Phi\left(-\frac{1}{2}(1+2\alpha)\sqrt{|x|}\right) & \text{if } x \le 0 \\ -\frac{1}{2}\beta^2\Phi\left(-\frac{1}{2}\beta\sqrt{x}\right) + \left(\xi + \frac{\beta^2}{2}\right)\exp\left(\frac{1}{2}(\phi+\xi)x\right)\Phi\left(-\left(\sqrt{\phi} + \frac{\beta}{2}\right)\sqrt{x}\right) & \text{if } x > 0 \end{cases}$$

where $\alpha := \xi/\phi$ and $\beta := \xi/\sqrt{\phi}$. The density functions are plotted for $\hat{\tau}_{OLS}$ (solid line) and $\hat{\tau}_{WLS}$ (dashed line) in Figure 1. Neither is symmetric around zero, with asymmetry induced when $\phi \neq 1$ and/or $\xi \neq 1$. Note that any form of weighting that is discontinuous at τ_0 will result in $\xi \neq 1$ and hence an asymmetric asymptotic distribution. This will hold regardless of the magnitude or the direction of the variance break, so long as the variance break occurs at τ_0.

The foregoing analysis begins to reveal that heteroskedasticity has a different effect on break fraction estimation (a non-standard statistical problem) than on the estimation of, for example, a classical linear regression coefficient. In the latter standard situation, unmodelled heteroskedasticity has no effect on the bias or asymptotic normality of least squares estimates but results only in loss of relative efficiency. By contrast, the effect of heteroskedasticity and weighting on break fraction estimation is not confined only to the variance of the estimator, but may affect various aspects of the entire sampling distribution, including its mean and/or symmetry and/or overall shape. This will be explained and illustrated in more detail once better asymptotic approximations are developed in section 3.3.

If the variance break does not occur at τ_0 (hence $\overline{\sigma}(\tau_0) = \underline{\sigma}(\tau_0)$ and both weighting functions are continuous at τ_0), the limiting distributions of $\hat{\tau}_{OLS}$ and $\hat{\tau}_{WLS}$ are exactly the same and are symmetric since $\phi = \xi = 1$. Also they coincide under homoskedasticity with the expression given in equation (5) of Jiang *et al.* (2018, p.158).

Figure 1: The density functions of $\arg\max_s Z(s)$ in (3.3)
$\hat{\tau}_{OLS}$ (solid line), $\hat{\tau}_{WLS}$ (dashed line)



3.3 Asymptotic Behaviour of $\hat{\tau}$ under Small Breaks

Elliott and Müller (2007) argue that the asymptotic behaviour of break fraction estimators such as $\hat{\tau}$ in (3.1) under "large" breaks is likely to provide a poor approximation to the finite sample properties of the estimator for the sort of break magnitudes typically encountered in practice. They argue that asymptotic theory based on the Pitman rate, $T^{-1/2}$, is likely to provide more accurate predictions for the behaviour of $\hat{\tau}$ in finite samples. They suggest that the asymptotics for $d = 1/2$ provides a continuous bridge between the no break case, $\delta = 0$, and the fixed magnitude break case considered in section 3.2. Accordingly, in Theorem 2 we now explore the asymptotic distribution theory for $\hat{\tau}$ in cases where the break magnitude can be "small" (i.e. $d \geq 1/2$) or, indeed, exactly zero ($\delta = 0$).

Theorem 2. *Let the conditions of Theorem 1 hold. Then for $d \geq 0$,*

$$\hat{\tau} \xrightarrow{d} \arg \max_{\tau \in [\tau_L, \tau_U]} Q(\tau; x(\cdot), \sigma(\cdot), \delta, d) \tag{3.4}$$

where

$$Q(\tau;x(\cdot),\sigma(\cdot),\delta,d) \;:=\; \left(1_{0\le d\le\frac{1}{2}}\frac{\delta}{\omega}(\chi(\tau_0)(1-\chi(\tau_0)))^{\frac{1}{2}}\left(\chi_1(\tau;\tau_0)\wedge\frac{1}{\chi_1(\tau;\tau_0)}\right)\right.$$
$$\left. - 1_{d\ge\frac{1}{2}}\frac{B_\eta(\tau)-\chi(\tau)B_\eta(1)}{(\chi(\tau)(1-\chi(\tau)))^{\frac{1}{2}}}\right)^2 \tag{3.5}$$

with $\omega^2 := (\int_0^1 x(s)^2 ds)^{-2}(\int_0^1 x(s)^4\sigma(s)^2 ds)$, $\chi(\tau) := \frac{\int_0^\tau x(s)^2 ds}{\int_0^1 x(s)^2 ds}$, $\eta(\tau) := \frac{\int_0^\tau x(s)^4\sigma(s)^2 ds}{\int_0^1 x(s)^4\sigma(s)^2 ds}$, *and* $\chi_1(\tau;\tau_0) := \left(\frac{\chi(\tau)/(1-\chi(\tau))}{\chi(\tau_0)/(1-\chi(\tau_0))}\right)^{1/2}$, *where* $B_\eta(\tau) = B(\eta(\tau))$, *with* $B(\cdot)$ *a standard Brownian motion, is a variance-transformed Brownian motion; see, for example, Davidson (1994).*

Theorem 2 establishes that $\hat{\tau}$ has a well-defined asymptotic distribution with support $\Lambda := [\tau_L, \tau_U]$ with its form depending on the increasing functions $\chi(\cdot) : [0,1] \mapsto [0,1]$ and $\eta(\cdot) : [0,1] \mapsto [0,1]$. The function $\chi(\tau)$ is the cumulative weighting function associated with the weighted regression (3.2). As regards $\eta(\tau)$, where $x_t = 1$, for all t, this function is the generalisation to weighted estimation of the *variance profile*, $(\int_0^1 \sigma(r)^2 dr)^{-1}\int_0^\tau \sigma(r)^2 dr$, of Cavaliere and Taylor (2007).

The constant ω^2 appearing in the first component of the right member of (3.5) is an asymptotic measure of the scaled disturbance variance in the weighted regression (3.2) and relates to the average level of the volatility in the weighted data. For $x_t = 1$ (the unweighted OLS estimator) it simplifies to $\omega^2 := \int_0^1 \sigma(r)^2 dr$ which, by Assumption $\mathcal{A}_2$, equals the limit of $T^{-1}\sum_{t=1}^T \sigma_t^2$, and may therefore be interpreted as the (asymptotic) average innovation variance. For $x_t = 1/\sigma_t$ (the infeasible WLS estimator), $\eta(\tau) = \omega^2\int_0^\tau \sigma(r)^{-2}dr$ and $\omega^2 = \left(\int_0^1 \sigma(r)^{-2}dr\right)^{-1}$. Notice that, for any given $\sigma(\cdot)$, the arithmetic/harmonic mean inequality implies that ω^2 is strictly greater for the OLS estimator than it is for the WLS estimator, with the exception of the case where $\sigma(s) = \sigma$ for all s, as holds under homoskedasticity, where they are equal. However this inequality need not imply an asymptotic *efficiency* gain for WLS relative to OLS, as it would in standard inference problems with asymptotic normal distribution theory. In this case the distributions are non-normal with unknown mean, so it is incomplete to consider only the variance as measure of estimator quality here. The sampling distributions for break fraction estimators under heteroskedasticity are considerably more complicated functions of nuisance parameters and such simple general conclusions cannot be drawn. Nevertheless, it will be shown by simulation in section 3.5 below that WLS can have substantially improved bias and general distributional properties than OLS under certain break location and heteroskedasticity configurations. However there are also particular cases in which OLS can be superior to WLS, even in the presence of heteroskedasticity, illustrating the complicated and non-standard nature of the distribution theory in Theorem 2.

Remark 3.4. In the case of the OLS estimator, $\hat{\tau}_{OLS}$, and under the Pitman drift rate, $T^{-1/2}$, the general result in Theorem 2 coincides under homoskedasticity with the expression given for $\hat{\tau}_{OLS}$ in Theorem 3 of Harvey *et al.* (2012, p.154). Notice also that the limiting function $Q(\tau;x(\cdot),\sigma(\cdot),\delta,d)$ appearing in Theorem 2 does not depend on any nuisance parameters arising from conditional heteroskedasticity in e_t satisfying the conditions in Assumption $\mathcal{A}_1$. □

Remark 3.5. As discussed in Remark 3.2, it is straightforward to extend the DGP to allow for autocorrelation in e_t. In that case the disturbances $e_t = \sigma_t u_t$ satisfy a heteroskedastic FCLT as usual, and ω^2 in Theorem 2 would become $\omega^2 = (\int_0^1 x(s)^2 ds)^{-2}(\int_0^1 x(s)^4\sigma(s)^2 ds)C(1)^2$. The implications of Theorem 2 are therefore qualitatively unchanged. □

Inspection of (3.5) shows that there are two components to the limiting $Q(\tau; x(\cdot), \sigma(\cdot), \delta, d)$ function. The first is non-stochastic and involves the true break fraction, τ_0, the ratio of the break magnitude parameter δ to ω, and the cumulative weighting function $\chi(\cdot)$. The second is stochastic and depends on the variance transformed Brownian motion $B_\eta(\cdot)$ (and hence the full volatility function $\sigma(.)$) and the cumulative weighting function, but not on either τ_0 or δ. Heuristically one may view these components as, respectively, the "signal" and the "noise" with respect to the estimation of τ_0. The relative importance of the two components of $Q(\tau; x(\cdot), \sigma(\cdot), \delta, d)$ depends on the localisation rate, d, and the break magnitude parameter, δ. We will now outline the following three possible cases of interest:

Case 1: $0 \le d < 1/2$, $\delta \neq 0$. This is a "large" break and hence the signal asymptotically dominates the noise. Theorem 2 implies that $\hat{\tau}$ converges to the maximiser of $\left(\chi_1(\tau;\tau_0) \wedge \chi_1(\tau;\tau_0)^{-1}\right)^2$, which is τ_0, which is the consistency result for $\hat{\tau}$ given in Theorem 1 for $0 \le d < 1/2$.

Case 2: $d = 1/2$, $\delta \neq 0$. The most interesting case is where the Pitman drift rate, $d = 1/2$, holds, and the "signal" and "noise" components have equivalent orders of magnitude. Here τ_0 cannot be consistently estimated, precisely because the signal does not dominate the noise, even asymptotically. The $Q(\tau; x(\cdot), \sigma(\cdot), \delta, d)$ function captures the trade-off between the signal and noise, and it is of course this trade-off that makes the Pitman-based local asymptotics useful for predicting the finite sample behaviour of $\hat{\tau}$. Now, because $\max_\tau \left(\chi_1(\tau;\tau_0) \wedge \chi_1(\tau;\tau_0)^{-1}\right)^2 = \chi(\tau_0;\tau_0)^2 = 1$, we may consider the scaling on the "signal" relative to the "noise" as being determined by the constant $\frac{\delta}{\omega}(\chi(\tau_0)(1-\chi(\tau_0)))^{\frac{1}{2}}$. In contrast to the "large" break asymptotics in section 3.2, the "small" break asymptotics predicts that the efficacy of $\hat{\tau}$ is not only related to the break size δ, but also to the average volatility across the whole sample (ω) (not just the level of volatility at the break location $\sigma(\tau_0)$) and to the form of the weighting scheme that determines $\chi(\tau)$. The constant $\chi(\tau_0)(1-\chi(\tau_0))$ is maximised for τ_0 satisfying $\chi(\tau_0) = \frac{1}{2}$, showing that the signal for weighted estimation of τ_0 is not necessarily highest at $\tau_0 = 0.5$, as it is for the unweighted estimator. Rather it is maximised at the value of τ_0 where the cumulative weighting reaches 0.5, i.e. $\int_0^{\tau_0} x(s)^2 ds = \frac{1}{2}\int_0^1 x(s)^2 ds$. In the supplementary appendix we provide calculations of these quantities for the two illustrative examples of a linear trend in variance and a single break in variance, together with some associated Monte Carlo simulation results for the latter example.

Case 3: $d > 1/2$ **and/or** $\delta = 0$. Consider finally the case where no trend break occurs (i.e. $\delta = 0$), or that the break is so small that the signal disappears from $Q(\tau; x(\cdot), \sigma(\cdot), \delta, d)$ asymptotically (i.e. $d > 1/2$). . Here the result in Theorem 2 implies that

$$\begin{aligned}\hat{\tau} &\xrightarrow{d} \arg\max_{\tau\in[\tau_L,\tau_U]} Q(\tau; x(\cdot), \sigma(\cdot), 0) \\ &= \arg\max_{\tau\in[\tau_L,\tau_U]} \frac{(B_\eta(\tau) - \chi(\tau)B_\eta(1))^2}{(\chi(\tau)(1-\chi(\tau)))} \qquad (3.6)\\ &= \arg\max_{\tau\in[\tau_L,\tau_U]} \frac{B_\eta(\tau)^2}{\chi(\tau)} + \frac{(B_\eta(1) - B_\eta(\tau))^2}{1-\chi(\tau)}. \qquad (3.7)\end{aligned}$$

The result in (3.7) coincides with the form of the distribution in part 1(a) of Theorem 3.1 of Nunes *et al.* (1995) specialised to the case of a level shift and generalised to allow for heteroskedasticity. The latter is also in the general form reported in Proposition 1 of Elliott and Müller (2007).

The OLS estimator, $\hat{\tau}_{OLS}$, applies equal weighting ($x_t = 1$) to the observations, implying $\chi(\tau) = \tau$. Under homoskedasticity ($\sigma_t = \sigma$) we have $\eta(\tau) = \tau$, in which case $Q(\tau; 1, \sigma(\cdot), 0)$ reduces to the square of a standard Brownian Bridge $B(\tau) - \tau B(1)$ divided by its standard deviation process, $(\tau(1-\tau))^{1/2}$. This scaled Brownian Bridge has a marginal standard normal distribution for each τ. In contrast, where unconditional heteroskedasticity is present, the limit $Q(\tau; 1, \sigma(\cdot), 0)$ in (3.6) involves the square of $(\tau(1-\tau))^{-1/2}(B_\eta(\tau) - \tau B_\eta(1))$ where $\eta(\tau) = \int_0^\tau \sigma(s)^2 ds / \int_0^1 \sigma(s)^2$ now differs from τ. Heuristically, this dependence suggests that the distribution of $\hat{\tau}_{OLS}$ will be significantly affected by the presence of unconditional heteroskedasticity. The WLS estimator, $\hat{\tau}_{WLS}$, applies weighting of the form $x_t = 1/\sigma_t$, implying that $\chi(\tau) = \eta(\tau) = \int_0^\tau \sigma(s)^{-2} ds / \int_0^1 \sigma(s)^{-2} ds$, and, hence, that $Q(\tau; 1/\sigma(\cdot), \sigma(\cdot), 0)$ is a function of the variance transformed Brownian Bridge $B_\eta(\tau) - \eta(\tau) B_\eta(1)$ divided by its standard deviation process, $(\eta(\tau)(1-\eta(\tau))^{1/2}$. As in the homoskedastic case, this latter scaled process has a marginal standard normal distribution for each τ. Although formally the asymptotic distribution of $\hat{\tau}_{WLS}$ depends on the joint distribution of $Q(.; 1/\sigma(\cdot), \sigma(\cdot), 0)$ on $[\tau_L, \tau_U]$, and, hence, will depend on $\sigma(\cdot)$ in some form, the marginal properties of the scaled process are suggestive that $\hat{\tau}_{WLS}$ will be less affected by any unconditional heteroskedasticity present in e_t than $\hat{\tau}_{OLS}$. This conjecture is supported by the simulation evidence reported in section 3.5.

3.4 A Feasible WLS Break Fraction Estimator

The WLS estimator, $\hat{\tau}_{WLS}$, outlined in section 3.1 is infeasible in practice because it requires knowledge of the volatility process, σ_t, $t = 1, ..., T$. It can, however, be made operational by replacing σ_t in the formulation of $\hat{\tau}_{WLS}$ by an estimate of σ_t. In practice the volatility process could be estimated either parametrically or non-parametrically. The former could be useful where the practitioner wishes to specify a particular model for the volatility process but of course has the drawback that an incorrectly specified model will likely give a very poor estimate of the volatility path. Given our focus in this paper is on setting up general assumptions on the heteroskedasticity present in the shocks without assuming a parametric model for the volatility process, it is more natural for us to consider a two-step approach based on a non-parametric (adaptive) estimator of the volatility process. In this approach the volatility, σ_t, is first estimated using the residuals from estimating the level break model as in (3.2) by standard OLS (i.e. treating the shocks as homoskedastic) and then substituting σ_t in the expression for $\hat{\tau}_{WLS}$ by the the resulting estimator, $\hat{\sigma}_t$, say. Our proposed estimator of σ_t is based on the approach developed in Hansen (1995) and Xu and Phillips (2008), which has recently been adapted to the unit root testing context by Boswijk and Zu (2018). We will demonstrate that the large sample behaviour of the resulting feasible weighted estimator coincides with that of the infeasible WLS estimator.

To that end, let $\hat{e}_{\hat{\tau},t} := y_t - \hat{\mu}_{\hat{\tau}} - \hat{\delta}_{\hat{\tau}} 1_{t > \lfloor \hat{\tau} T \rfloor}$, $t = 1, ..., T$, denote the standard OLS residuals which obtain from estimating (2.1) under the assumption that e_t is homoskedastic. In doing so an initial estimate of the level break location is needed. This could be provided by any form of the generic estimator $\hat{\tau}$ given in (3.1) such that the consistency result in Theorem 1 holds and a natural choice would be the simple OLS estimator, $\hat{\tau}_{OLS}$. Next let $K(\cdot)$ be a kernel function, and let $K_h(t) := K(t/h)$ with $h > 0$ a bandwidth. Then, given the residuals $\hat{e}_{\hat{\tau},t}$, and $K_h(t)$, a kernel smoothing estimator for σ_t^2 can be defined as

$$\hat{\sigma}_t^2 := \frac{\sum_{i=1}^{T} K_h\left(\frac{t-i}{T}\right)\hat{e}_{\hat{\tau},i}^2}{\sum_{i=1}^{T} K_h\left(\frac{t-i}{T}\right)}. \tag{3.8}$$

By choosing different kernel functions one can obtain either one-sided or two-sided smoothing. We will follow Xu and Phillips (2008) and set $K_h(0) = 0$, and also avoid the need for boundary value adjustments to (3.8) of the type discussed in Hansen (1995) by assuming the use of two-sided smoothing in what follows. In particular, we will assume that $K(\cdot)$ is a bounded non-negative function defined on the real line and is such that $\int_{-\infty}^{\infty} K(x)dx = 1$ and $0 < \int_0^{\infty} K(x)dx < 1$. The bandwidth, $h := h(T)$, is assumed to satisfy the (standard) rate condition that $h + (Th^2)^{-1} \to 0$ as $T \to \infty$. The practical implementation of the estimator $\hat{\sigma}_t^2$ depends on the choice of kernel function, $K(\cdot)$, and the bandwidth, h. Commonly used kernels which satisfy the stated conditions include the uniform, Epanechnikov, biweight and Gaussian functions. The bandwidth condition implies that $h \to 0$ but at a slower rate than $T^{-1/2}$. In practice bandwidth selection can be crucial to performance, and cross-validation and plug-in rules can be defined for h. The latter is used in the simulations in section 3.5 below.

If $\sigma(s)$ was continuous in $s \in [0, 1]$, then it would be possible to establish that $\hat{\sigma}_t^2$ in (3.8) was a uniformly consistent estimator for σ_t^2. However, we do not want to impose continuity on $\sigma(s)$ and we will show below that even without doing so the resulting feasible weighted break fraction estimator will have the same large sample properties as the infeasible estimator under the conditions stated above for the kernel function and bandwidth.

Based on the adaptive estimate $\hat{\sigma}_t^2$ we can define the corresponding feasible WLS estimator

$$\hat{\tau}_{FWLS} := \arg \min_{\tau \in [\tau_L, \tau_U]} \sum_{t=1}^{T} \tilde{e}_{\tau,t}^{*2}$$

where $\tilde{e}_{\tau,t}^*$, $t = 1, ..., T$, are the OLS residuals from the weighted regression

$$\frac{y_t}{\hat{\sigma}_t} = \tilde{\mu}_\tau \frac{1}{\hat{\sigma}_t} + \tilde{\delta}_\tau \left(1_{t > \lfloor \tau T \rfloor} \cdot \frac{1}{\hat{\sigma}_t}\right) + \tilde{e}_{\tau,t}^*. \tag{3.9}$$

We now detail the large sample properties of the feasible WLS estimator, $\hat{\tau}_{FWLS}$. As in Xu and Phillips (2008), in order to do so we need to assume conditional homoskedasticity in ε_t and appropriately strengthen the moment condition in part (ii) of Assumption $\mathcal{A}_1$.

Theorem 3. *Let the conditions of Theorem 1 hold with $d \geq 0$. Assume further that $E(\varepsilon_t^2 | \mathcal{F}_{t-1}) = 1$ and that Assumption $\mathcal{A}_1$(ii) is replaced by $\sup_t E(\varepsilon_t^8) < \infty$. If the kernel function $K(\cdot)$ and bandwidth h satisfy the conditions stated below equation* (3.8)*, then $\hat{\tau}_{FWLS} - \hat{\tau}_{WLS} \xrightarrow{p} 0$.*

Remark 3.6. The result in Theorem 3 demonstrates that the feasible WLS level break estimator, $\hat{\tau}_{FWLS}$, based on the adaptive estimation of σ_t is asymptotically equivalent to the infeasible WLS estimator $\hat{\tau}_{WLS}$. □

Remark 3.7. It is straightforward to show that the adaptive estimator for σ_t remains consistent (except, as usual, at the points of discontinuity of $\sigma(s)$) in the presence of serial correlation in e_t of the form mentioned in Remark 3.2. The result in Theorem 3 will continue to hold in such cases. Boswijk and Zu (2018) also discuss the kernel estimation of variances in the presence of autocorrelation in a related unit root testing context. □

Remark 3.8. Following equation (18) of Bai (1997, p.555), it is possible to use the result in (3.3) of Theorem 1 to construct confidence intervals for the true break fraction, τ_0, based on either $\hat{\tau}_{OLS}$ or $\hat{\tau}_{FWLS}$. For a generic break fraction estimator $\hat{\tau}$ equal to either $\hat{\tau}_{OLS}$ or $\hat{\tau}_{FWLS}$, it is straightforward to show that

$$\widehat{\underline{\sigma}}^2_{\lfloor T\hat{\tau}\rfloor} := \frac{\sum_{i=1}^{\lfloor T\hat{\tau}\rfloor} K_h\left(\frac{\lfloor T\hat{\tau}\rfloor - i}{T}\right)\hat{e}^2_{\hat{\tau},i}}{\sum_{i=1}^{\lfloor T\hat{\tau}\rfloor} K_h\left(\frac{\lfloor T\hat{\tau}\rfloor - i}{T}\right)} \quad \text{and} \quad \widehat{\bar{\sigma}}^2_{\lfloor T\hat{\tau}\rfloor+1} := \frac{\sum_{i=\lfloor T\hat{\tau}\rfloor+1}^{T} K_h\left(\frac{\lfloor T\hat{\tau}\rfloor+1-i}{T}\right)\hat{e}^2_{\hat{\tau},i}}{\sum_{i=\lfloor T\hat{\tau}\rfloor+1}^{T} K_h\left(\frac{\lfloor T\hat{\tau}\rfloor+1-i}{T}\right)}$$

are consistent estimates of $\bar{\sigma}(\tau_0)^2$ and $\underline{\sigma}(\tau_0)^2$, respectively. The parameters ϕ and ξ can then be estimated using a standard plug-in method. Letting c_1 and c_2 respectively denote the $(\alpha/2)^{\text{th}}$ and $(1-\alpha/2)^{\text{th}}$ quantiles of $\arg\max_s Z(s)$, computed from equations B.2 and B.3 of Bai (1997, p.563), an approximate $100(1-\alpha)\,\%$ confidence interval for τ_0 can then be constructed as

$$\left[\hat{\tau} - c_2\left(\frac{T\hat{\delta}^2_{\hat{\tau}}}{\widehat{\underline{\sigma}}^2_{\lfloor T\hat{\tau}\rfloor}}\right)^{-1} - \frac{1}{T}\,,\ \hat{\tau} - c_1\left(\frac{T\hat{\delta}^2_{\hat{\tau}}}{\widehat{\underline{\sigma}}^2_{\lfloor T\hat{\tau}\rfloor}}\right)^{-1} + \frac{1}{T}\right]$$

with $\hat{\delta}_{\hat{\tau}}$ obtained from (3.2) with $x_t = 1$ for $\hat{\tau}_{OLS}$ or $x_t = 1/\hat{\sigma}_t$ for $\hat{\tau}_{FWLS}$.

3.5 Finite Sample Properties

We now provide a Monte Carlo comparison of the finite sample behaviour of the OLS and feasible WLS break fraction estimators, $\hat{\tau}_{OLS}$ and $\hat{\tau}_{FWLS}$ respectively, from section 3 under both homoskedasticity and a variety of heteroskedastic environments. We also explore how useful the large sample results from the previous section are in predicting their finite sample behaviour.

All simulation results are based on 10,000 Monte Carlo replications programmed in Gauss 15 using the `rndn` random number generator. For both $\hat{\tau}_{OLS}$ and $\hat{\tau}_{FWLS}$ we set $\tau_L = 0.2$ and $\tau_U = 0.8$ in (3.1), thereby defining the set of possible breakpoints to be searched over as $\{T/5, ..., 4T/5\}$. For the kernel variance estimator for $\hat{\tau}_{FWLS}$ we used a QS kernel and plug-in bandwidth $h = sT^{-0.2}$ where s is the sample standard deviation of the regressor $1, \ldots, T$ (see section 2.2.1 of Li and Racine, 2007); the results were found to be quite insensitive to reasonable variations of this choice.

The Monte Carlo simulations reported in this section are based on the level break DGP:

$$y_t = \mu + \delta \cdot 1_{t>\lfloor T\tau_0\rfloor} + \sigma_t\varepsilon_t,\ t = 1, ..., T, \ \text{ with }\ \varepsilon_t \sim \text{i.i.d.}N(0,1). \tag{3.10}$$

Data were generated from this DGP allowing for both the no break case, $\delta = 0$, and for level breaks occurring at $\tau_0 \in \{0.3, 0.5, 0.7\}$. The volatility process, σ_t, was varied among the following models:

$$\begin{aligned}
\text{SD0} : \sigma_t &= 1,\ t = 1, ..., T \\
\text{SD1} : \sigma_t &= 1 + \kappa \cdot 1_{t>\lfloor T\lambda_0\rfloor},\ \ \text{SD2} : \sigma_t = 1 + \kappa \cdot 1_{t<\lfloor T\lambda_0\rfloor},\ \text{with } \kappa \in \{1,2\} \text{ and } \lambda_0 \in \{0.3, 0.5, 0.7\} \\
\text{SD3} : \sigma_t &= 1 + \kappa \cdot (1_{t<\lfloor T\lambda_0\rfloor} + 1_{t>\lfloor T(1-\lambda_0)\rfloor}),\ \text{with } \lambda_0 = 0.3 \text{ and } \kappa \in \{1,2\} \\
\text{SD4} : \sigma_t &= 1 + \kappa \cdot t/T,\ \text{with } \kappa \in \{1,2\}.
\end{aligned}$$

SD0 is the case of unconditional homoskedasticity. SD1 (SD2) allows for an increase (decrease) in volatility at break fraction λ_0 from 1 to $(1+\kappa)$ ($(1+\kappa)$ to 1). SD3 allows for a double change in volatility from $(1+\kappa)$ to 1 at break fraction λ_0 reverting back to $(1+\kappa)$ at $(1-\lambda_0)$. Finally SD4 generates a volatility process which follows a positive linear trend between 1 at the start of the sample and $(1+\kappa)$ at the end of the sample.

Tables 1–4 report the mean and standard deviation and, when $\delta \neq 0$, the root mean squared error [RMSE] from the simulated distributions of $\hat{\tau}_{OLS}$ (Panel A) and $\hat{\tau}_{FWLS}$ (Panel B) for samples of size 100 and 300 and for level break magnitudes $\delta \in \{0, 0.5, 1\}$. Figures 2–5 report corresponding plots of the empirical density functions of $\hat{\tau}_{OLS}$ and $\hat{\tau}_{FWLS}$ for samples of size $100, 200$ and 300 and break magnitudes $\delta \in \{0, 0.5\}$, organised so that Figure 2 presents results for the no level break case, while Figures 3, 4 and 5 present results for the case where a level break occurs at $\tau_0 = 0.3$, 0.5 and 0.7, respectively. A brief summary of the main conclusions is as follows.

(i) The efficacy of both $\hat{\tau}_{OLS}$ and $\hat{\tau}_{FWLS}$ in estimating τ_0 improves with larger sample sizes (and/or larger break magnitudes), illustrating the consistency property from Theorem 1.

(ii) There is a tendency for $\hat{\tau}_{OLS}$ to be drawn towards periods of high volatility in a time series, regardless of the presence and location of a level break, which can produce substantial finite sample bias in the estimator if the level break does not occur in such periods.

(iii) This tendency can be counteracted by using the weighted estimator $\hat{\tau}_{FWLS}$, which down-weights the data in periods of high volatility, and hence substantially reduces the finite sample bias of $\hat{\tau}_{OLS}$ in the worst cases.

(iv) These latter two findings are not predicted by the asymptotic approximation of Theorem 1, but can be reasonably well explained by the results in Theorem 2.

We now discuss the results and conclusions in more detail.

Finite sample properties: level break not present

Consider the results in Table 1 and Figure 2 where no level break occurs, $\delta = 0$. Here we see that for the homoskedastic case $\hat{\tau}_{OLS}$ and $\hat{\tau}_{FWLS}$ behave almost identically with a relatively uniform empirical density across the search interval with slight pile-up effects at the ends of the search set, $\tau_L = 0.2$ and $\tau_U = 0.8$. Both have an empirical mean of about 0.5.

When heteroskedasticity is present the two estimators behave quite differently. While the behaviour of $\hat{\tau}_{FWLS}$ is seen to be relatively unchanged from the homoskedastic case in all of the heteroskedastic cases considered, the behaviour of $\hat{\tau}_{OLS}$ varies considerably across the different non-constant volatility cases. In particular we see that the mass of the distribution of the estimator is redistributed towards high volatility periods *vis-à-vis* the homoskedastic case. This phenomenon is most obviously seen in Figure 2(g) which relates to the case where the volatility increases by a factor of 3 at $\lambda_0 = 0.7$. Here we see that a large bulk of the mass of the empirical density of $\hat{\tau}_{OLS}$ is now spread out across the high volatility period in the data, with the empirical mean of $\hat{\tau}_{OLS}$ now very close to 0.8, the upper limit of the search set. In contrast, the empirical density of $\hat{\tau}_{FWLS}$ in Figure 2(h) is seen to be almost unchanged from the homoskedastic case. This is of course to be expected as, by construction, $\hat{\tau}_{FWLS}$ down-weights the data in periods of high volatility, thereby reducing the tendency of the break estimator to be drawn towards such periods.

Finite sample properties: level break present

When a level break occurs ($\delta \neq 0$), the tendency of $\hat{\tau}_{OLS}$ to be drawn towards high volatility periods in the data persists. Substantial bias can result, especially if the level break occurs during a low volatility period. The weighting inherent in $\hat{\tau}_{FWLS}$ can ameliorate this bias. To illustrate, consider Figures 3e and 3f relative to Figures 3a and 3b — in each case a level break of magnitude $\delta = 0.5$ occurs at $\tau_0 = 0.3$. In Figures 3a and 3b, where volatility is constant, both $\hat{\tau}_{OLS}$ and $\hat{\tau}_{FWLS}$ are centred on τ_0 with the estimated densities becoming increasingly concentrated around τ_0 as the sample size increases. However, in Figures 3e and 3f where the volatility increases threefold at $\lambda_0 = 0.5$, although the density of $\hat{\tau}_{FWLS}$ is almost identical to that seen in Figure 3b, the density of $\hat{\tau}_{OLS}$ is radically altered. A relative peak still exists at τ_0, at least for the larger sample sizes, but it can be observed that, as also happens when no level break is present (see Figure 2e), a large mass of the density has shifted into the high volatility region with a relative peak seen at $\tau_U = 0.8$. Notice also that the performance of the $\hat{\tau}_{OLS}$ estimator is little improved between $T = 100$ and $T = 300$ here. Further illustration of these effects can also be seen from the associated results in Table 2, where the empirical mean of $\hat{\tau}_{OLS}$ is seen to be as high as 0.678 (relative to $\tau_0 = 0.3$) for $T = 100$, an example of the substantial bias referred to above.

The results also show that the weighted estimator is not a panacea and can in some cases display apparently inferior finite sample performance to $\hat{\tau}_{OLS}$. This can occur in cases where the level break occurs in a high volatility period of the data, and especially so where the period of high volatility is short-lived. Where the level break occurs within an extended period of high volatility, weighting is relatively innocuous and there is little difference seen between $\hat{\tau}_{OLS}$ and $\hat{\tau}_{FWLS}$. This phenomenon occurs because here, as we have already observed, some of the mass of the unweighted $\hat{\tau}_{OLS}$ estimator is attracted to the high volatility regime, regardless of whether a level break occurs or not. In contrast, $\hat{\tau}_{FWLS}$ down-weights the high volatility period and, as a result, where a level break occurs within the high volatility regime $\hat{\tau}_{FWLS}$ will have less mass in the vicinity of the level break than the $\hat{\tau}_{OLS}$ estimator. However, for $\hat{\tau}_{OLS}$ this mass will be spread across the high volatility regime and so one will still see reduced performance relative to the homoskedastic case (even where the level and volatility break locations coincide) and increasingly so the longer the duration of the high volatility period. A good illustration of this phenomenon is seen in Figures 5a-5h relating to the case where a level break occurs at $\tau_0 = 0.7$. In the homoskedastic case, $\hat{\tau}_{OLS}$ and $\hat{\tau}_{FWLS}$ perform similarly well. However, in cases where the volatility increases by a factor 3 at λ_0 we see that the performance of both estimators deteriorates. For $\hat{\tau}_{FWLS}$ the performance is roughly similar regardless of where in the sample the volatility break occurs. For $\hat{\tau}_{OLS}$ the pile up of mass in the high volatility region is evident (see also Figures 2c, 2e and 2f) and so it has more mass in the vicinity of the level break - increasingly so as λ_0 increases, such that the duration of the high volatility region decreases. Indeed, for the case of the longest period of high volatility where this regime starts at $\lambda_0 = 0.3$ the empirical densities of $\hat{\tau}_{OLS}$ and $\hat{\tau}_{FWLS}$ are relatively similar.

Finite sample properties and the asymptotic approximations

We can also use the results in Figures 2-5 and Tables 1-4 to explore further how well the finite sample behaviour of $\hat{\tau}_{OLS}$ and $\hat{\tau}_{FWLS}$ conform to the predictions of the large sample theory given in Theorem 1 for level breaks of fixed magnitude and Theorem 2 for level breaks whose magnitude is local-to-zero at the Pitman rate, $d = 1/2$. Recall that Theorem 1 predicts that both $\hat{\tau}_{OLS}$ and $\hat{\tau}_{FWLS}$ will be consistent for τ_0 regardless of the pattern of heteroskedasticity present. Looking at the results for the constant volatility case in Table 2 and Figures 3-5 we see this prediction being borne out for both $\hat{\tau}_{OLS}$ and $\hat{\tau}_{FWLS}$ with each of the empirical bias, standard deviation and RMSE of the estimators decreasing, other things equal, the larger the sample size, T, for a fixed break magnitude, δ. These quantities also all decrease as the break magnitude increases while keeping T constant, as anticipated by the result in Theorem 2 when $d = 1/2$.

A key prediction from Theorem 2 is that for a level break whose magnitude is modelled as local-to-zero at the Pitman rate, the asymptotic distributions of $\hat{\tau}_{OLS}$ and $\hat{\tau}_{FWLS}$ will differ from one another, and that their form will depend on the pattern of unconditional heteroskedasticity present. In contrast, Theorem 1 provides an asymptotic approximation based on a "large" break magnitude, and this predicts that the two estimators will be identically behaved and that it is only the volatility in the neighbourhood of the level break that matters for the efficacy of the estimators. That the finite sample behaviour of $\hat{\tau}_{OLS}$ and $\hat{\tau}_{FWLS}$ differs significantly, and also varies according to the form of heteroskedasticity, has been discussed in some detail above, and this clearly demonstrates the superiority of the asymptotic approximation provided by Theorem 2. An implication of this is that Theorem 1 would be practically unsound as a basis for any further research on formal inference for break fractions, such as the confidence interval construction described in Remark 3.8, in the presence of heteroskedasticity. Theorem 2 would be superior in its finite sample relevance, but poses the challenging question of addressing its complicated nuisance parameter dependency.

Theorem 2 also predicts that the efficacy of the two estimators will depend on the break magnitude, δ, considered relative to the parameter ω. We recall from the discussion in section 3.3 that ω provides a measure of the average volatility in the weighted data and is a function of the volatility path $\sigma(\cdot)$ and of the weighting function used (and therefore differs between $\hat{\tau}_{OLS}$ and $\hat{\tau}_{FWLS}$). To illustrate the role of ω, consider Figures 3m-3p together with Table 4, which relate to the case where a level break occurs at $\tau_0 = 0.3$ and the volatility displays an upward linear trend through the sample (SD4). We can see that relative to the homoskedastic case (see Figures 3a and 3b and Table 1) the efficacy of both $\hat{\tau}_{OLS}$ and $\hat{\tau}_{FWLS}$ is considerably reduced when a trend in volatility is present, and increasingly so as the magnitude of the linear trend, κ, is increased. It is also seen that the peaks in the empirical densities at τ_0 are somewhat smaller for $\hat{\tau}_{OLS}$ than for $\hat{\tau}_{FWLS}$. Noting that ω increases as the magnitude of the linear trend increases and is higher for $\hat{\tau}_{OLS}$ than for $\hat{\tau}_{FWLS}$[3] and that the level break occurs near the start of the series (where the volatility at that point is relatively small compared to the average volatility), we clearly see that the efficacy of the estimators in finite samples is related to the average volatility across the whole sample rather than just to the volatility level near the level break, and to the weighting function used in constructing the level break fraction estimator, in each case as Theorem 2 predicts.

To illustrate further the usefulness of the asymptotic approximation provided by Theorem 2,

[3] In this example the parameter $\omega^2 = 1$ when $\kappa = 0$ (the homoskedastic case) for both $\hat{\tau}_{OLS}$ and $\hat{\tau}_{(F)WLS}$, but for $\hat{\tau}_{OLS}$, $\omega^2 = 2\frac{1}{3}$ when $\kappa = 1$ and $\omega^2 = 4\frac{1}{3}$ when $\kappa = 2$, while for $\hat{\tau}_{(F)WLS}$, $\omega^2 = 2$ when $\kappa = 1$ and $\omega^2 = 3$ when $\kappa = 2$.

Figure 6 graphs simulations of the distribution of $Q(\tau; x(\cdot), \sigma(\cdot), \delta, d)$ with comparisons to the finite sample distributions of $\hat{\tau}_{OLS}$ and $\hat{\tau}_{FWLS}$ from the same DGPs. Figure 6a shows, in the broken lines, the simulated sampling distributions of $\hat{\tau}_{OLS}$ for $T = 100, 200, 300$ from a DGP with no level shift ($\delta = 0$) and heteroskedasticity of the form SD2 with $\kappa = 2$ and $\lambda_0 = 0.7$. The solid line shows the asymptotic approximation for this same DGP, obtained using a 2000 step discretisation. Clearly in this case the distribution of $\hat{\tau}_{OLS}$ is seen to be essentially the same across these sample sizes. Figure 6b shows the same information for $\hat{\tau}_{FWLS}$. The asymptotic approximation remains very accurate here, other than a minor divergence around the time of the break in variance ($\lambda_0 = 0.7$) arising from the differences of the finite sample properties of the kernel variance estimator used for finite T and the true variance process that is used in $Q(\tau; x(\cdot), \sigma(\cdot), \delta, d)$. These two figures illustrate the applicability of the stochastic component of $Q(\tau; x(\cdot), \sigma(\cdot), \delta, d)$ for predicting the finite sample behaviour of the estimators when no level shift occurs.

Figures 6c and 6d graph the simulated finite sample and asymptotic distributions when a level shift of magnitude $\delta_T = \delta T^{-1/2}$ at $\tau_0 = 0.5$ is present. Both figures show that the approximation provided by the asymptotic distribution given in Theorem 2 is very accurate where both a level shift and unconditional heteroskedasticity are present in the DGP. The level shift magnitude in the previous simulations was held fixed, while here it becomes smaller as T increases. Figure 4g and 4h show the finite sample distributions with fixed level shift magnitude of 0.5, and the asymptotic approximations given in Figures 6c and 6d evidently match well with this for $T = 300$ in particular, since for $T = 300$ the implied level shift magnitude $\delta_T = 8T^{-1/2} = 0.46$ is close to 0.5.

4 An Application to the Unit Root Testing Problem

As we have shown, non-stationary volatility can affect the asymptotic and finite sample properties of the OLS and (feasible) WLS estimators of a level break location. However, such estimation is rarely the ultimate goal of the analysis of the data; rather, it is an input into subsequent inference. We now illustrate the relevance of these findings for the case where the estimated level break is used to date a possible trend break in a time series prior to running a unit root test.

4.1 Unit Root Tests allowing for a Possible Trend Break

Consider the time series process y_t generated according to the following DGP,

$$y_t = \begin{cases} \mu_{0,0} + \mu_{1,0} t + z_t, & t = 1, \ldots, \lfloor \tau_0 T \rfloor \\ \mu_{0,1} + \mu_{1,1} t + z_t, & t = \lfloor \tau_0 T \rfloor + 1, \ldots, T \end{cases} \tag{4.1}$$

where

$$z_t = \phi_T z_{t-1} + e_t, \tag{4.2}$$

and where e_t is generated according to (2.2) and is taken to satisfy the conditions of Assumption $\mathcal{A}$.[4] As is common in this literature, we assume that the initial condition satisfies $T^{-1/2} z_0 \xrightarrow{p} 0$. In (4.2)

[4] For simplicity we assume that e_t is serially uncorrelated. Where e_t admits serial correlation of the form given in Remark 3.2, provided the standard invertibility condition that $C(z) \neq 0$ for all $|z| \leq 1$ holds, this can be accounted for in the usual way using an augmented DF statistic, whereby the right hand side of (4.5) is augmented with the lagged differences, $\{\Delta \hat{z}_{\tau,t-j}\}_{j=1}^{p}$, with p satisfying the rate condition that $1/p + p^3/T \to 0$, as $T \to \infty$.

we will follow the convention in the unit root testing literature and focus on the near-integrated autoregressive model, $H_c : \phi_T := 1 + c/T$ with $-\infty < c \leq 0$. We will therefore be concerned with testing the unit root null hypothesis, $H_0 : c = 0$, against local alternatives, H_c where $c < 0$.

The observation equation in (4.1) allows for a linear trend in y_t and a possible break in both intercept and slope occurring at time $\lfloor \tau_0 T \rfloor$. Following Harris *et al.* (2009) and Cavaliere *et al.* (2011), among others, we will focus on the situation where the trend function is restricted to be continuous at the break point, so that the coefficients satisfy $\mu_{0,0} + \mu_{1,0}\lfloor \tau_0 T \rfloor = \mu_{0,1} + \mu_{1,1}\lfloor \tau_0 T \rfloor$. In this case the trend specification can be written as[5]

$$y_t = \alpha + \mu t + \delta_T 1_{t>\lfloor \tau_0 T \rfloor}(t - \lfloor \tau_0 T \rfloor) + z_t \tag{4.3}$$

with $\alpha := \mu_{0,0}$, $\mu := \mu_{1,0}$ and $\delta_T := \mu_{1,1} - \mu_{1,0}$ (allowing for the magnitude of the break to depend on T as the previous sections). Taking first differences we obtain

$$\Delta y_t = \mu + \delta_T 1_{t>\lfloor \tau_0 T \rfloor} + \Delta z_t, \tag{4.4}$$

where $\Delta := (1 - L)$ denotes the first difference operator. Under the unit root null hypothesis, H_0, (4.4) can be seen to coincide with (2.1) on replacing y_t by Δy_t in the latter. Consequently, the results obtained in section 3 relating to the estimation of the level break location continue to apply in this context, so that we estimate the trend break location via level break estimation applied to the first differences of the data.

We will base our unit root test on Dickey-Fuller [DF] type statistics which model the trend break. These statistics are based on a two step procedure whereby the data are de-trended in the first step and in the second step a standard DF test is applied to the de-trended data. We will follow the recent literature and use the quasi-difference [QD] de-trending approach of Elliott *et al.* (1996) in what follows, although OLS de-trending could alternatively be used. For a generic trend break location, τ, the QD de-trended data are given by $\hat{z}_{\tau,t} := y_t - X_t(\tau)' \hat{\theta}_{\bar{c}}$, where $X_t(\tau) := \left(1, t, (t - \lfloor T\tau \rfloor) \cdot 1_{t>\lfloor T\tau \rfloor}\right)'$ and $\hat{\theta}_{\bar{c}}$ the vector of OLS parameter estimates from the regression of $y_{\bar{c},t}$ on $X_{\bar{c},t}(\tau)$, with $y_{\bar{c},1} := y_1$, $y_{\bar{c},t} := y_t - \bar{\phi}_T y_{t-1}$, $t = 2, ..., T$; $X_{\bar{c},1}(\tau) := X_1(\tau)$, $X_{\bar{c},t}(\tau) := X_t(\tau) - \bar{\phi}_T X_{t-1}(\tau)$, $t = 2, \ldots, T$, and where $\bar{\phi}_T := 1 + \bar{c}/T$, where $\bar{c}$ is the QD parameter. The QD de-trended data $\hat{z}_{\tau,t}$ can then be used to estimate the DF regression

$$\hat{z}_{\tau,t} = \hat{\phi}_\tau \hat{z}_{\tau,t-1} + \hat{e}_{\tau,t} \tag{4.5}$$

and hence to obtain the usual DF t-statistic

$$t_\tau := \frac{\hat{\phi}_\tau - 1}{\text{s.e}(\hat{\phi}_\tau)}. \tag{4.6}$$

DF unit root tests can then be based on (4.6) evaluated at either the OLS break fraction estimate, $\tau = \hat{\tau}_{OLS}$, or the corresponding WLS estimate, $\tau = \hat{\tau}_{FWLS}$. We will denote the resulting ADF tests by the simplified notation t_{OLS} and t_{FWLS} in what follows. We will also consider the DF test that

[5]The imposition of continuity on the trend function makes the connection to the level shift results clear and simple. The restriction is not compulsory, however, as without it the equation corresponding to (4.4) would be given by $\Delta y_t = \mu + \lambda 1_{t=\lfloor \tau_0 T \rfloor} + \gamma 1_{t>\lfloor \tau_0 T \rfloor} + \Delta z_t$, and the effect of the additional impulse dummy variable $1_{t=\lfloor \tau_0 T \rfloor}$ is asymptotically negligible.

obtains when allowing only for a constant and linear trend in the QD de-trending, by replacing $X_t(\tau)$ with $X_t := (1, t)'$ in the de-trending step; this statistic will be denoted t_0 in what follows.

Theorem S.1 in the supplementary appendix derives the limiting distribution of t_τ under the local alternative H_c when evaluated at the true break fraction $\tau = \tau_0$, and shows that for the case of a "large" magnitude trend break, i.e. such that $\delta_T = \delta T^{-d}$, $0 \leq d < 1/2$, with $\delta \neq 0$, this limit also holds for t_{OLS} and t_{FWLS}.[6] The (common) limiting null distribution of t_{OLS} and t_{FWLS}, depends on the volatility process, $\sigma(\cdot)$. Consequently ADF tests need to be based on either the simulated critical value approach outlined in section 4.2 of Cavaliere and Taylor (2007) or a wild bootstrap approach, the latter outlined for the t_0 statistic in section 4.1 of Cavaliere and Taylor (2008a), and for the trend break case in Algorithm 1 of Cavaliere *et al.* (2011, p.971). Further discussion on the large sample validity of these methods is provided in the supplementary appendix.

In practice it will not be known for sure if a trend break has occurred. Allowing for a non-existent trend break (and, hence, estimating a phantom break date) results in both t_{OLS} and t_{FWLS} converging to limiting distributions whose form depends on the random outcomes of $\hat{\tau}_{OLS}$ and $\hat{\tau}_{FWLS}$, respectively, within the search set $[\tau_L, \tau_U]$. In order to control asymptotic size the tests must be based on the no break asymptotic critical value; using a critical value based on the estimated break fraction leads to over-sized tests when no trend break occurs. This leads to a loss in test power, even asymptotically, both where a break occurs because a conservative critical value is being used, and where a break does not occur because the inclusion of a redundant trend break regressor leads to a considerable power loss relative to the corresponding unit root test that does not allow for trend break; see, for example, the numerical results presented in section 5 of Harris *et al.* (2009) and sections 3.2 and 5 of Cavaliere *et al.* (2011).

In order to overcome these issues a modified version of the usual Schwarz (1978) criterion [SC] can be used to select between the trend break and no trend break versions of the unit root tests. Analogous procedures based on any consistent information criterion, such as Hannan-Quinn [HQ] where $\log(T-1)$ is replaced by $2\log\log(T-1)$ in the SC penalty functions outlined below, could also be used and would have the same large sample properties as the SC-based procedures. To that end, consider calculating the SC for break selection based on the representation for y_t provided by equations (4.2) and (4.3). For the calculation excluding the break, define the OLS residuals $\tilde{e}_{0,t}$ from an OLS regression of y_t on an intercept, trend (t) and y_{t-1}, with associated residual variance $s_0^2 := (T-1)^{-1}\sum_{t=2}^{T}\tilde{e}_{0,t}^2$. The SC for the model excluding the trend break is therefore

$$SC_0 := (T-1)\log(s_0^2) + 3\log(T-1) \tag{4.7}$$

the "3" appearing in the penalty function derives from the estimation of the coefficients on the intercept, trend and y_{t-1} regressors. Similarly the calculation for the model including a trend break at break fraction τ involves the residuals $\tilde{e}_{\tau,t}$ from an OLS regression of y_t on an intercept, trend (t), y_{t-1} and also the break regressors $1_{t>\lfloor\tau T\rfloor}$ and $1_{t>\lfloor\tau T\rfloor}(t-\lfloor\tau T\rfloor)$, giving residual variance

[6]For $d = 1/2$ results comparable to those given in section 5 of Harvey *et al.* (2012), but generalised by the non-stationary volatility allowed for under Assumption $\mathcal{A}_2$, would be obtained. For $d > 1/2$, as discussed in Case 1 in section 3.3, the magnitude of the trend break would be such that it would lead to trend break estimators which behave asymptotically the same as in the no break case.

$s_\tau^2 := (T-1)^{-1}\sum_{t=2}^T \tilde{e}_{\tau,t}^2$, and SC

$$SC_\tau := (T-1)\log(s_\tau^2) + 6\log(T-1). \tag{4.8}$$

The penalty of 6 presumes that the break fraction τ is an estimated parameter, as it will be in our applications.[7] If a fixed τ were used then the penalty would become 5. The SC decision rule is to include a trend break at time $t = \lfloor \tau T \rfloor$ if $SC_\tau < SC_0$, and to exclude the trend break otherwise. We evaluate below an implementation of this decision rule with τ replaced by the OLS estimator $\hat{\tau}_{OLS}$, taking no account for heteroskedasticity.

The evidence of section 3 suggests that $\hat{\tau}_{FWLS}$ can be substantially superior to $\hat{\tau}_{OLS}$ under certain forms of non-stationary volatility, and so we also consider its use in the SC. In addition in this case, since weighting for heteroskedasticity was found to be effective for break point estimation, we also consider its effectiveness for break selection by including weighting in the SC calculation. The weighted residuals $\tilde{e}_{0,t}^*$ are calculated from a regression of $y_t/\hat{\sigma}_t$ on $1/\hat{\sigma}_t$, $t/\hat{\sigma}_t$ and $y_{t-1}/\hat{\sigma}_t$, where $\hat{\sigma}_t^2$ is defined in (3.8). Similarly the residuals $\tilde{e}_{\hat{\tau}_{FWLS},t}^*$ are calculated from a regression of $y_t/\hat{\sigma}_t$ on $1/\hat{\sigma}_t$, $t/\hat{\sigma}_t$, $y_{t-1}/\hat{\sigma}_t$, $1_{t>\lfloor \hat{\tau}_{FWLS} T\rfloor}/\hat{\sigma}_t$ and $1_{t>\lfloor \hat{\tau}_{FWLS} T\rfloor}(t - \lfloor \hat{\tau}_{FWLS} T\rfloor)/\hat{\sigma}_t$. The weighted SC analogues of (4.7) and (4.8) are then given by

$$SC_0^* := (T-1)\log(s_0^{*2}) + 3\log(T-1), \text{ and } SC_{\hat{\tau}_{FWLS}}^* := (T-1)\log(s_{\hat{\tau}_{FWLS}}^{*2}) + 6\log(T-1),$$

respectively, where $s_0^{*2} := (T-1)^{-1}\sum_{t=2}^T \tilde{e}_{0,t}^{*2}$ and $s_{\hat{\tau}_{FWLS}}^{*2} := (T-1)^{-1}\sum_{t=2}^T \tilde{e}_{\hat{\tau}_{FWLS},t}^{*2}$.

We will use the unweighted and weighted SC decision rules outlined above to choose whether or not to include a trend break in the de-trending regression used in the first step of computing the unit root statistics outlined above. Our proposed weighted and unweighted SC-based DF test statistics are then defined as,

$$t_{SC} := \begin{cases} t_0 & \text{if } SC_0 < SC_{\hat{\tau}_{OLS}} \\ t_{OLS} & \text{if } SC_0 \geq SC_{\hat{\tau}_{OLS}} \end{cases} \text{ and } t_{WSC} := \begin{cases} t_0 & \text{if } SC_0^* < SC_{\hat{\tau}_{FWLS}}^* \\ t_{FWLS} & \text{if } SC_0^* \geq SC_{\hat{\tau}_{FWLS}}^* \end{cases} \tag{4.9}$$

respectively, where we recall that t_0 is the DF test that obtains when allowing only for a constant and linear trend in the QD de-trending step.

Theorem S.2 in the supplementary appendix establishes the large sample properties of the weighted and unweighted SC-based procedures, for the case were the trend break magnitude is either zero or "large". These results show that the tests from both SC procedures are asymptotically correctly sized when using the appropriate asymptotic critical value, obtained using either the simulated critical value approach of Cavaliere and Taylor (2007) or a wild bootstrap approach, regardless of whether a trend break occurs or not. Moreover, the asymptotic local power of the SC tests is identical to that of the (size-adjusted) infeasible test which assumes knowledge of whether a break has occurred or not, and knowledge of the true break fraction, τ_0, in the former case.

[7] It is also worth noting that both the unweighted and weighted SC penalties given above assign a penalty of 1 to the unknown breakpoint parameter. Theoretical results provided in Zhang and Siegmund (2007), Kurozumi and Tuvaandorj (2011) and Kim (2012) suggest that a stricter penalty of 2 might be appropriate for this parameter.

4.2 Finite Sample Simulations

We now use Monte Carlo simulation methods to investigate whether the superior finite sample behaviour observed for the feasible weighted break fraction estimator, $\hat{\tau}_{FWLS}$, over the unweighted estimator, $\hat{\tau}_{OLS}$, seen in the simulation results in section 3, carries over to the unit root test procedures based on $\hat{\tau}_{FWLS}$ and the feasible weighted model selection criteria outlined above, relative to unit root tests based on the corresponding unweighted quantities.

The results reported in this section are based on the DGP:

$$y_t = \alpha + \mu t + \delta(t - \lfloor \tau_0 T \rfloor) \cdot 1_{t > \lfloor \tau_0 T \rfloor} + z_t \tag{4.10}$$

$$z_t = \phi_T z_{t-1} + \sigma_t \varepsilon_t, \ \ \varepsilon_t \sim \text{i.i.d.} N(0,1). \tag{4.11}$$

We set $\alpha = \mu = 0$ in our experiments because all of the unit root tests considered are exact invariant to these parameters. For the volatility process, σ_t, we considered the same set of models as outlined in section 3.5. Again we report only a representative selection here with the full set of results available from the authors on request. In particular, Figures 7 and 8 for $T = 100$ and $T = 200$, respectively, report results for the homoskedastic case $\kappa = 0$, and for a one-time break in volatility occurring at $\lfloor \lambda_0 T \rfloor$ for $\lambda_0 \in \{0.3, 0.5, 0.7\}$. Results are reported for the no trend break case, $\delta = 0$, and where a trend break of magnitude $\delta = 0.5$ occurs at $\lfloor \tau_0 T \rfloor$ for $\tau_0 \in \{0.3, 0.5, 0.7\}$.

Figures 7 and 8 compare the empirical rejection frequencies, for $\phi_T := (1 + c/T)$ with $c \in \{0, -1, -2, ..., -50\}$, of the t_{SC} and t_{WSC} SC-based unit root test procedures of (4.9), comparing each with a number of benchmark tests that are also required in the definition of t_{SC} and t_{WSC}. First t_0, the DF test which does not allow for a trend break in the de-trending step and where we used $\bar{c} = -13.5$ in the QD de-trending procedure. Second, in cases where a trend break occurs in the DGP, t_{τ_0} the infeasible DF test based on including a trend break in the de-trending step at the true break fraction τ_0. Finally, we also report t_{OLS} and t_{FWLS}, the DF tests which always including a trend break located at $\hat{\tau}_{OLS}$ and $\hat{\tau}_{FWLS}$, respectively, in the de-trending step. For all of the tests which include a trend break we set $\bar{c}$ in the QD de-trending procedure according to the relevant entry from Table 1 from Cavaliere *et al.* (2011, p.964). In all cases the tests were run at the nominal 5% level using the Gaussian wild bootstrap with 499 bootstrap replications. For the t_{SC} and t_{WSC} procedures the SC rule with the penalties outlined in section 4.1 are used. Also shown under the labels 'SC' and 'WSC', respectively, are the empirical frequencies with which the unweighted and weighted SC decision rules select the model which allows for a trend break.

The finite sample properties of t_{SC} relative to t_{WSC}, and of t_{OLS} relative to t_{FWLS} generally mirror the corresponding differences seen between the unweighted and weighted break fraction estimators, $\hat{\tau}_{OLS}$ and $\hat{\tau}_{FWLS}$, seen in the results for these models for σ_t in section 3. In all of the Figures relating to a trend break, the differences between the weighted and unweighted SC decision rules and tests are generally rather smaller, other things equal, for $T = 200$ than for $T = 100$. This is to be expected, given that both approaches are consistent and δ is fixed and non-zero.

Consider first the homoskedastic cases in Figures 7a, 7e, 7i and 7m and Figures 8a, 8e, 8i and 8m. Here we see no discernible differences between the behaviour of t_{SC} and t_{WSC} and between t_{OLS} and t_{FWLS}, even for $T = 100$. Where no trend break is present (Figures 7a and 8a), both the weighted and unweighted SC decision rules select the no trend break model with high probability and, as a result, both t_{SC} and t_{WSC} lie very close to the (near-) efficient t_0 test. Notice that a degree of over-sizing is seen here for both t_{OLS} and t_{FWLS} and, as a consequence, also for t_{SC} and t_{WSC}, although this is reduced for $T = 200$ *vis-à-vis* $T = 100$. The power gains from using the SC-based t_{SC} and t_{WSC} tests, relative to the t_{OLS} and t_{FWLS} tests which always include a trend break (at the fitted break fractions $\hat{\tau}_{OLS}$ and $\hat{\tau}_{FWLS}$, respectively), when no break occurs can also clearly be seen for both sample sizes. Where a trend break is present (Figures 7e, 7i and 7m and Figures 8e, 8i and 8m) the power of the t_0 test is effectively zero, regardless of the value of c. Consequently, we want the t_{SC} and t_{WSC} procedures to select the no break case, and hence t_0, as infrequently as possible. The results show that both the weighted and unweighted SC rules perform well in this regard, with t_{SC} and t_{WSC} generally lying reasonably close to t_{OLS} and t_{FWLS} respectively, the more so the later in the sample the trend break occurs, which in turn lie close to the infeasible efficient benchmark t_{τ_0} test. An interesting feature seen for both SC decision rules is that their efficacy to select the trend break model improves the further the AR parameter ϕ_T lies into the stationarity region (i.e. the bigger is c). This phenomenon is clearly beneficial to the finite sample performance of the t_{SC} and t_{WSC} procedures, and is to be expected given that it is well known that a trend break is more easily detected in stationary noise than it is in noise which contains a unit root; see, for example, Harvey *et al.* (2009).

Consider next the cases where σ_t is heteroskedastic. Where no trend break occurs ($\delta = 0$), it is seen in Figures 7b, 7c and 7d and Figures 8b, 8c and 8d that although the weighted SC decision rule is marginally more efficacious in selecting the no trend break model than the unweighted SC rule, and increasingly so as λ_0 increases, in selecting the no break model, there is almost nothing to choose between the resulting t_{SC} and t_{WSC} procedures, each of which again performs well lying very close to the t_0 test, as in the homoskedastic case. However, where a trend break occurs ($\delta \neq 0$) this picture changes considerably. The most dramatic differences between the weighted and unweighted tests are seen for precisely those cases where $\hat{\tau}_{FWLS}$ was observed in the simulations in section 3 to be significantly more efficacious than $\hat{\tau}_{OLS}$. These are the cases where the trend break occurs in a low volatility regime and correspond with Figures 7g, 7h and 7l and Figures 8g, 8h and 8l. Here the superior finite sample performance of t_{FWLS} over t_{OLS} is clearly seen with the former lying very close to the infeasible efficient benchmark t_{τ_0} test, while the latter lies some considerable distance from this benchmark. As these tests differ only in the break fraction estimator used, the power improvement of t_{FWLS} over t_{OLS} can be attributed to the superior properties of $\hat{\tau}_{FWLS}$ in these situations. In particular, the results of Theorem 2 and the simulation results of section 3.5 document and explain the tendency of $\hat{\tau}_{OLS}$ to be potentially badly biased when the trend break lies in a low volatility regime. It has been well known since Perron (1989) that not properly accounting for a trend break results in unit root tests with very low power, and that fitting a trend break at the wrong location is essentially no better than not fitting a trend break at all. Similarly, the results of Theorem 2 and the simulation results of section 3.5 document and explain how the weighting used in $\hat{\tau}_{FWLS}$ works to counteract the bias in $\hat{\tau}_{OLS}$ due to the heteroskedasticity.

It is also seen in the examples discussed above that the weighted SC decision rule is considerably more efficacious than the unweighted SC decision rule in (correctly) selecting the trend break model for the de-trending step. This is crucial to explaining the differences in the behaviour of t_{SC} relative to t_{WSC}. Too often in these cases, the unweighted SC rule wrongly selects the no break model and hence selects the inappropriate no break t_0 test and, as such, is heavily compromised. The superior performance of both the weighted SC decision rule and the DF test based on the weighted break fraction estimator translate into very significant power gains for t_{WSC} over t_{SC} in these cases, especially so for $T = 100$. For example, in Figure 7g the empirical power of t_{WSC} for $T = 100$ is around 90% for $c = -40$ while that of t_{SC} is only about 35%. Interestingly, the weighted SC decision rule often outperforms the unweighted SC rule, and t_{WSC} accordingly outperforms t_{SC}, even in cases where $\hat{\tau}_{FWLS}$ was seen to be no more efficacious than $\hat{\tau}_{FWLS}$ in the simulations in section 3. Examples of this can be seen in Figures 7f and 7k and 8f and 8k where the location of the trend and volatility breaks coincides. In these examples t_{FWLS} also performs better than t_{OLS}. The intuition for the advantage of the weighted SC rule over the unweighted one is more traditional than for the break fraction estimators – the SC method is essentially a likelihood ratio criterion for break inclusion, except that a "penalty" term is applied in place of a critical value. The weighted SC is effectively providing superior "power" for break detection, just as would be expected in a standard formal hypothesis test in the presence of heteroskedasticity.

Finally, in those cases where $\hat{\tau}_{FWLS}$ performed least well relative to $\hat{\tau}_{OLS}$, which are the cases where the trend break lies in a high volatility regime (see Figures 7j, 7n and 7o and Figures 8j, 8n and 8o) the unweighted SC decision rule is seen to perform slightly better than the weighted SC rule. In these examples t_{OLS} correspondingly also performs slightly better than t_{FWLS} as does t_{SC} over t_{WSC}. However it is clear the cost of using the weighting methods in these cases is very much smaller than the gains to using them in the preceding cases discussed, so that in general the weighted methods are to be preferred for practice.

In unreported simulations we also explored corresponding procedures based on the HQ information criterion, and procedures using the stricter double penalty on the estimated break fraction; cf. footnote 7. These govern the strength of the penalty (the SC penalty is stricter than the HQ penalty) imposed on including the trend break. The weaker the penalty, the higher the frequency with which the trend break will be retained in the de-trending step, other things equal. As we have seen, the break retention frequency affects the finite sample size and power properties of the resulting unit root tests. We found that the stricter the penalty used the better the finite sample size control of the information criteria based test procedures (so that, for example, using the SC with a double penalty on the trend break reduced the over-sizing in t_{SC} over t_{WSC} relative to that seen in Figures 7 and 8), but came at the expense of lower finite sample power where a trend break is present. However, the qualitative conclusions drawn above regarding the relative finite sample performance of the unweighted and weighted information criteria and associated unit root tests were unaltered between these different possible penalties.

5 An Empirical Illustration to U.S. and U.K. GDP

We next provide an illustration of the methods discussed in this paper with a practical application to data on GDP in the U.S. and the U.K. The inter-related questions concerning whether GDP admits an autoregressive unit root and/or a broken deterministic linear trend date back to at least Perron (1989). We revisit these questions using both standard methods and the corresponding (adaptive) weighted methods proposed here. The U.S. dataset we consider has previously been analysed in Eo and Morley (2015) and constitutes a measure of quarterly real U.S. GDP, obtained from the Bureau of Economic Analysis website. It was downloaded from James Morley's website, `https://sites.google.com/site/jamescmorley/research/code`. The quarterly U.K. GDP dataset, obtained from the IMF Outlook, was downloaded from Benjamin Wong's website, `https://sites.google.com/site/benjaminwongshijie/research` and was previously analysed in Kamber, Morley and Wong (2018). Full details on the construction of the U.S. and U.K. datasets are provided in Eo and Morley (2015) and Kamber *et al.* (2018), respectively. Graphs of the logarithms of the U.S. and U.K. GDP series covering the sample periods considered, namely 1958Q3 to 2012Q1 and 1961Q3 to 2016Q2, respectively, are provided in Figure S.2 in the supplement.

To visualise the possible presence of unconditional heteroskedasticity in these data, part (a) of Figures 9 and 10 plot the annualised quarterly real GDP growth rates for the U.S. and the U.K., respectively. Also plotted are the broken level functions for the growth rate series corresponding to a level break estimated from the growth rate series by either the standard OLS estimator $\hat{\tau}_{OLS}$ (the blue dashed line) or by our proposed FWLS estimator $\hat{\tau}_{FWLS}$ (the dashed red line); for both $\hat{\tau}_{OLS}$ and $\hat{\tau}_{FWLS}$ we set $\tau_L = 0.1$ and $\tau_U = 0.9$ in (3.1). Part (b) of Figures 9 and 10 plot the adaptive estimate $\hat{\sigma}_t^2$ obtained according to (3.8) using exactly the same practical implementation settings as used in the simulations in section 3.5. Part (c) of Figures 9 and 10 plots sample variance profiles of the OLS residuals, denoted $\tilde{\varepsilon}_t$, obtained from the regression of the first differences of the log GDP series onto an intercept and $1_{t>\lfloor \hat{\tau}_{FWLS} T \rfloor}$. The sample variance profiles, see Cavaliere and Taylor (2008b), are plots of $\hat{\eta}\,(u) := (\sum_{t=2}^{T} \tilde{\varepsilon}_t^2)^{-1} \sum_{t=2}^{\lfloor Tu \rfloor} \tilde{\varepsilon}_t^2$ against $u \in [0, 1]$. In large samples, $\hat{\eta}\,(u) \approx (\int_0^1 \sigma^2\,(s)\,ds)^{-1} \int_0^u \sigma^2\,(s)\,ds$, which equals u when the unconditional volatility is constant; that is, when there is no unconditional heteroskedasticity. Consequently, under conditional homoskedasticity or, more generally, under stationary conditional heteroskedasticity, $\hat{\eta}(u)$ should be close to the 45 degree line, and significant deviations of this function from the 45 degree line point to the presence of persistent changes in volatility; in particular, in a period of relatively high (low) volatility in the data the slope of $\hat{\eta}(u)$ will tend to exceed (be less than) 45 degrees. These deviations, along with the corresponding 95% confidence bands[8], are reported in part (d) of Figures 9 and 10. The pattern of a period of relatively high volatility followed by a decline in unconditional volatility associated with the Great Moderation from the mid 1980s onwards, and a subsequent increase in volatility again after the Great Recession, discussed in section 1, is apparent for both the U.S. and U.K. GDP data in Figures 9 and 10.

To formally investigate for the presence of non-constant volatility, we report in Table 5 the $\mathcal{H}_{\mathrm{R}}$, $\mathcal{H}_{\mathrm{KS}}$, $\mathcal{H}_{\mathrm{CvM}}$, and $\mathcal{H}_{\mathrm{AD}}$ stationary volatility tests of Cavaliere and Taylor (2008b, p. 312) applied to $\tilde{\varepsilon}_t$ for both the U.S. and U.K. real GDP series. These are tests of the null of stationary volatility, i.e. allowing in particular for conditional heteroskedasticity under the null, against the alternative of non-stationary volatility (unconditional heteroskedasticity). The results demonstrate that both series display strong statistical evidence of unconditional heteroskedasticity.

[8]The confidence bands are obtained as suggested by Cavaliere and Taylor (2008b). This requires estimation of the long-run variance of $\tilde{\varepsilon}_t^2$ under the null hypothesis, which is done here using an autoregressive spectral density estimator with lag length chosen by a standard SC starting from an initial maximum of 4 lags.

Table 5: Application of the Stationary Volatility Tests
of Cavaliere and Taylor (2008b) to U.S. and U.K. real GDP

	$\mathcal{H}_{\text{R}}$	$\mathcal{H}_{\text{KS}}$	$\mathcal{H}_{\text{CvM}}$	$\mathcal{H}_{\text{AD}}$
U.S.	2.030***	1.860***	0.780***	3.541**
U.K.	2.037***	1.946***	1.187***	5.375***

Note: The superscripts *, **, and *** denote significance at the 10%, 5%, and 1% nominal (asymptotic) levels, respectively.

The OLS level break estimate $\hat{\tau}_{OLS}$ for the U.S. GDP growth rate series (graphed in Figure 9(a)) gives a break date of 1966Q1, while for U.K. GDP growth rates (graphed in Figure 10(a)) $\hat{\tau}_{OLS}$ implies a break date of 1973Q1. In each case a trend break at these dates is therefore implied in the levels GDP series. In both cases these estimated trend breaks lie in a high volatility period of the time series. Moreover, these locations are close to those found in earlier studies in the literature based on OLS break date estimation; for example, Kim and Perron estimate a trend break in U.S. GDP located at 1965Q2 (for a sample period of 1947Q1 to 2004Q2). In contrast, the FWLS estimator, $\hat{\tau}_{FWLS}$, places the trend breaks much later: for the U.S. at 2000Q2, and for the U.K. at 2005Q4, both of which lie in a relatively low volatility phase of the respective GDP series.

In order to investigate the significance of the magnitude of these estimated trend breaks we next use the weighted and unweighted information criteria-based rules from section 4 to select between the trend break and no trend break models for the U.S. and U.K. GDP series. In order to allow for serial correlation of unknown order in the GDP series these criteria were generalised in the obvious way (see Ng and Perron, 2005) to jointly minimise with respect to the autoregressive lag order and between the break and no break models. To that end, in Table 6 we report the outcomes of the unweighted SC-based criteria allowing for the no break and trend break models, SC_0 and $SC_{\hat{\tau}_{OLS}}$, respectively, along with the corresponding weighted criteria, SC_0^* and $SC_{\hat{\tau}_{FWLS}}^*$. We also report the corresponding unweighted and weighted criteria based on the HQ penalty, denoted with an obvious notation by HQ_0, $HQ_{\hat{\tau}_{OLS}}$, HQ_0^* and $HQ_{\hat{\tau}_{FWLS}}^*$. In each case the values reported in Table 6 are the most negative values that each of the criteria takes across all possible autoregressive lag lengths up to a maximum lag length of $p_{\max} = \lfloor 16(\frac{T}{100})^{0.25}) \rfloor$. All of the entries in Table 6 have been scaled by $(T - p_{\max} - 1)$ to aid readability.

Table 6: Standard and Adaptive SC and HQ Information Criteria for U.S. and U.K. real GDP for selecting between trend break and no trend break models.

	SC_0	$SC_{\hat{\tau}_{OLS}}$	SC_0^*	$SC_{\hat{\tau}_{FWLS}}^*$	HQ_0	$HQ_{\hat{\tau}_{OLS}}$	HQ_0^*	$HQ_{\hat{\tau}_{FWLS}}^*$
U.S.	**-9.578**	-9.504	**-12.081**	-12.061	**-9.628**	-9.584	-12.131	**-12.141**
U.K.	**-9.289**	-9.216	**-12.077**	-12.058	**-9.348**	-9.305	-12.126	**-12.137**

We can see from the results in Table 6 that for both the U.S. and the U.K. the SC penalty favours the no break model, regardless of whether the trend break is fitted at the location identified by $\hat{\tau}_{OLS}$ or $\hat{\tau}_{FWLS}$. When the HQ penalty is used there is also no evidence to accept the presence of a trend break at 1966Q1 for the U.S. or at 1973Q1 for the U.K., the dates implied by the $\hat{\tau}_{OLS}$ estimates. However, the weighted IC with the HQ penalty favours the model with a trend break at 2000Q2 for the U.S. and the model with a trend break at 2005Q4 for the U.K., the dates implied by the respective $\hat{\tau}_{FWLS}$ estimates.

Finally, to investigate if the differing estimates of the trend break location have an impact on inference on the unit root hypothesis, we next consider the application of standard unit root tests to the data, allowing for either no trend break, or for a trend break at the locations identified by $\hat{\tau}_{OLS}$ and $\hat{\tau}_{FWLS}$. In Table 7 we report results for the QD detrended augmented DF [ADF] (see footnote 4) tests t_0, t_{OLS} and t_{FWLS} from section 4.1, together with the corresponding ADF tests based on OLS detrending, which we denote by t_0^{ols}, t_{OLS}^{ols} and t_{FWLS}^{ols}, respectively. The autoregressive lag length used in these ADF unit root tests was selected by the usual SC with a maximum lag length of $p_{\max} = \lfloor 16(\frac{T}{100})^{0.25})\rfloor$, and is reported under $\hat{p}$. Wild bootstrap p-values for each test obtained using the algorithms in Cavaliere and Taylor (2008a) and Cavaliere *et al.* (2011), in each case using 499 bootstrap replications, are reported in parentheses.

Table 7: Unit Root Tests for U.S. and U.K. real GDP

	$\hat{p}$	t_0	t_{OLS}	t_{FWLS}	t_0^{ols}	t_{OLS}^{ols}	t_{FWLS}^{ols}
U.S.	2	-0.956 (0.886)	-2.592 (0.249)	-2.128 (0.465)	-1.881 (0.679)	-2.612 (0.325)	-3.094 (0.351)
U.K.	3	-2.209 (0.247)	-2.815 (0.247)	-3.329 (0.038)	-2.913 (0.251)	-2.848 (0.361)	-3.844 (0.046)

For U.S. GDP no evidence is found against the unit root null hypothesis at standard significance levels, regardless of whether we allow for a trend break or not, and regardless of whether the break is placed at the location identified by $\hat{\tau}_{OLS}$ or $\hat{\tau}_{FWLS}$. Hence, although for the U.S. data the HQ-based criterion favours the model with a trend break at 2000Q2, the omission of this trend break from the unit root test procedure does not alter the decision to accept the unit root null hypothesis. In the case of U.K. GDP, when either no trend break is included or a trend is included at 1973Q1 (the date estimated by $\hat{\tau}_{OLS}$) there is again no evidence against the unit root null hypothesis at standard significance levels. However, when a trend break is included at 2005Q4 (the date estimated by $\hat{\tau}_{FWLS}$) both the QD and OLS detrended ADF tests now deliver significant rejections of the unit root null hypothesis at the 5% level with p-values of 0.038 and 0.046, respectively. The evidence therefore suggests that while the magnitudes of the trend breaks in U.S. and U.K. GDP are both sufficiently large for the HQ-based criterion to select the trend break model, it is only in the case of U.K. GDP that this break is of sufficient magnitude that failing to account for it in the unit root test procedure alters the decision made on whether to accept the unit root null hypothesis or not.

6 Conclusions

We have investigated the properties of RSS-based estimators, including OLS and feasible WLS estimators, the latter formed using a non-parametric kernel-based estimate of the volatility process, for the location of a level break in series driven by shocks displaying non-stationary volatility. Consistency rates were derived against breaks of fixed magnitude and shown to coincide with those obtained under homoskedasticity. Distribution theory for these estimators was also derived for cases where the break magnitude was either local-to-zero or exactly zero. Under Pitman drift these limiting distributions were shown to depend on nuisance parameters deriving from the non-stationary volatility and on the location and magnitude of the level break and the bounds of the search set. Monte Carlo evidence demonstrated that these Pitman limits closely predict the finite sample behaviour of both the OLS and feasible WLS estimators, and highlighted the potential for the feasible WLS estimator to deliver significant improvements over the OLS estimator in certain heteroskedastic environments. The feasible WLS level break fraction estimator can be used in the context of the problem of unit root testing when trend and/or volatility breaks may be present in the data by applying it to the first differences of the data. This was shown to have the potential to deliver significant improvements in the finite sample properties of the resulting unit root tests relative to using an OLS break fraction estimate. We also discussed feasible weighted information criteria, based on the same estimate of the volatility process, to select between the trend break and no trend break models. Again these were shown to have the potential to deliver unit root tests with considerably improved finite sample behaviour under heteroskedasticity relative to the use of standard information criteria. An empirical illustration to U.S. and U.K. real GDP highlighted the practical relevance of these methods. For both series, OLS estimation estimated an early break date in a high volatility regime, whereas for both series the feasible WLS estimator estimated a much later break date in a relatively low volatility regime. The positioning of the trend break was shown to be important in the case of the U.K. data, with a rejection of the unit root null hypothesis possible when based on feasible WLS break date, but not when based on the OLS break date.

Although our focus in this paper has been on a single level break, the ideas we have presented naturally extend to the case of multiple level breaks and to structural breaks in the parameters of more general time series regression settings. Moreover, the procedures we develop here should extend to the multivariate case and so would be anticipated to improve inference on determining the co-integration rank in the case of multiple time series potentially subject to breaks in both trend and volatility. These issues are currently being investigated by the authors.

References

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

Bai, J. (1997). Estimation of a change point in multiple regressions. Review of Economics and Statistics 79, 551–653.

Boswijk, H.P. and Y. Zu (2018). Adaptive wild bootstrap testing for a unit root with nonstationary volatility. Econometrics Journal 2, 87–113.

Carrion-i-Silvestre, J.L., D. Kim, and P. Perron (2009). GLS-based unit root tests with multiple structural breaks under both the null and the alternative hypotheses. Econometric Theory 25, 1754—1792.

Cavaliere, G. and A.M.R. Taylor (2007). Testing for unit roots in time series models with non-stationary volatility. Journal of Econometrics 140, 919–947.

Cavaliere, G. and A.M.R. Taylor (2008a). Bootstrap unit root tests for time series with non-stationary volatility. Econometric Theory 24, 43–71.

Cavaliere, G. and A.M.R. Taylor (2008b), Time-transformed unit root tests for models with non-stationary volatility, Journal of Time Series Analysis 29, 300–330.

Cavaliere, G. D.I. Harvey, S.J. Leybourne and A.M.R. Taylor (2011). Testing for unit roots in the presence of a possible break in trend and non-stationary volatility. Econometric Theory 27, 957-991.

Clark, T. (2009). Is the Great Moderation over? An empirical analysis. Federal Reserve Bank of Kansas City Economic Review 94(4), 5–42.

Davidson J. (1994). Stochastic Limit Theory. Oxford: Oxford University Press.

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

Elliott, G., T.J. Rothenberg and J.H. Stock (1996). Efficient tests for an autoregressive unit root. Econometrica 64, 813-836.

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

Hansen, B.E. (1995). Regression with nonstationary volatility. Econometrica 63, 1113–1132.

Harris, D., D.I. Harvey, S.J. Leybourne and A.M.R. Taylor (2009). Testing for a unit root in the presence of a possible break in trend. Econometric Theory 25, 1545—1588.

Harris, D. and H. Kew (2014). Portmanteau Autocorrelation Tests Under Q-Dependence and Heteroskedasticity. Journal of Time Series Analysis 35, 203-217.

Harris, D., and H. Kew (2017). Adaptive Long Memory Testing under Heteroskedasticity. Econometric Theory 33, 755-778.

Harvey, D.I., S.J. Leybourne and A.M.R. Taylor (2009). Simple, robust, and powerful tests of the breaking trend hypothesis. Econometric Theory 25, 995–1029.

Harvey, D.I., S.J. Leybourne and A.M.R. Taylor (2012). Unit root testing under a local break in trend. Journal of Econometrics 167, 140–167.

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

Kamber, G., J. Morley and B. Wong (2018). Intuitive and reliable estimates of the output gap from a Beveridge-Nelson filter. Review of Economics and Statistics 100, 550–566.

Kim, D. and P. Perron (2009). Unit root tests allowing for a break in the trend function at an unknown time under both the null and alternative hypotheses. Journal of Econometrics 148, 1-13.

Kim J-Y. (2012). Model selection in the presence of nonstationarity. Journal of Econometrics 169, 247-257.

Kurozumi, E. and P. Tuvaandorj (2011). Model selection criteria in multivariate models with multiple structural changes. Journal of Econometrics 164, 218–238.

Li, Q. and J.S. Racine (2007). Nonparametric Econometrics: Theory and Practice. Princeton: Princeton University Press.

McConnell, M.M. and G. Perez-Quiros (2000). Output fluctuations in the United States: what has changed since the early 1980s? American Economic Review 90, 1464–1476.

Ng, S. and P. Perron (2005). A note on the selection of time series models. Oxford Bulletin of Economics and Statistics 67, 115–134.

Nunes, L.C., C.M. Kuan, and P. Newbold (1995). Spurious break. Econometric Theory 11, 736–749.

Patilea, V. and H. Raïssi (2012). Adaptive estimation of vector autoregressive models with time varying variance: Application to testing linear causality in mean. Journal of Statistical Planning and Inference 142, 2891–2912.

Patilea, V. and H. Raïssi (2013). Corrected portmanteau tests for VAR models with time-varying variance. Journal of Multivariate Analysis 116, 190–207.

Perron, P. (1989). The Great Crash, the oil price shock, and the unit root hypothesis. Econometrica 57, 1361—401.

Perron, P. (1997). Further evidence of breaking trend functions in macroeconomic variables. Journal of Econometrics 80, 355—385.

Perron, P. and G. Rodríguez (2003). GLS detrending, efficient unit root tests and structural change. Journal of Econometrics 115, 1-27.

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

Phillips, P.C.B. and K.-L. Xu (2006). Inference in Autoregression Under Heteroskedasticity. Journal of Time Series Analysis 27, 289-308.

Schwarz, G. (1978). Estimating the dimension of a model. Annals of Statistics 6, 461–464.

Stock, J.H. and M.W. Watson (1996). Evidence on structural instability in macroeconomic time series relations. Journal of Business and Economic Statistics 14, 11—30.

Stock, J.H. and M.W. Watson (1999). A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. In R.F. Engle and H. White (eds.), Cointegration, Causality and Forecasting: A Festschrift in Honour of Clive W.J. Granger, pp. 1–44. OUP.

Stock, J. and M.W. Watson (2005). Implications of Dynamic Factor Analysis for VAR Models. NBER Working paper 11467.

Stock, J.H. and M.W. Watson (2012), Disentangling the channels of the 2007-09 recession. Brookings Papers on Economic Activity Spring Volume, 81–156.

Xu, K.-L. and P.C.B. Phillips (2008). Adaptive estimation of autoregressive models with time-varying variances. Journal of Econometrics 142, 265–280.

Zhang, N.R. and D.O. Siegmund (2007). A modified Bayes information criterion with applications to the analysis of comparative genomic hybridization data. Biometrics 63, 22–32.

Table 1: Finite Sample Properties of Break Fraction Estimators. No Level Break. Volatility Models SD0-SD4.

T	Mean	SD	Mean	SD	Mean	SD
			Panel A: $\hat{\tau}_{\text{OLS}}$			
			SD1 : $\kappa = 2$			
	$\lambda_0 = 0.3$		$\lambda_0 = 0.5$		$\lambda_0 = 0.7$	
100	0.667	0.187	0.730	0.148	0.782	0.138
300	0.669	0.187	0.733	0.143	0.788	0.130
			SD2 : $\kappa = 2$			
	$\lambda_0 = 0.3$		$\lambda_0 = 0.5$		$\lambda_0 = 0.7$	
100	0.229	0.148	0.277	0.146	0.336	0.182
300	0.215	0.134	0.270	0.146	0.332	0.185
	SD3		**SD4**		**SD0**	
	$\kappa = 2,\ \lambda_0 = 0.3$		$\gamma = 2$			
100	0.359	0.281	0.699	0.201	0.509	0.272
300	0.341	0.276	0.707	0.196	0.502	0.275
			Panel B: $\hat{\tau}_{\text{FWLS}}$			
			SD1 : $\kappa = 2$			
	$\lambda_0 = 0.3$		$\lambda_0 = 0.5$		$\lambda_0 = 0.7$	
100	0.533	0.280	0.515	0.271	0.537	0.261
300	0.490	0.290	0.479	0.269	0.510	0.258
			SD2 : $\kappa = 2$			
	$\lambda_0 = 0.3$		$\lambda_0 = 0.5$		$\lambda_0 = 0.7$	
100	0.478	0.263	0.505	0.270	0.493	0.282
300	0.493	0.258	0.525	0.268	0.519	0.290
	SD3		**SD4**		**SD0**	
	$\kappa = 2,\ \lambda_0 = 0.3$		$\kappa = 2$			
100	0.486	0.261	0.549	0.269	0.510	0.272
300	0.492	0.251	0.523	0.275	0.504	0.275

Table 2: Finite Sample Properties of Break Fraction Estimators. Break size δ, Break fraction τ_0. Volatility Models SD0 and SD1.

		Mean	SD	RMSE	Mean	SD	RMSE	Mean	SD	RMSE
δ	T	$\tau_0 = 0.3$			$\tau_0 = 0.5$			$\tau_0 = 0.7$		
					Panel A: $\hat{\tau}_{\text{OLS}}$					
					SD0					
0.5	100	0.375	0.195	0.209	0.502	0.166	0.166	0.633	0.188	0.200
	300	0.312	0.094	0.094	0.500	0.084	0.084	0.687	0.094	0.095
1	100	0.309	0.074	0.074	0.500	0.064	0.064	0.694	0.067	0.067
	300	0.300	0.019	0.019	0.500	0.019	0.019	0.700	0.020	0.020
					SD1 : $\kappa = 2,\ \lambda_0 = 0.3$					
0.5	100	0.614	0.209	0.378	0.631	0.188	0.229	0.669	0.178	0.181
	300	0.536	0.213	0.318	0.585	0.174	0.193	0.670	0.161	0.164
1	100	0.507	0.210	0.295	0.569	0.166	0.180	0.672	0.153	0.156
	300	0.391	0.144	0.170	0.521	0.115	0.117	0.684	0.114	0.115
					SD1 : $\kappa = 2,\ \lambda_0 = 0.5$					
0.5	100	0.678	0.193	0.425	0.694	0.155	0.248	0.722	0.140	0.142
	300	0.590	0.230	0.370	0.646	0.146	0.207	0.715	0.122	0.123
1	100	0.543	0.235	0.338	0.626	0.141	0.189	0.709	0.118	0.119
	300	0.371	0.164	0.179	0.563	0.097	0.116	0.702	0.086	0.086
					SD1 : $\kappa = 2,\ \lambda_0 = 0.7$					
0.5	100	0.695	0.226	0.455	0.729	0.172	0.286	0.774	0.127	0.147
	300	0.566	0.264	0.375	0.663	0.177	0.241	0.768	0.101	0.122
1	100	0.503	0.257	0.327	0.626	0.171	0.212	0.761	0.095	0.113
	300	0.334	0.130	0.135	0.531	0.098	0.103	0.742	0.061	0.074
					Panel B: $\hat{\tau}_{\text{FWLS}}$					
					SD0					
0.5	100	0.376	0.196	0.210	0.504	0.168	0.168	0.630	0.192	0.204
	300	0.313	0.094	0.095	0.499	0.084	0.084	0.686	0.096	0.097
1	100	0.309	0.075	0.075	0.500	0.064	0.064	0.693	0.069	0.070
	300	0.300	0.019	0.019	0.500	0.019	0.019	0.700	0.020	0.020
					SD1 : $\kappa = 2,\ \lambda_0 = 0.3$					
0.5	100	0.491	0.260	0.323	0.525	0.254	0.255	0.555	0.264	0.302
	300	0.414	0.219	0.247	0.503	0.221	0.221	0.568	0.253	0.285
1	100	0.420	0.205	0.238	0.517	0.196	0.197	0.609	0.222	0.240
	300	0.347	0.109	0.119	0.505	0.124	0.124	0.657	0.159	0.164
					SD1 : $\kappa = 2,\ \lambda_0 = 0.5$					
0.5	100	0.430	0.237	0.270	0.529	0.242	0.244	0.549	0.257	0.298
	300	0.326	0.131	0.134	0.521	0.200	0.201	0.569	0.239	0.273
1	100	0.335	0.140	0.144	0.545	0.179	0.184	0.611	0.214	0.232
	300	0.300	0.023	0.023	0.534	0.097	0.103	0.668	0.140	0.144
					SD1 : $\kappa = 2,\ \lambda_0 = 0.7$					
0.5	100	0.416	0.222	0.250	0.522	0.199	0.200	0.574	0.250	0.280
	300	0.318	0.106	0.108	0.496	0.107	0.107	0.603	0.226	0.246
1	100	0.320	0.105	0.107	0.503	0.101	0.101	0.642	0.204	0.212
	300	0.300	0.019	0.019	0.499	0.021	0.021	0.700	0.117	0.117

Table 3: Finite Sample Properties of Break Fraction Estimators. Break size δ, Break fraction τ_0. Volatility Model SD2.

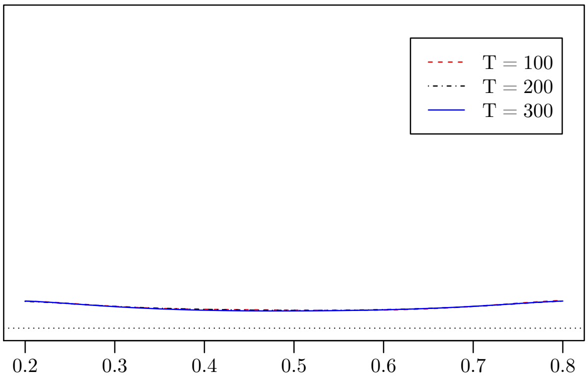
		Mean	SD	RMSE	Mean	SD	RMSE	Mean	SD	RMSE
δ	T	$\tau_0 = 0.3$			$\tau_0 = 0.5$			$\tau_0 = 0.7$		
					Panel A: $\hat{\tau}_{\text{OLS}}$					
					SD2 : $\kappa = 2,\ \lambda_0 = 0.3$					
0.5	100	0.233	0.132	0.148	0.277	0.170	0.281	0.314	0.228	0.448
	300	0.230	0.101	0.123	0.337	0.177	0.241	0.437	0.264	0.372
1	100	0.244	0.104	0.118	0.384	0.168	0.205	0.510	0.252	0.316
	300	0.256	0.062	0.076	0.469	0.098	0.103	0.664	0.134	0.139
					SD2 : $\kappa = 2,\ \lambda_0 = 0.5$					
0.5	100	0.283	0.136	0.137	0.309	0.150	0.243	0.328	0.193	0.419
	300	0.282	0.120	0.122	0.352	0.146	0.208	0.411	0.231	0.369
1	100	0.292	0.115	0.115	0.373	0.139	0.188	0.467	0.234	0.330
	300	0.296	0.087	0.087	0.434	0.098	0.118	0.627	0.168	0.183
					SD2 : $\kappa = 2,\ \lambda_0 = 0.7$					
0.5	100	0.335	0.173	0.177	0.371	0.183	0.224	0.387	0.204	0.374
	300	0.326	0.161	0.163	0.415	0.173	0.193	0.462	0.213	0.319
1	100	0.330	0.152	0.155	0.432	0.162	0.176	0.493	0.206	0.291
	300	0.313	0.113	0.114	0.477	0.114	0.117	0.607	0.146	0.173
					Panel B: $\hat{\tau}_{\text{FWLS}}$					
					SD2 : $\kappa = 2,\ \lambda_0 = 0.3$					
0.5	100	0.440	0.252	0.288	0.489	0.195	0.195	0.592	0.215	0.241
	300	0.399	0.227	0.247	0.503	0.107	0.108	0.680	0.108	0.110
1	100	0.364	0.208	0.218	0.499	0.097	0.097	0.684	0.096	0.097
	300	0.299	0.120	0.120	0.501	0.023	0.023	0.700	0.021	0.021
					SD2 : $\kappa = 2,\ \lambda_0 = 0.5$					
0.5	100	0.468	0.258	0.308	0.487	0.241	0.241	0.586	0.229	0.256
	300	0.432	0.239	0.273	0.479	0.200	0.201	0.671	0.135	0.138
1	100	0.397	0.217	0.238	0.463	0.177	0.181	0.673	0.127	0.130
	300	0.332	0.143	0.146	0.465	0.099	0.105	0.700	0.023	0.023
					SD2 : $\kappa = 2,\ \lambda_0 = 0.7$					
0.5	100	0.463	0.268	0.314	0.490	0.254	0.254	0.526	0.258	0.311
	300	0.429	0.252	0.283	0.500	0.221	0.221	0.584	0.219	0.248
1	100	0.403	0.227	0.249	0.488	0.194	0.194	0.585	0.201	0.232
	300	0.341	0.160	0.165	0.497	0.126	0.126	0.652	0.111	0.121

Table 4: Finite Sample Properties of Break Fraction Estimators. Break size δ, Break fraction τ_0. Volatility Models SD3 and SD4.

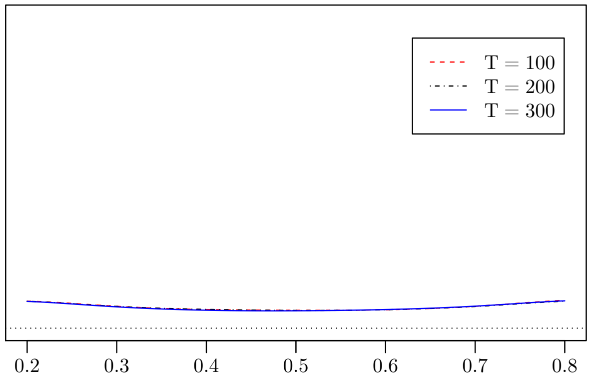
		Mean	SD	RMSE	Mean	SD	RMSE	Mean	SD	RMSE
δ	T	$\tau_0 = 0.3$			$\tau_0 = 0.5$			$\tau_0 = 0.7$		
					Panel A: $\hat{\tau}_{\text{OLS}}$					
					SD3 : $\kappa = 2,\ \lambda_0 = 0.3$					
0.5	100	0.339	0.257	0.260	0.379	0.267	0.293	0.427	0.293	0.401
	300	0.294	0.207	0.207	0.395	0.235	0.257	0.512	0.287	0.343
1	100	0.300	0.198	0.198	0.427	0.214	0.226	0.563	0.266	0.299
	300	0.265	0.093	0.099	0.473	0.112	0.115	0.680	0.150	0.152
					SD4 : $\kappa = 1$					
0.5	100	0.527	0.252	0.339	0.592	0.204	0.224	0.670	0.193	0.195
	300	0.403	0.203	0.227	0.543	0.150	0.156	0.689	0.140	0.140
1	100	0.377	0.178	0.194	0.532	0.130	0.134	0.691	0.121	0.122
	300	0.307	0.053	0.054	0.505	0.052	0.052	0.701	0.057	0.057
					SD4 : $\kappa = 2$					
0.5	100	0.622	0.240	0.402	0.653	0.200	0.252	0.702	0.181	0.181
	300	0.514	0.246	0.326	0.601	0.180	0.206	0.704	0.153	0.153
1	100	0.474	0.235	0.292	0.581	0.168	0.187	0.701	0.143	0.143
	300	0.343	0.132	0.138	0.525	0.097	0.100	0.705	0.088	0.088
					Panel B: $\hat{\tau}_{\text{FWLS}}$					
					SD3 : $\kappa = 2,\ \lambda_0 = 0.3$					
0.5	100	0.449	0.252	0.293	0.493	0.210	0.210	0.547	0.250	0.293
	300	0.402	0.224	0.246	0.501	0.121	0.120	0.633	0.195	0.206
1	100	0.372	0.213	0.225	0.501	0.121	0.121	0.647	0.187	0.194
	300	0.301	0.122	0.122	0.500	0.026	0.026	0.708	0.064	0.064
					SD4 : $\kappa = 1$					
0.5	100	0.443	0.241	0.280	0.529	0.217	0.219	0.596	0.235	0.257
	300	0.347	0.158	0.165	0.509	0.149	0.150	0.642	0.185	0.194
1	100	0.342	0.144	0.150	0.512	0.128	0.128	0.665	0.153	0.157
	300	0.303	0.039	0.039	0.501	0.050	0.050	0.695	0.066	0.066
					SD4 : $\kappa = 2$					
0.5	100	0.478	0.255	0.310	0.539	0.238	0.241	0.576	0.251	0.280
	300	0.374	0.195	0.208	0.514	0.189	0.190	0.601	0.228	0.249
1	100	0.375	0.185	0.200	0.524	0.172	0.173	0.634	0.201	0.212
	300	0.308	0.067	0.067	0.504	0.087	0.087	0.678	0.124	0.126

Figure 2: Simulated Sampling Density Functions of $\hat{\tau}_{OLS}$ and $\hat{\tau}_{FWLS}$. No Level Break.

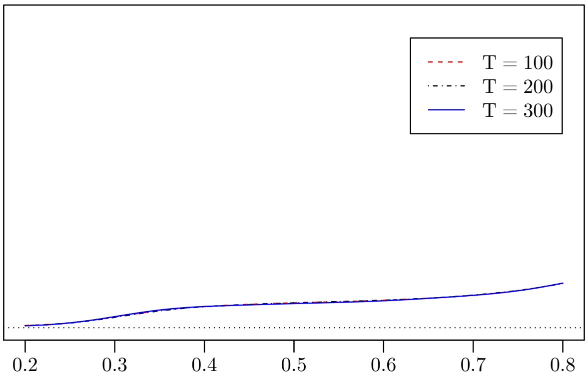
(a) $\hat{\tau}_{\mathrm{OLS}}$: $\delta = 0$, SD0, $\kappa = 0$, $\lambda_0 = 0$



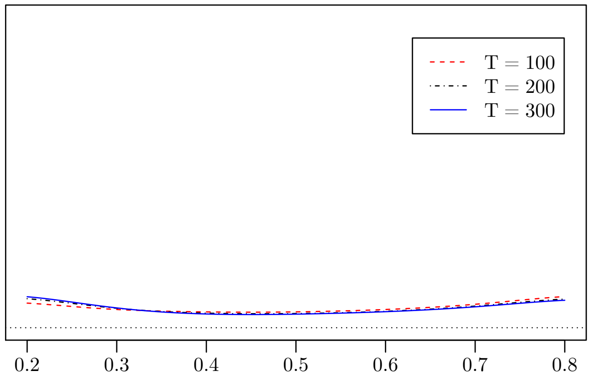
(b) $\hat{\tau}_{\mathrm{FWLS}}$: $\delta = 0$, SD0, $\kappa = 0$, $\lambda_0 = 0$



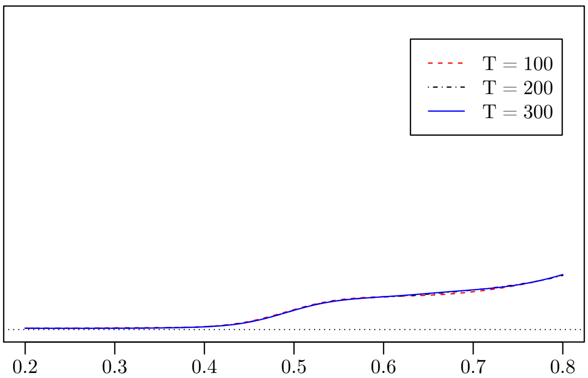
(c) $\hat{\tau}_{\mathrm{OLS}}$: $\delta = 0$, SD1, $\kappa = 2$, $\lambda_0 = 0.3$



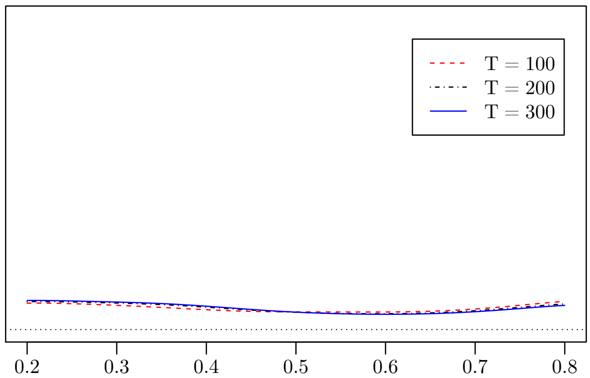
(d) $\hat{\tau}_{\mathrm{FWLS}}$: $\delta = 0$, SD1, $\kappa = 2$, $\lambda_0 = 0.3$



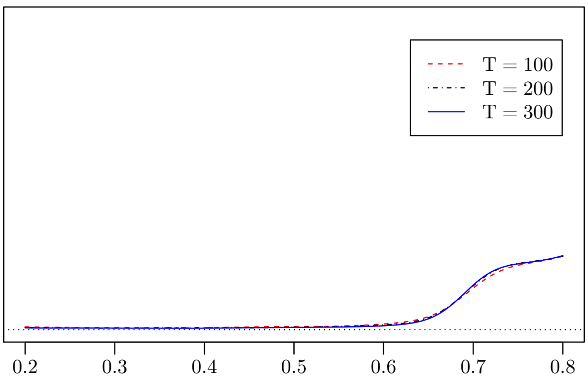
(e) $\hat{\tau}_{\mathrm{OLS}}$: $\delta = 0$, SD1, $\kappa = 2$, $\lambda_0 = 0.5$



(f) $\hat{\tau}_{\mathrm{FWLS}}$: $\delta = 0$, SD1, $\kappa = 2$, $\lambda_0 = 0.5$



(g) $\hat{\tau}_{\mathrm{OLS}}$: $\delta = 0$, SD1, $\kappa = 2$, $\lambda_0 = 0.7$



(h) $\hat{\tau}_{\mathrm{FWLS}}$: $\delta = 0$, SD1, $\kappa = 2$, $\lambda_0 = 0.7$

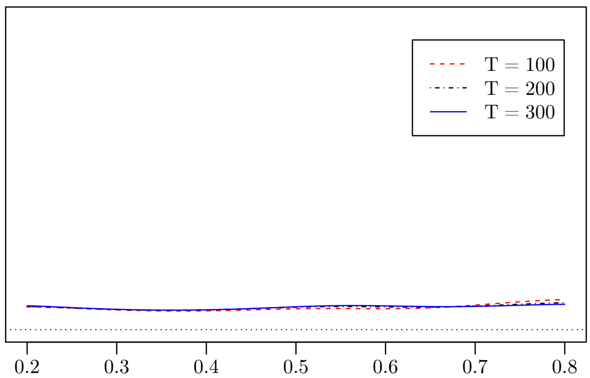
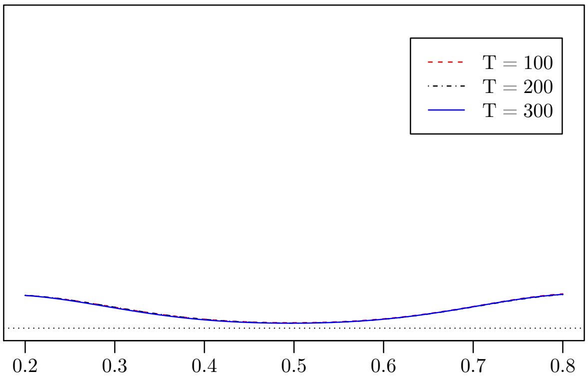
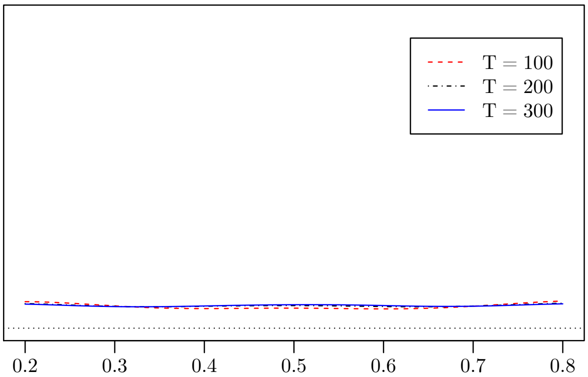


Figure 2: continued ...

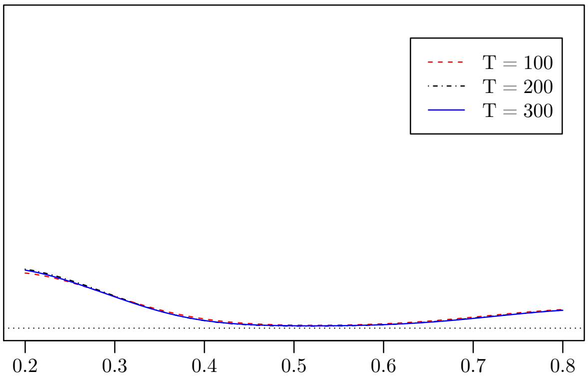
(i) $\hat{\tau}_{\text{OLS}} : \delta = 0$, SD3, $\kappa = 1$, $\lambda_0 = 0.3$



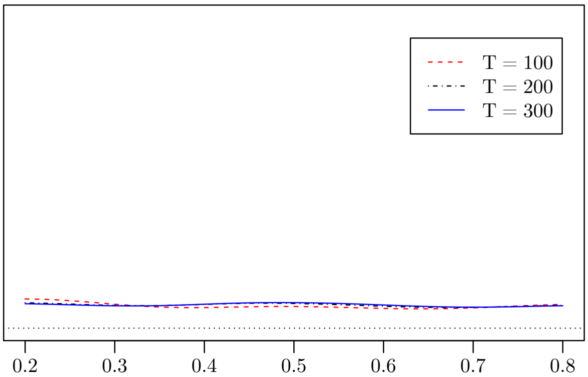
(j) $\hat{\tau}_{\text{FWLS}} : \delta = 0$, SD3, $\kappa = 1$, $\lambda_0 = 0.3$



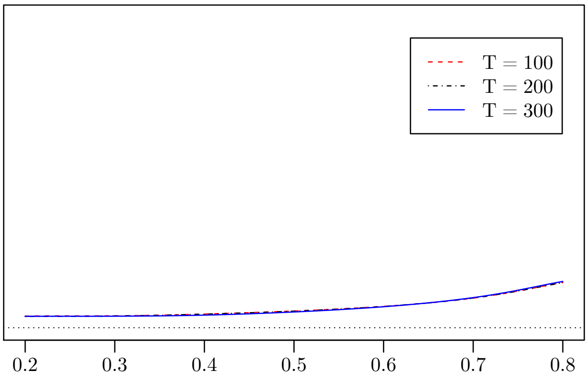
(k) $\hat{\tau}_{\text{OLS}} : \delta = 0$, SD3, $\kappa = 2$, $\lambda_0 = 0.3$



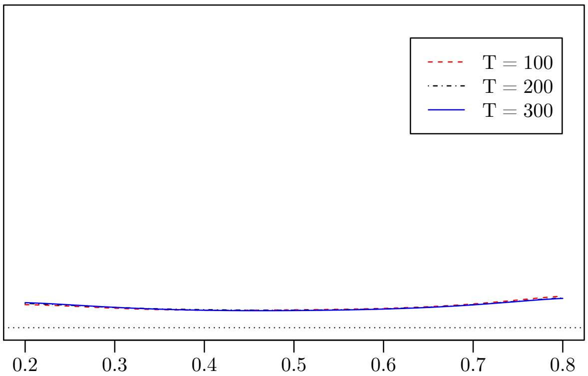
(l) $\hat{\tau}_{\text{FWLS}} : \delta = 0$, SD3, $\kappa = 2$, $\lambda_0 = 0.3$



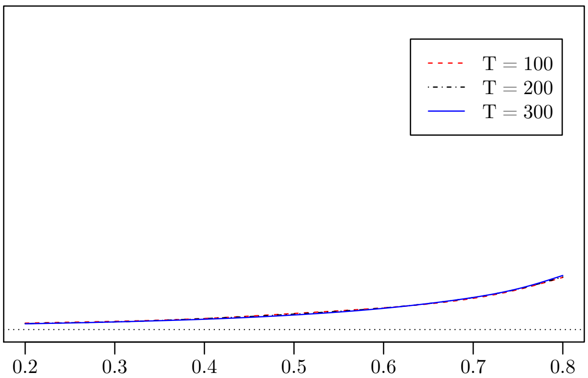
(m) $\hat{\tau}_{\text{OLS}} : \delta = 0$, SD4, $\kappa = 1$, $\lambda_0 = 0$



(n) $\hat{\tau}_{\text{FWLS}} : \delta = 0$, SD4, $\kappa = 1$, $\lambda_0 = 0$



(o) $\hat{\tau}_{\text{OLS}} : \delta = 0$, SD4, $\kappa = 2$, $\lambda_0 = 0$



(p) $\hat{\tau}_{\text{FWLS}} : \delta = 0$, SD4, $\kappa = 2$, $\lambda_0 = 0$

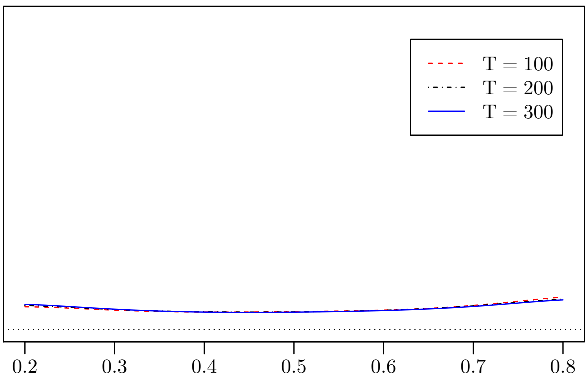
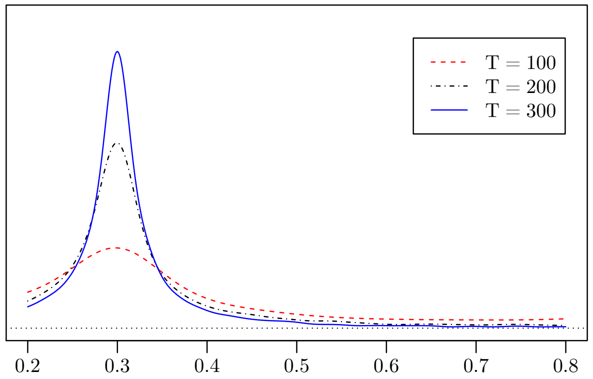
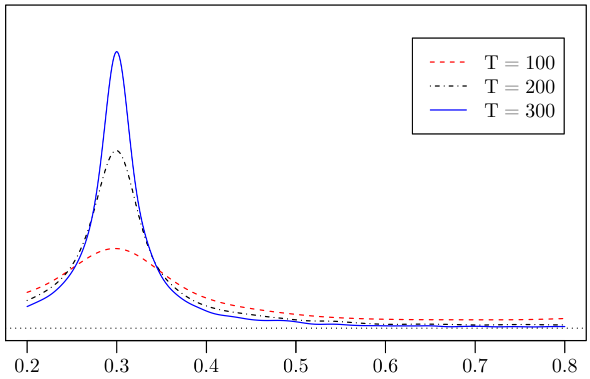


Figure 3: Simulated Sampling Density Functions of $\hat{\tau}_{OLS}$ and $\hat{\tau}_{FWLS}$. Level Break at $\tau_0 = 0.3$.

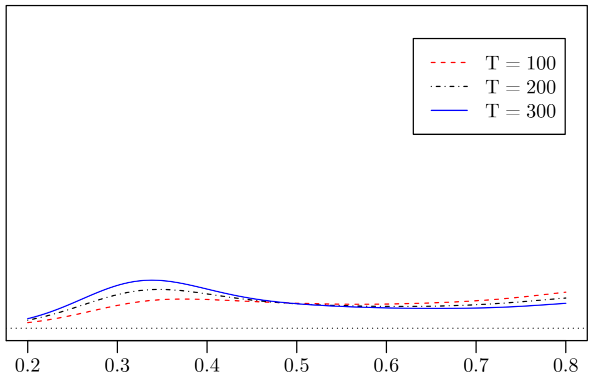
(a) $\hat{\tau}_{\text{OLS}}$: $\delta = 0.5$, $\tau_0 = 0.3$; SD0, $\kappa = 0$, $\lambda_0 = 0$



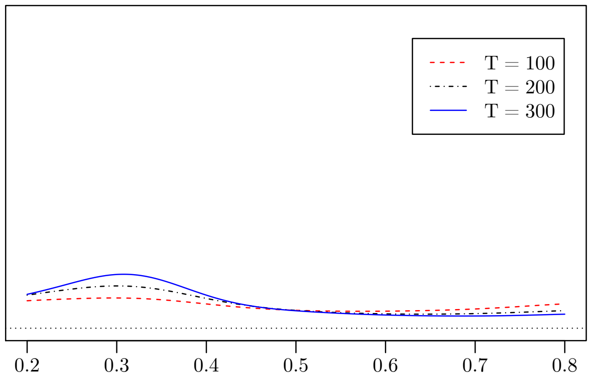
(b) $\hat{\tau}_{\text{FWLS}}$: $\delta = 0.5$, $\tau_0 = 0.3$; SD0, $\kappa = 0$, $\lambda_0 = 0$



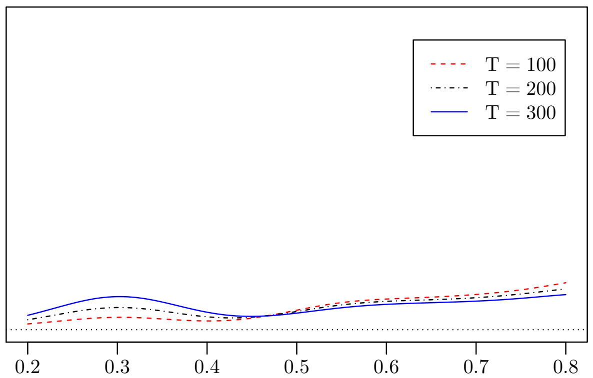
(c) $\hat{\tau}_{\text{OLS}}$: $\delta = 0.5$, $\tau_0 = 0.3$; SD1, $\kappa = 2$, $\lambda_0 = 0.3$



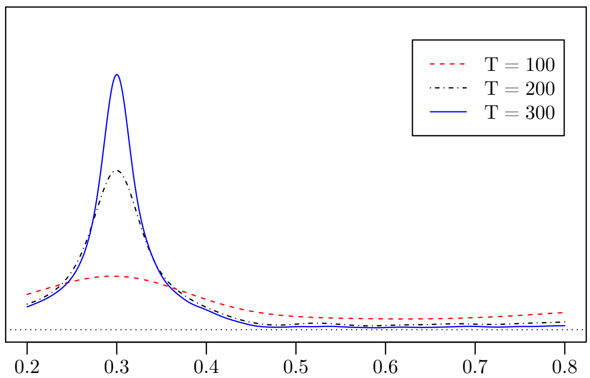
(d) $\hat{\tau}_{\text{FWLS}}$: $\delta = 0.5$, $\tau_0 = 0.3$; SD1, $\kappa = 2$, $\lambda_0 = 0.3$



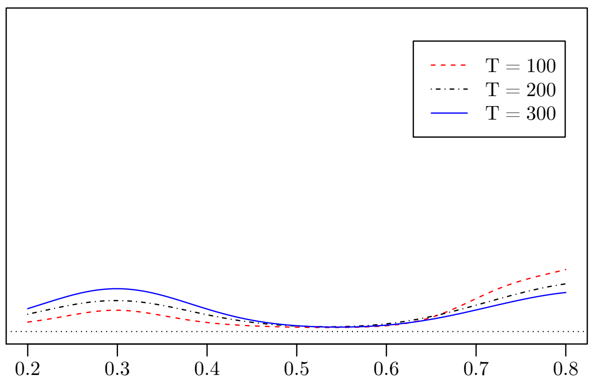
(e) $\hat{\tau}_{\text{OLS}}$: $\delta = 0.5$, $\tau_0 = 0.3$; SD1, $\kappa = 2$, $\lambda_0 = 0.5$



(f) $\hat{\tau}_{\text{FWLS}}$: $\delta = 0.5$, $\tau_0 = 0.3$; SD1, $\kappa = 2$, $\lambda_0 = 0.5$



(g) $\hat{\tau}_{\text{OLS}}$: $\delta = 0.5$, $\tau_0 = 0.3$; SD1, $\kappa = 2$, $\lambda_0 = 0.7$



(h) $\hat{\tau}_{\text{FWLS}}$: $\delta = 0.5$, $\tau_0 = 0.3$; SD1, $\kappa = 2$, $\lambda_0 = 0.7$

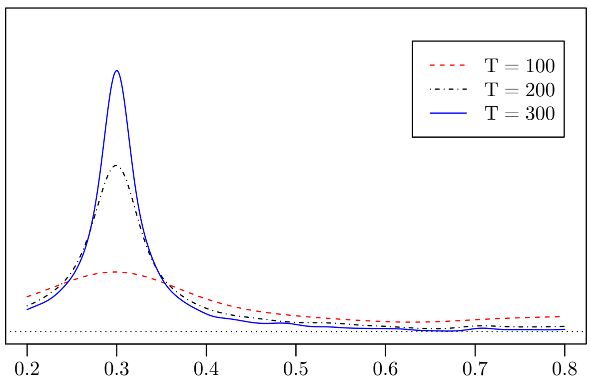
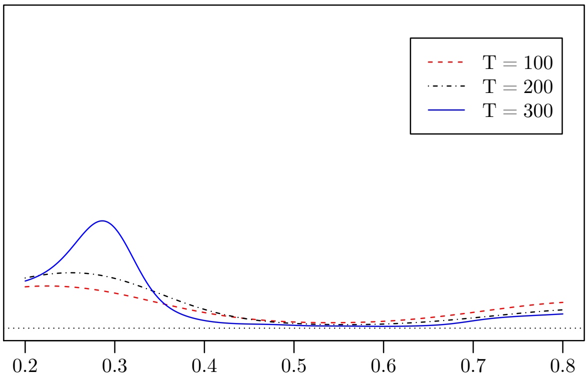
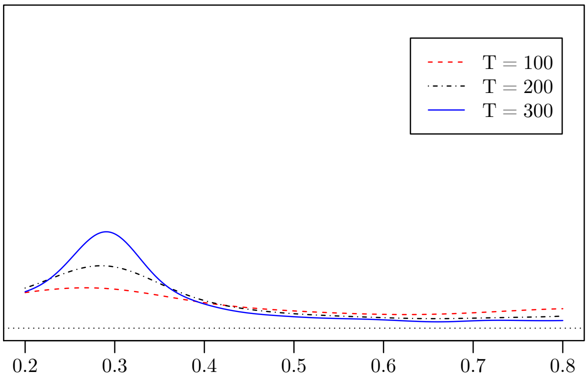


Figure 3: continued ...

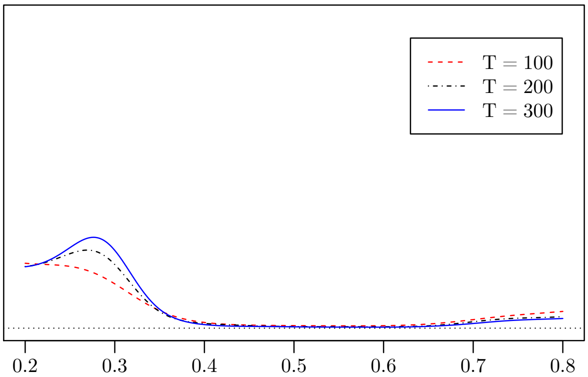
(i) $\hat{\tau}_{\text{OLS}}$: $\delta = 0.5$, $\tau_0 = 0.3$; SD3, $\kappa = 1$, $\lambda_0 = 0.3$



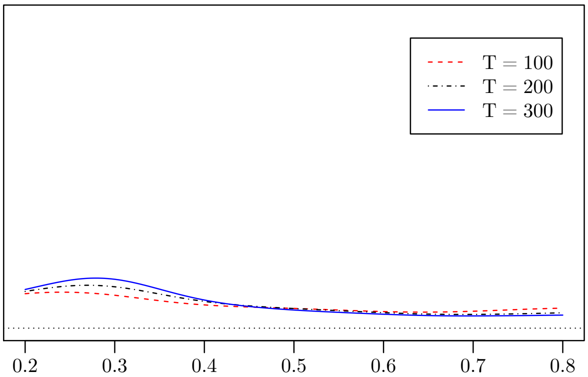
(j) $\hat{\tau}_{\text{FWLS}}$: $\delta = 0.5$, $\tau_0 = 0.3$; SD3, $\kappa = 1$, $\lambda_0 = 0.3$



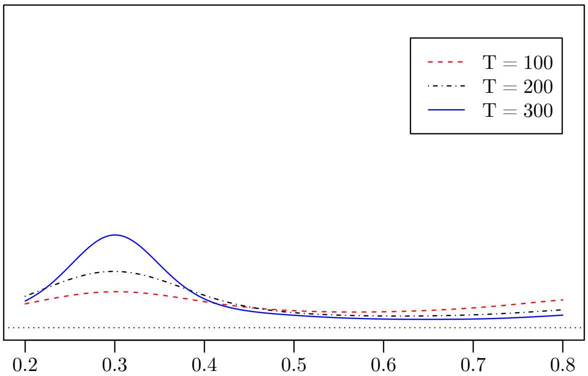
(k) $\hat{\tau}_{\text{OLS}}$: $\delta = 0.5$, $\tau_0 = 0.3$; SD3, $\kappa = 2$, $\lambda_0 = 0.3$



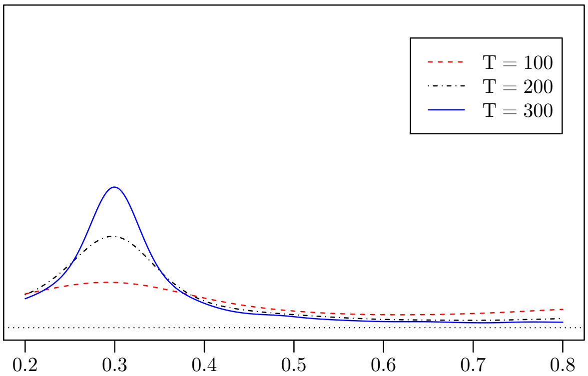
(l) $\hat{\tau}_{\text{FWLS}}$: $\delta = 0.5$, $\tau_0 = 0.3$; SD3, $\kappa = 2$, $\lambda_0 = 0.3$



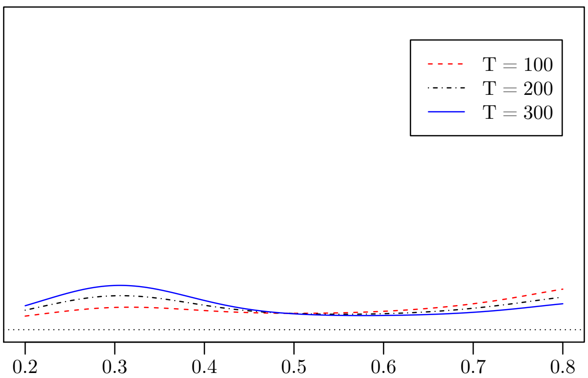
(m) $\hat{\tau}_{\text{OLS}}$: $\delta = 0.5$, $\tau_0 = 0.3$; SD4, $\kappa = 1$, $\lambda_0 = 0$



(n) $\hat{\tau}_{\text{FWLS}}$: $\delta = 0.5$, $\tau_0 = 0.3$; SD4, $\kappa = 1$, $\lambda_0 = 0$



(o) $\hat{\tau}_{\text{OLS}}$: $\delta = 0.5$, $\tau_0 = 0.3$; SD4, $\kappa = 2$, $\lambda_0 = 0$



(p) $\hat{\tau}_{\text{FWLS}}$: $\delta = 0.5$, $\tau_0 = 0.3$; SD4, $\kappa = 2$, $\lambda_0 = 0$

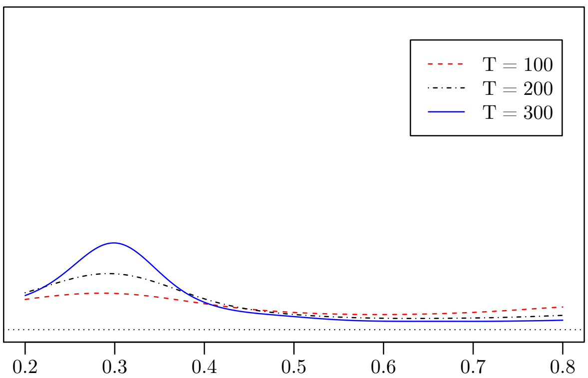


Figure 4: Simulated Sampling Density Functions of $\hat{\tau}_{OLS}$ and $\hat{\tau}_{FWLS}$. Level Break at $\tau_0 = 0.5$.

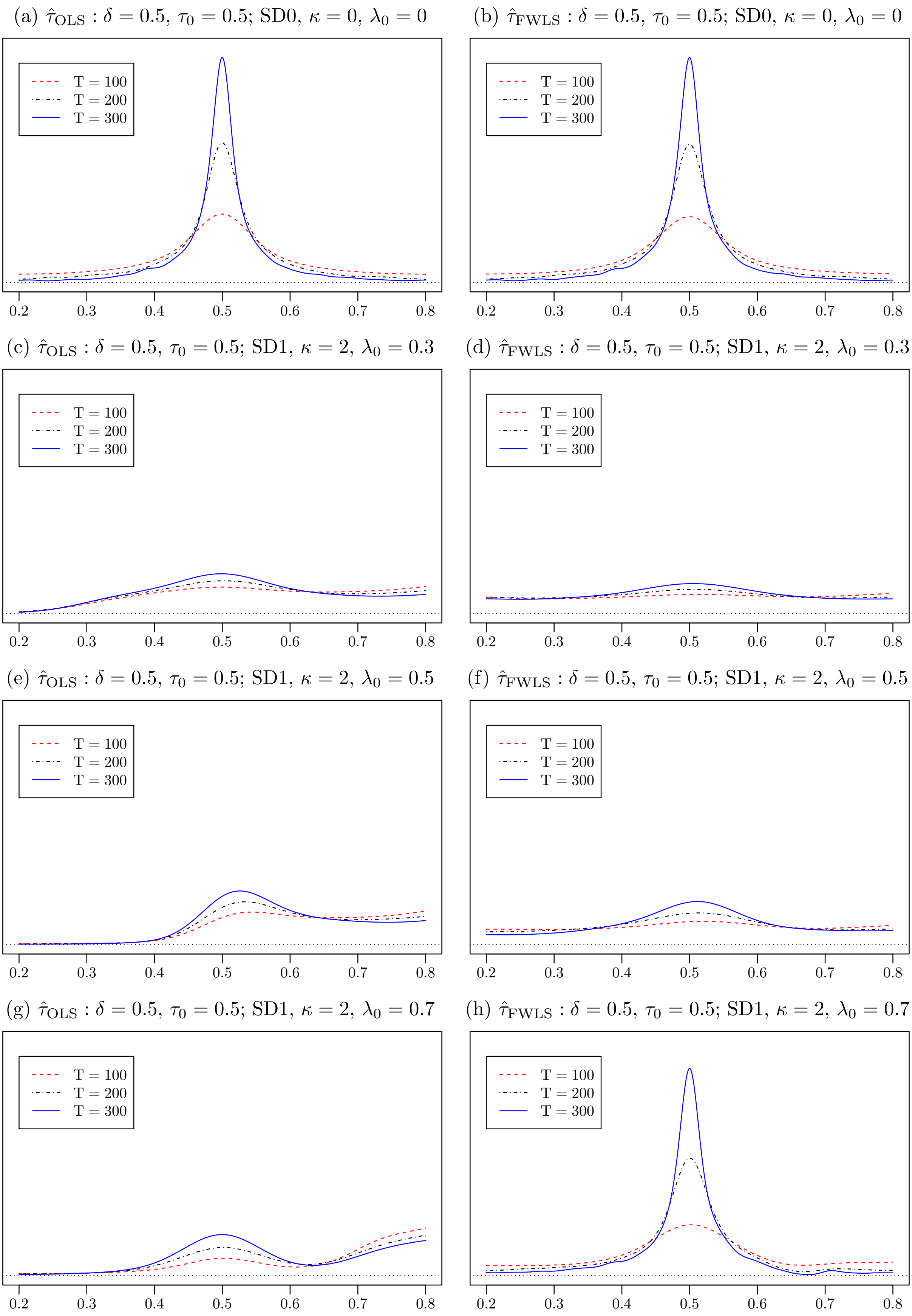
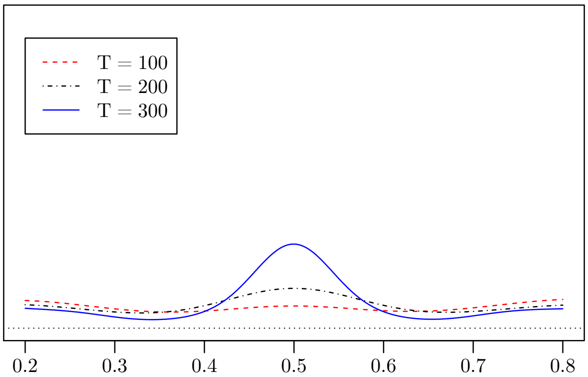
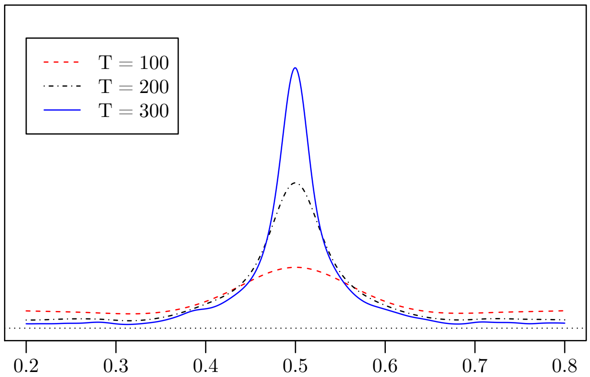


Figure 4: continued ...

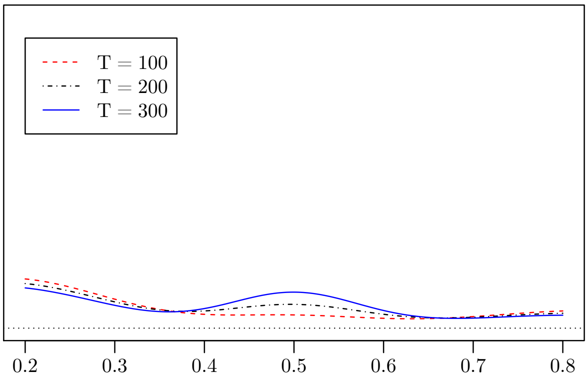
(i) $\hat{\tau}_{\mathrm{OLS}}$: $\delta = 0.5$, $\tau_0 = 0.5$; SD3, $\kappa = 1$, $\lambda_0 = 0.3$



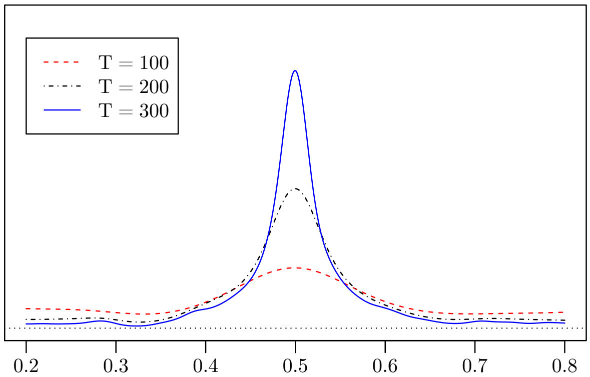
(j) $\hat{\tau}_{\mathrm{FWLS}}$: $\delta = 0.5$, $\tau_0 = 0.5$; SD3, $\kappa = 1$, $\lambda_0 = 0.3$



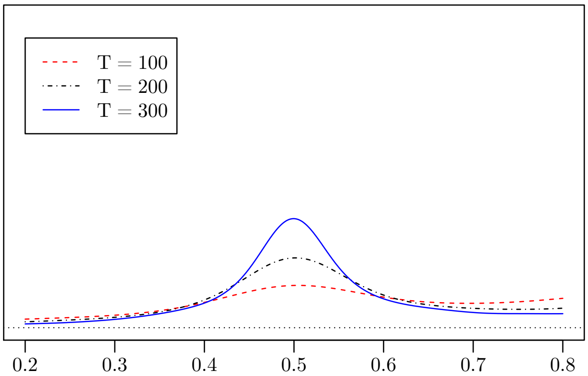
(k) $\hat{\tau}_{\mathrm{OLS}}$: $\delta = 0.5$, $\tau_0 = 0.5$; SD3, $\kappa = 2$, $\lambda_0 = 0.3$



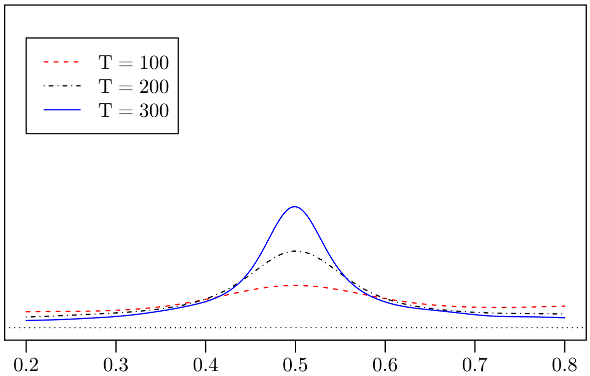
(l) $\hat{\tau}_{\mathrm{FWLS}}$: $\delta = 0.5$, $\tau_0 = 0.5$; SD3, $\kappa = 2$, $\lambda_0 = 0.3$



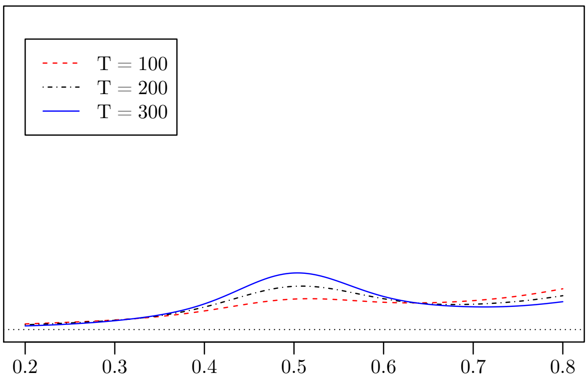
(m) $\hat{\tau}_{\mathrm{OLS}}$: $\delta = 0.5$, $\tau_0 = 0.5$; SD4, $\kappa = 1$, $\lambda_0 = 0$



(n) $\hat{\tau}_{\mathrm{FWLS}}$: $\delta = 0.5$, $\tau_0 = 0.5$; SD4, $\kappa = 1$, $\lambda_0 = 0$



(o) $\hat{\tau}_{\mathrm{OLS}}$: $\delta = 0.5$, $\tau_0 = 0.5$; SD4, $\kappa = 2$, $\lambda_0 = 0$



(p) $\hat{\tau}_{\mathrm{FWLS}}$: $\delta = 0.5$, $\tau_0 = 0.5$; SD4, $\kappa = 2$, $\lambda_0 = 0$

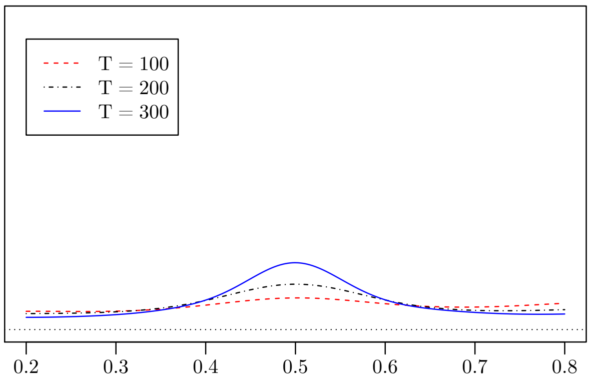


Figure 5: Simulated Sampling Density Functions of $\hat{\tau}_{OLS}$ and $\hat{\tau}_{FWLS}$. Level Break at $\tau_0 = 0.7$.

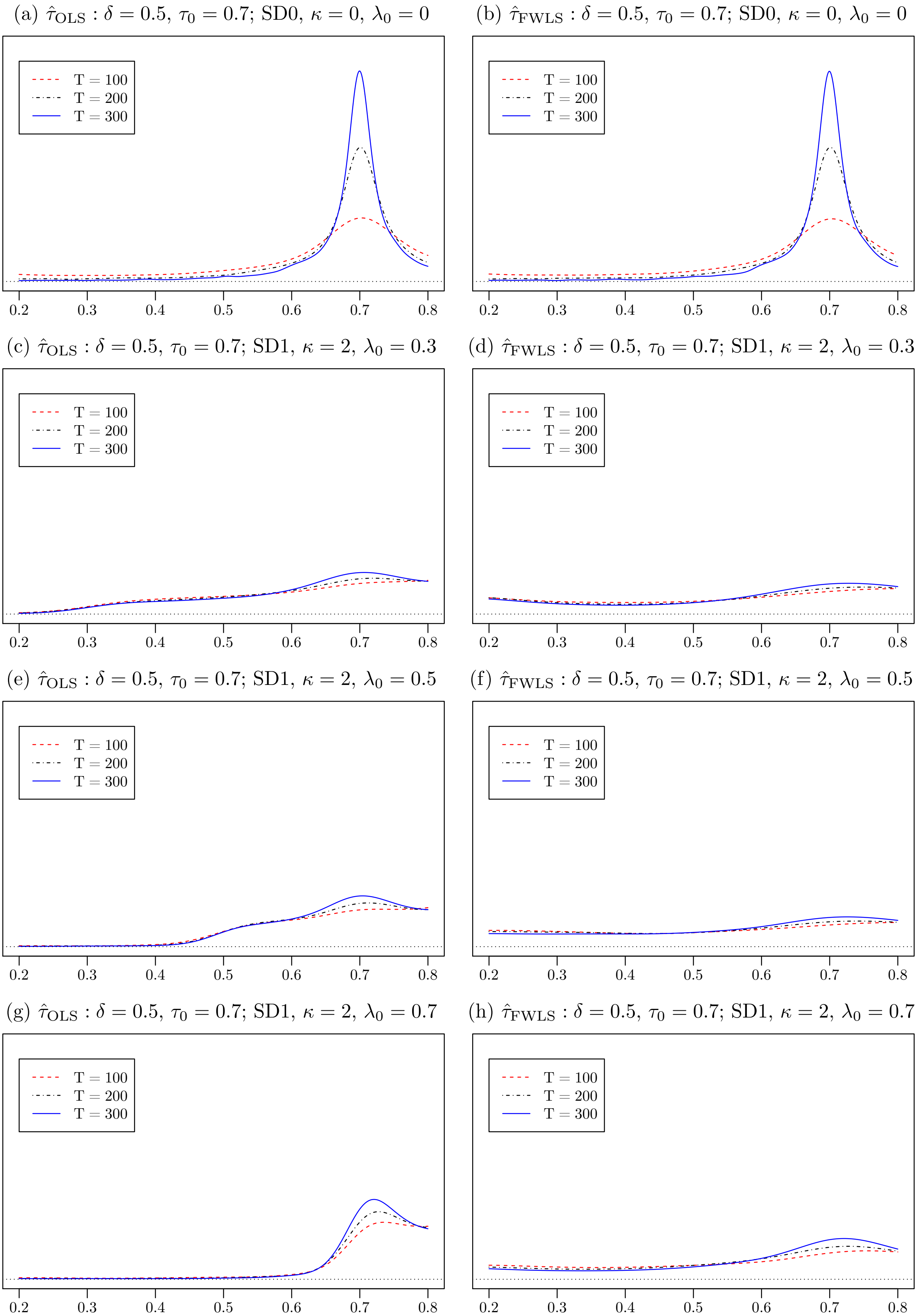
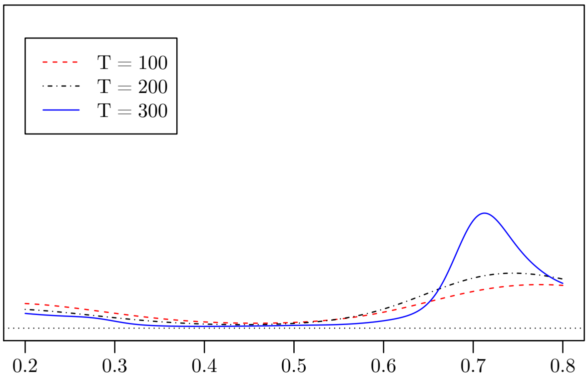
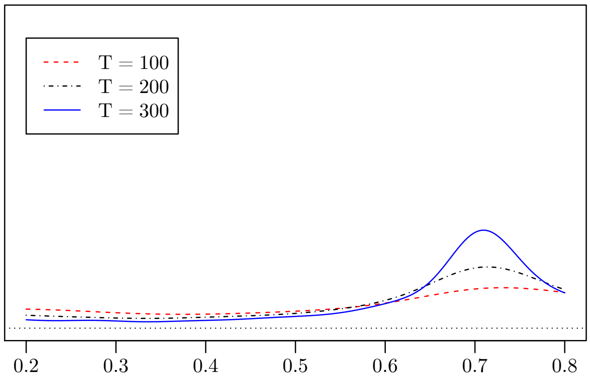


Figure 5: continued ...

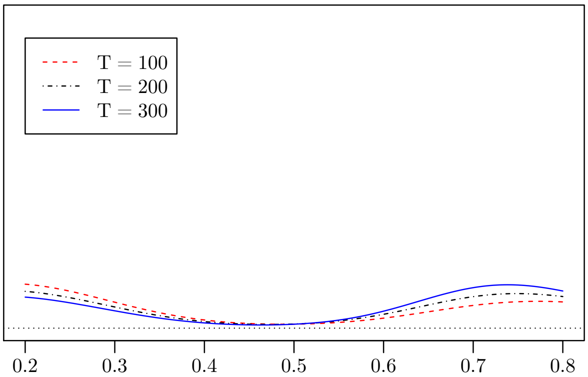
(i) $\hat{\tau}_{\text{OLS}} : \delta = 0.5,\ \tau_0 = 0.7$; SD3, $\kappa = 1,\ \lambda_0 = 0.3$



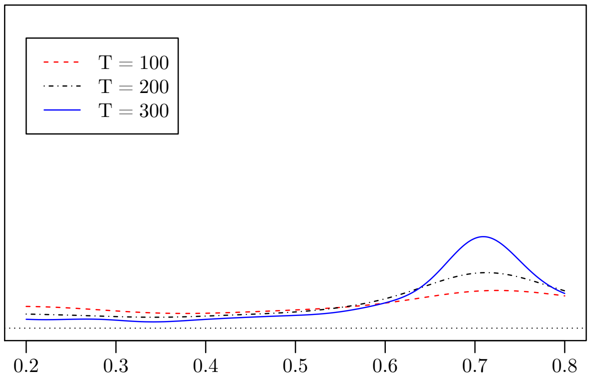
(j) $\hat{\tau}_{\text{FWLS}} : \delta = 0.5,\ \tau_0 = 0.7$; SD3, $\kappa = 1,\ \lambda_0 = 0.3$



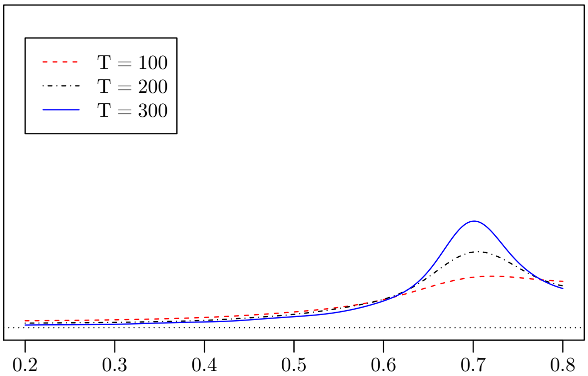
(k) $\hat{\tau}_{\text{OLS}} : \delta = 0.5,\ \tau_0 = 0.7$; SD3, $\kappa = 2,\ \lambda_0 = 0.3$



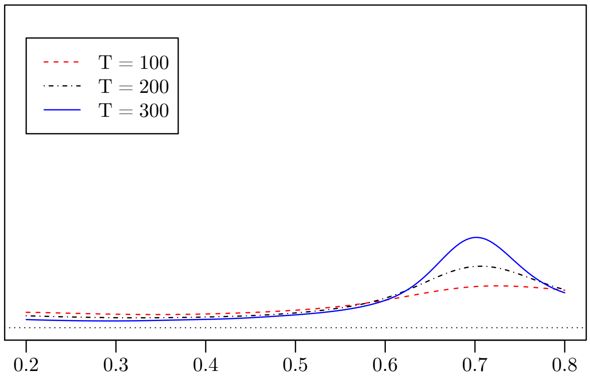
(l) $\hat{\tau}_{\text{FWLS}} : \delta = 0.5,\ \tau_0 = 0.7$; SD3, $\kappa = 2,\ \lambda_0 = 0.3$



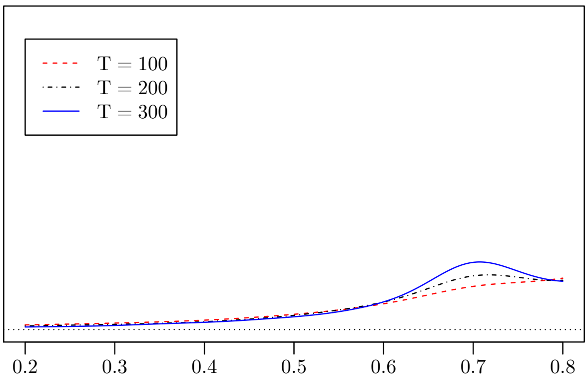
(m) $\hat{\tau}_{\text{OLS}} : \delta = 0.5,\ \tau_0 = 0.7$; SD4, $\kappa = 1,\ \lambda_0 = 0$



(n) $\hat{\tau}_{\text{FWLS}} : \delta = 0.5,\ \tau_0 = 0.7$; SD4, $\kappa = 1,\ \lambda_0 = 0$



(o) $\hat{\tau}_{\text{OLS}} : \delta = 0.5,\ \tau_0 = 0.7$; SD4, $\kappa = 2,\ \lambda_0 = 0$



(p) $\hat{\tau}_{\text{FWLS}} : \delta = 0.5,\ \tau_0 = 0.7$; SD4, $\kappa = 2,\ \lambda_0 = 0$

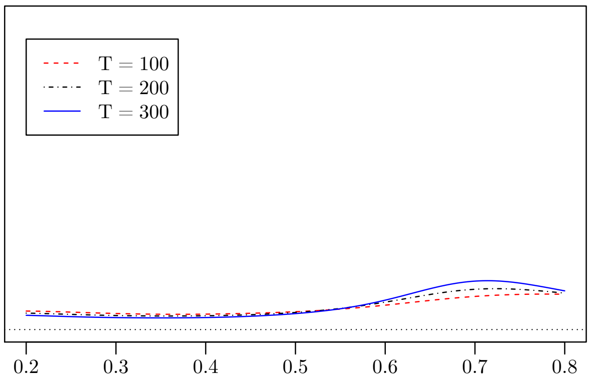


Figure 6: Simulated sampling density functions with $\delta_T = \delta T^{-1/2}$

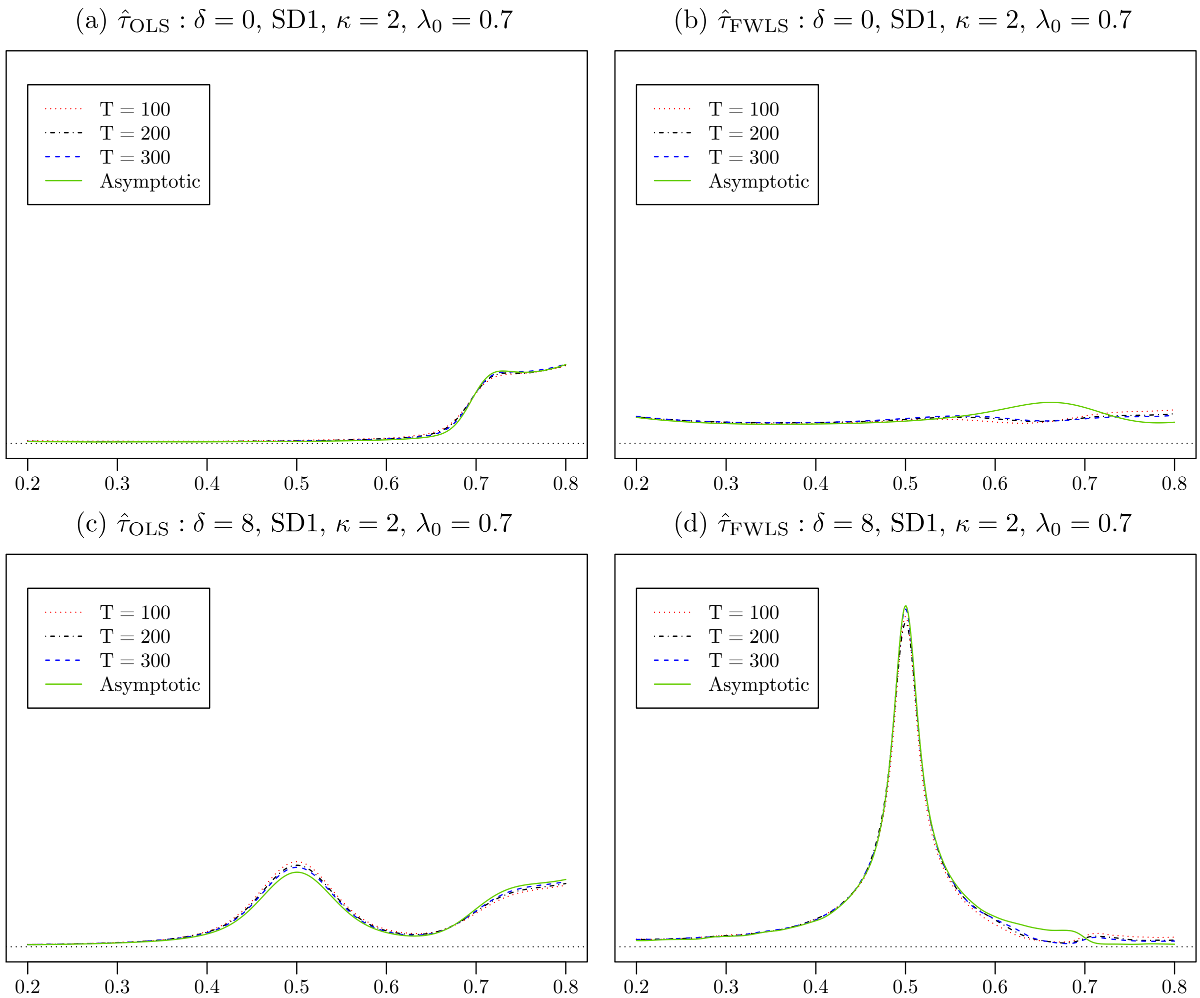


Figure 7: Finite Sample Local Power Comparisons, $T = 100$

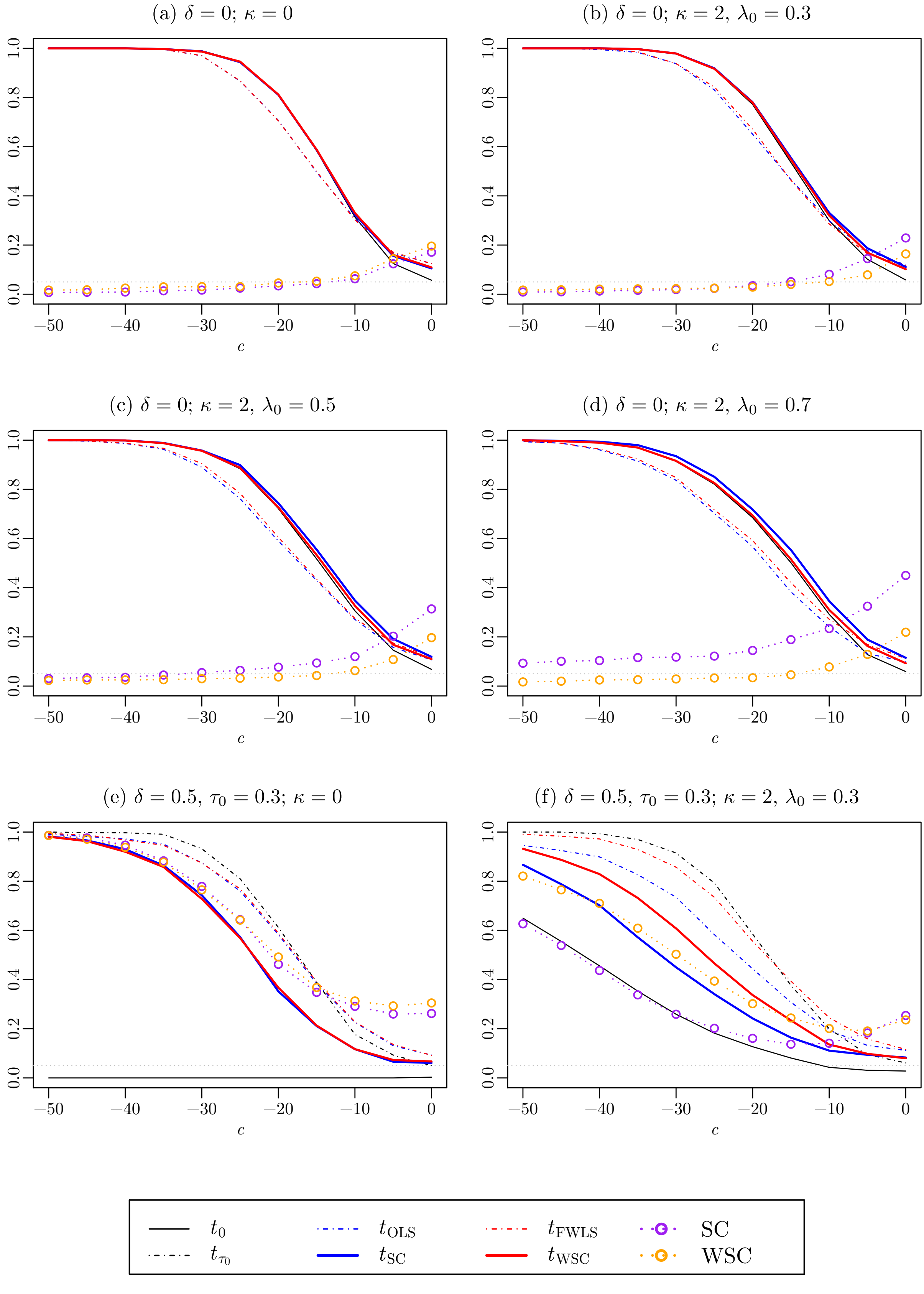


Figure 7: continued ...

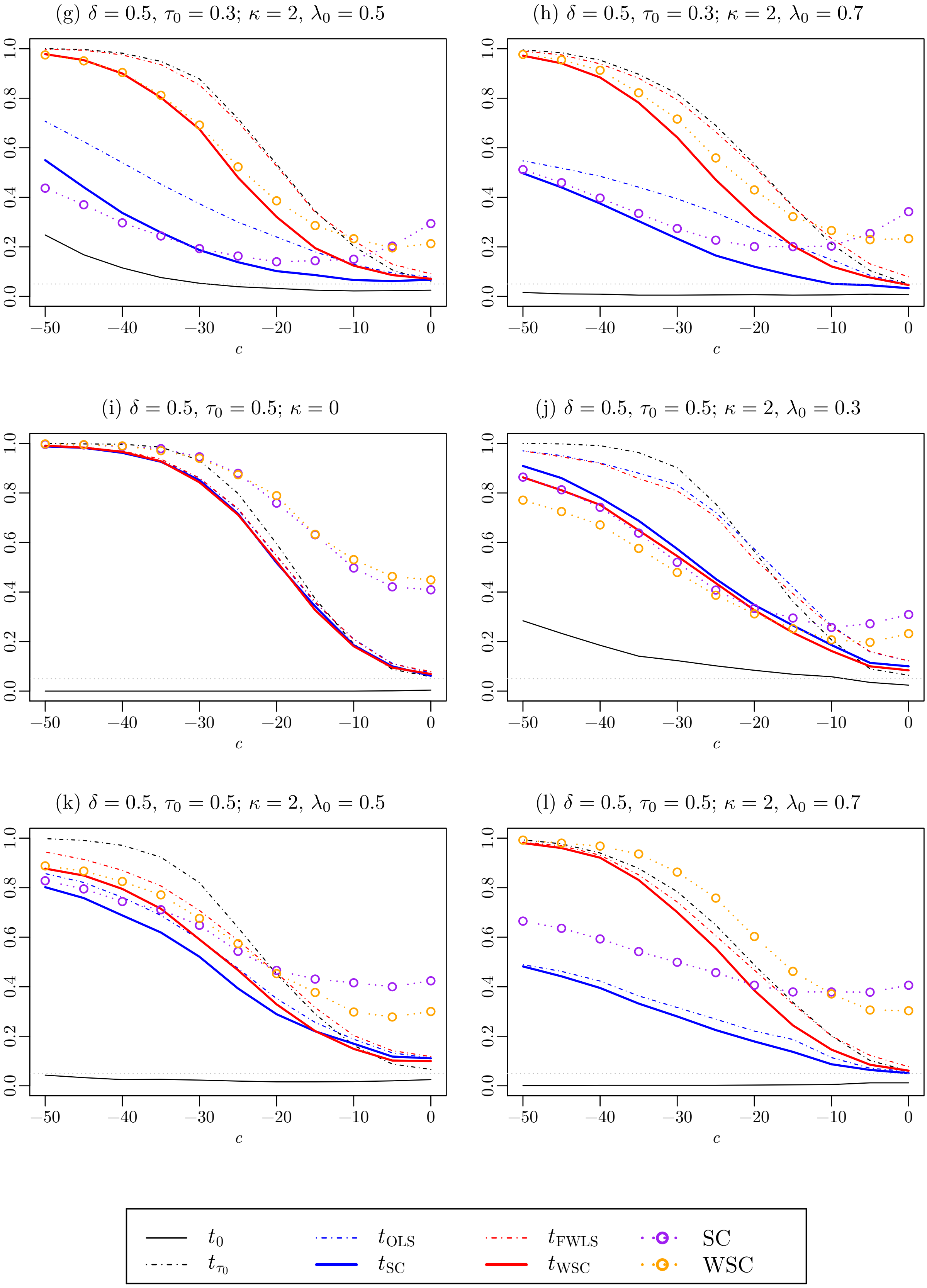


Figure 7: continued ...

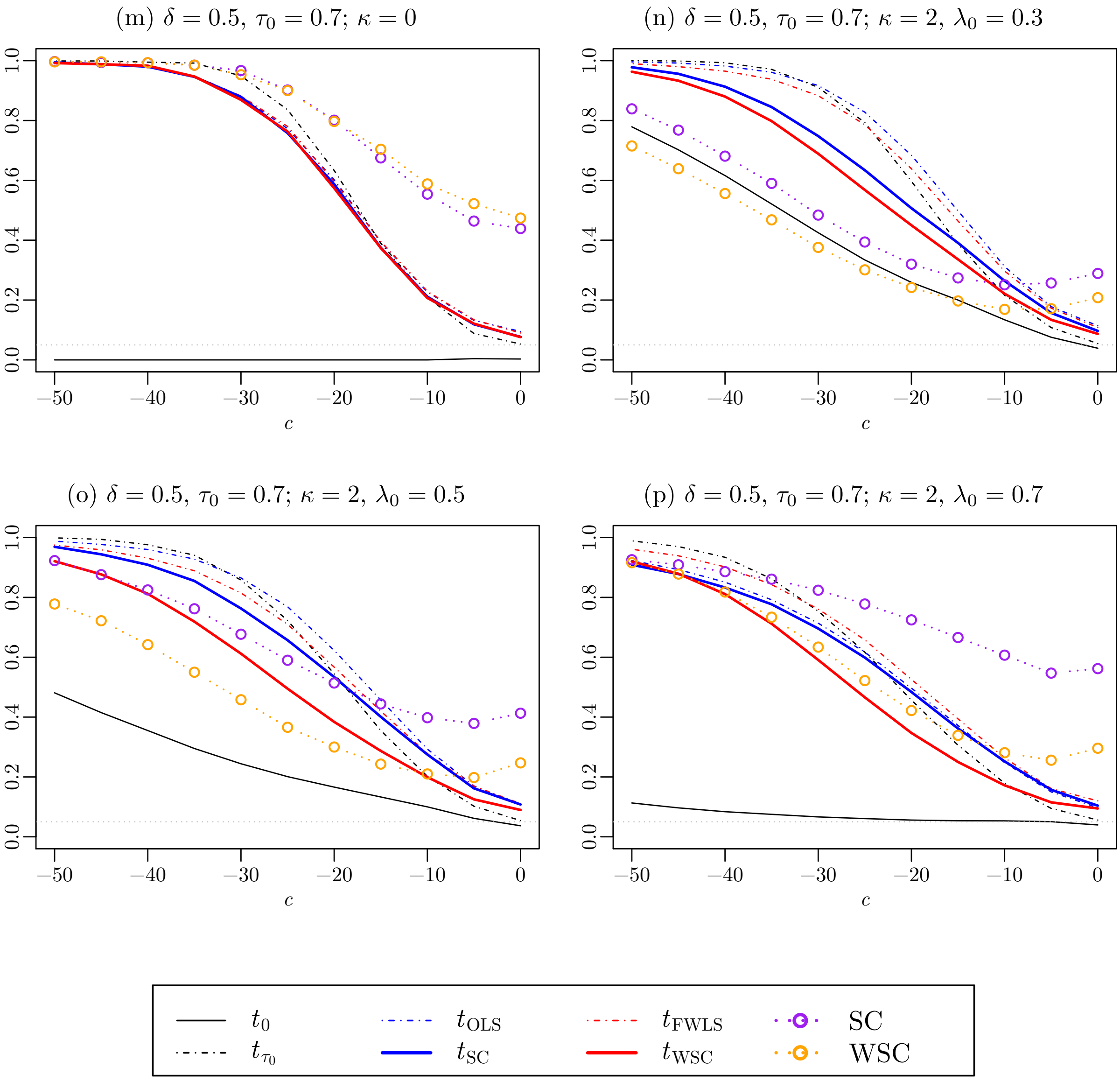


Figure 8: Finite Sample Local Power Comparisons, $T = 200$

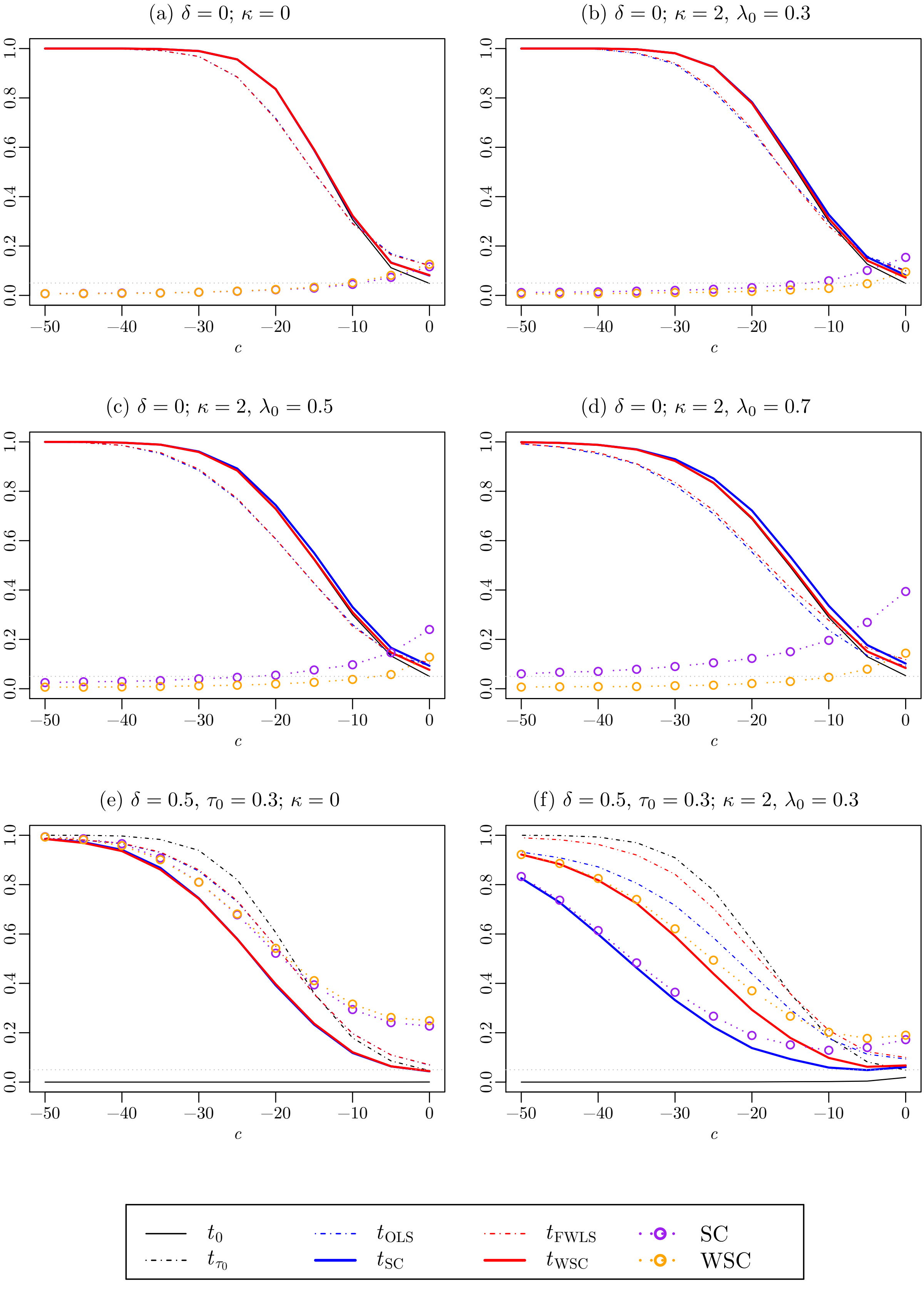


Figure 8: continued ...

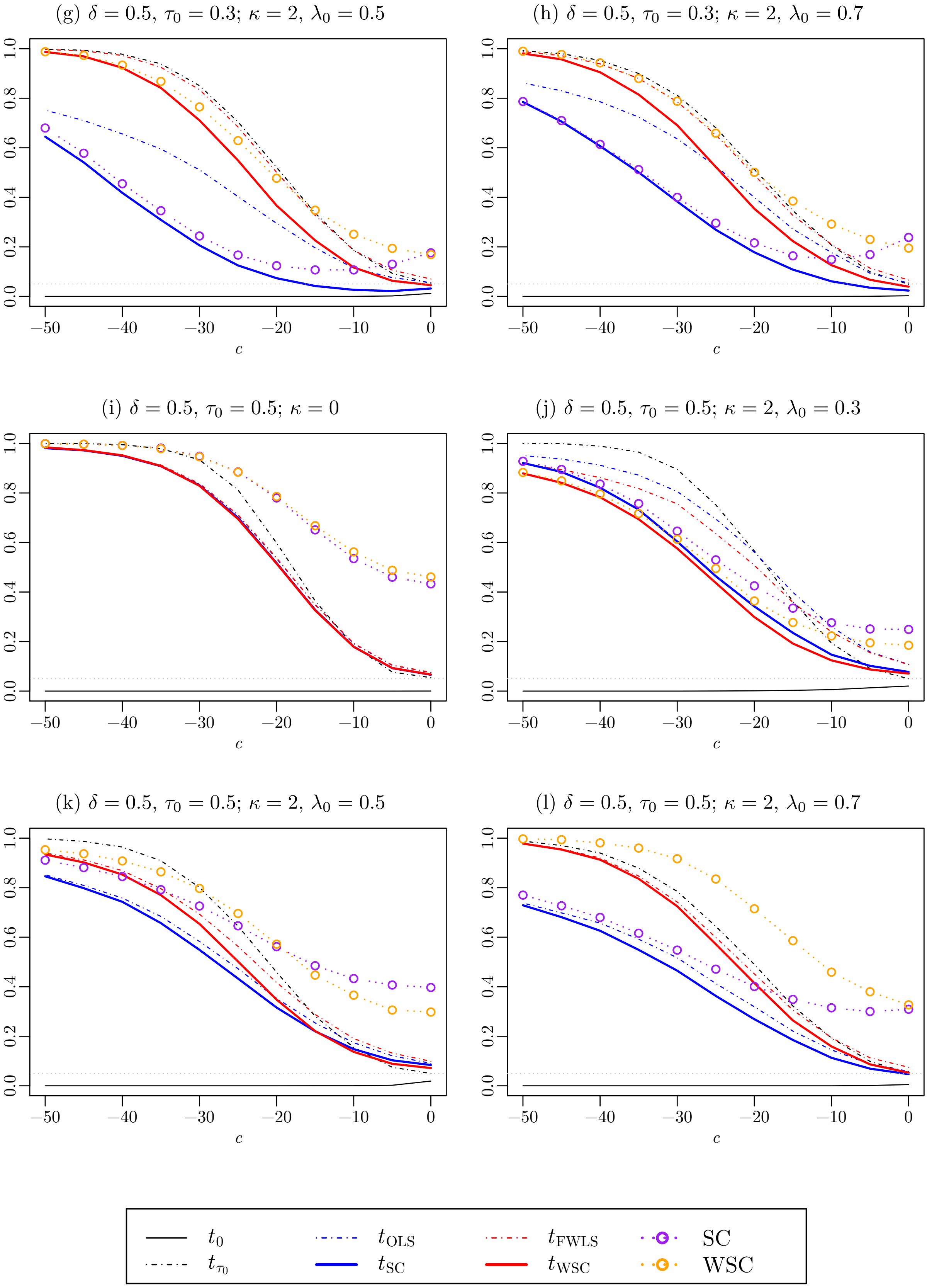


Figure 8: continued ...

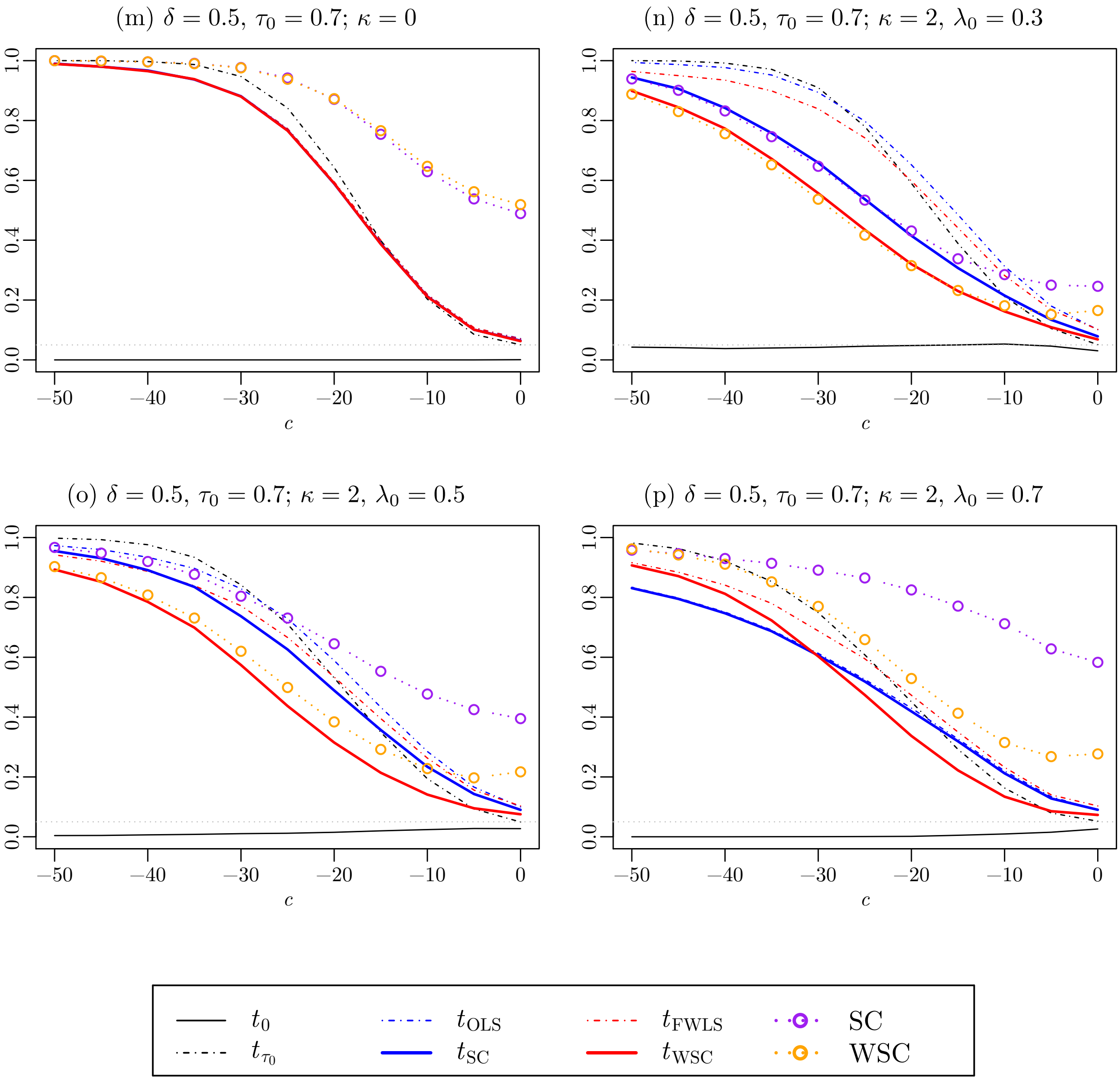


Figure 9: (a) Annualised quarterly U.S. real GDP growth rates with fitted OLS and FWLS level break path estimates; (b) estimated volatility path; (c) estimated variance profile; (d) centered variance profile estimate.

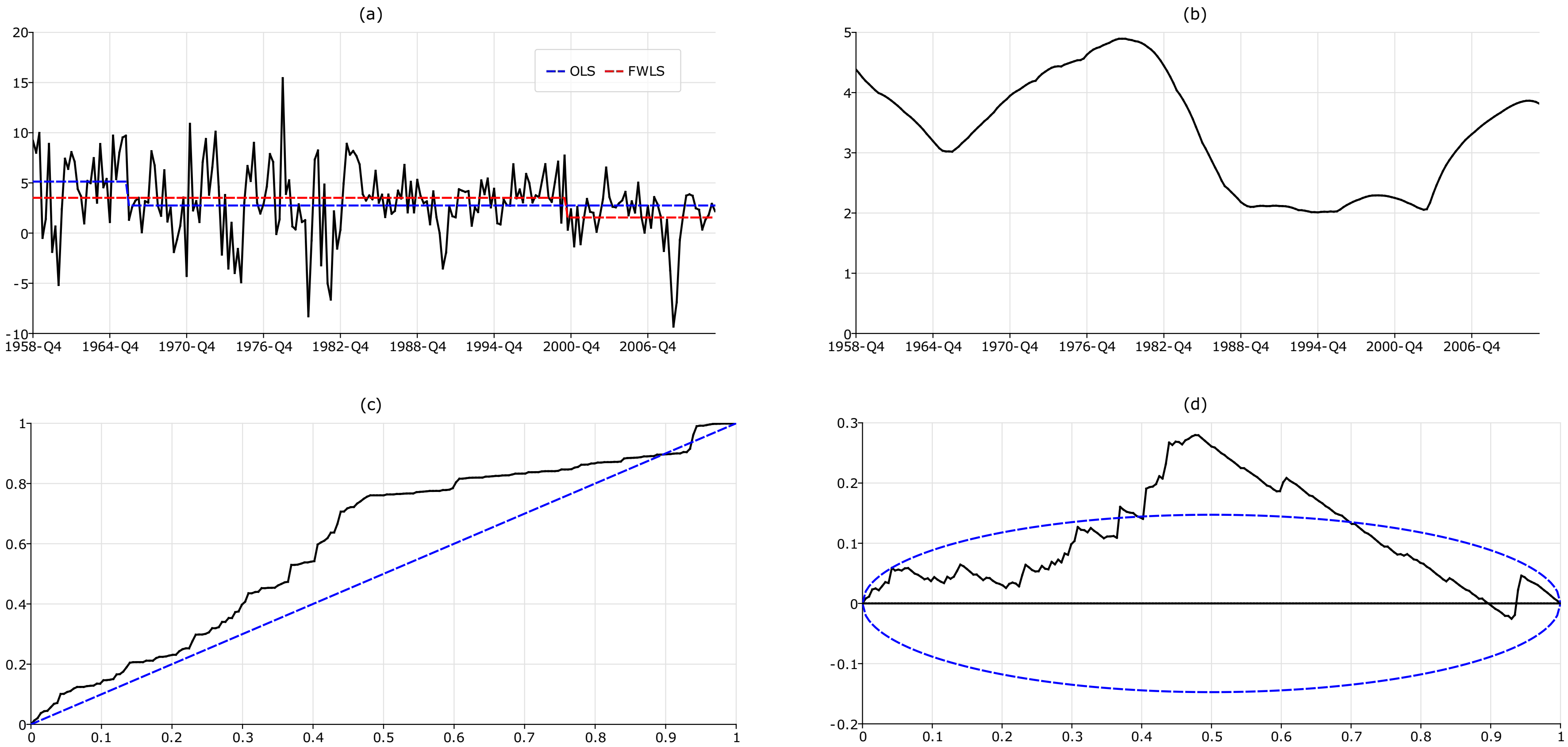
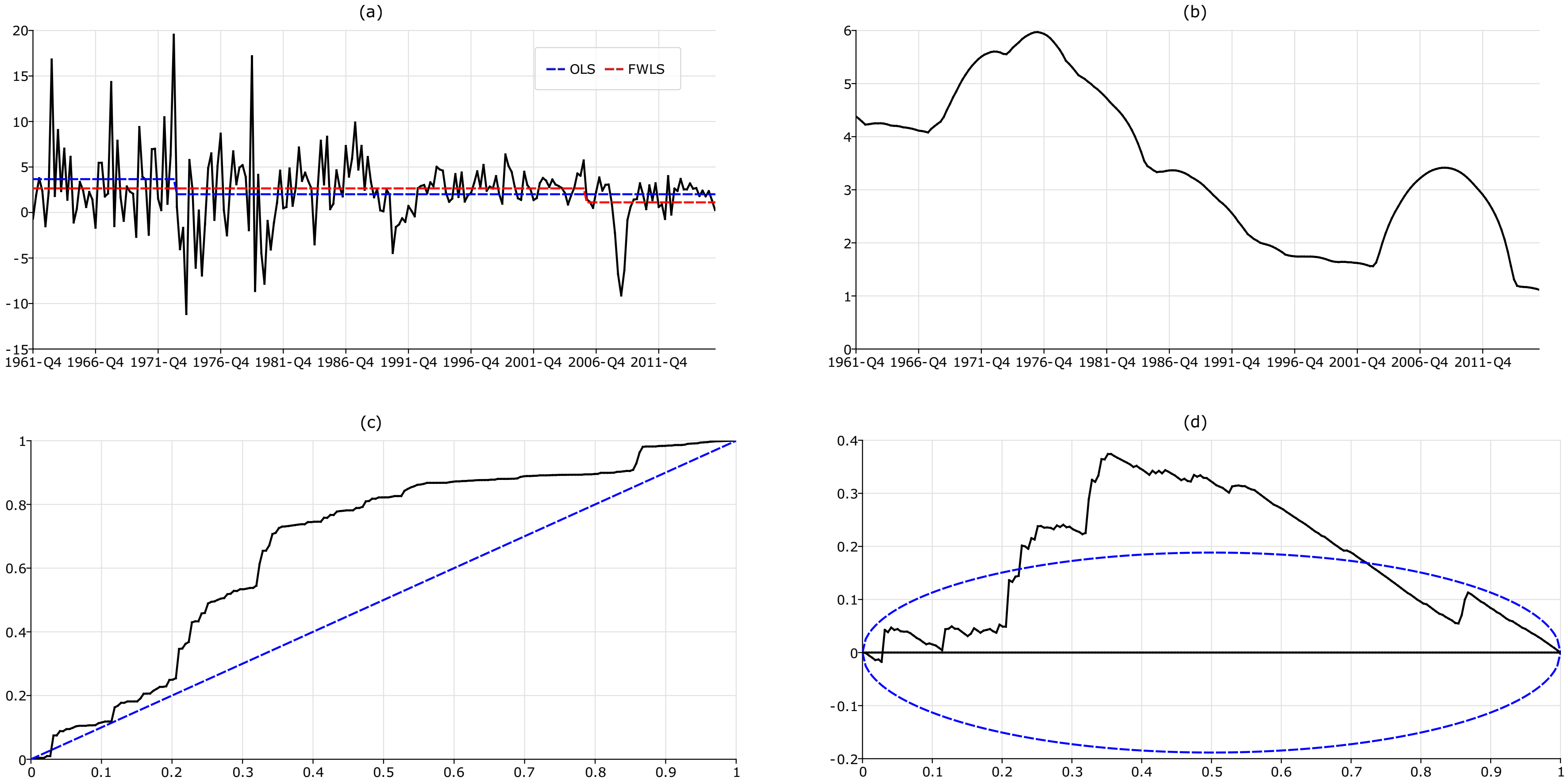


Figure 10: (a) Annualised quarterly U.K. real GDP growth rates with fitted OLS and FWLS level break path estimates; (b) estimated volatility path; (c) estimated variance profile; (d) centered variance profile estimate.



On-Line Supplementary Appendix

to

Level Shift Estimation in the Presence of Non-stationary Volatility with an Application to the Unit Root Testing Problem

by

D. Harris, H. Kew and A.M.R. Taylor

Contents: Section S.1 provides some example calculations relating to Case 2 in section 3.3. Section S.2 of this supplement reports the large sample properties of the unit root tests and IC-based trend break selection procedures. Section S.3 contains graphs of the U.K. and U.S. real GDP series considered in section 5. Section S.4 provides proofs of Theorems 1-3 and Theorems S.1 and S.2. Additional references, where not cited within the main paper, are provided at the end of this supplementary appendix.

S.1 Example Calculations Relating to Case 2 of Section 3.3

In this section we provide two numerical examples of the calculations discussed in Case 2 in section 3.3 relating to the Pitman rate local asymptotic distribution theory from Theorem 2. Example 1 relates to the case where the variance follows a linear trend path, while in Example 2 a one-time change in volatility occurs. Some additional Monte Carlo simulations are also included to investigate how well Theorem 2 predicts the finite sample behaviour of the $\hat{\tau}_{FWLS}$ estimator in this regard.

Example 1: Consider the case where the weighted estimator is formed on the assumption that the variance follows the linear trend path $\sigma_t^2 = 1 + t/T$. The corresponding weighted estimator obtains setting $x_t = 1/(1+t/T)^{1/2}$, and hence $x(s) = (1+s)^{-1/2}$ and $\chi(\tau) = \frac{\log(1+\tau)}{\log 2}$. Then $\chi(\tau_0) = \frac{1}{2}$ gives $\tau_0 = \sqrt{2} - 1 \approx 0.414$. Consequently, when weighting is used appropriate for a linear trend in the variance the position of a break fraction that maximises the asymptotic signal in $Q(\tau; x(\cdot), \sigma(\cdot), \delta, d)$ is $\tau_0 \approx 0.414$, rather than $\tau_0 = 0.5$. Notice that this result obtains regardless of whether this weighting leads to the true WLS estimator; that is, the result holds regardless of the true variance process, σ_t. □

Example 2: As a second example, suppose that it is assumed that there is one-time change in variance at time $\lfloor T\lambda \rfloor$; that is, under the assumption that $\sigma_t = 1 + \kappa 1_{\frac{t}{T} > \lambda}$. Here $x_t = 1/\sigma_t$, so that

$$x(s) = \begin{cases} 1, & s \leq \lambda \\ (1+\kappa)^{-1}, & s > \lambda \end{cases} \quad \text{and, hence,} \quad \chi(\tau) = \begin{cases} \frac{\tau}{\lambda + (1+\kappa)^{-2}(1-\lambda)}, & \tau \leq \lambda \\ \frac{\lambda + (1+\kappa)^{-2}(\tau - \lambda)}{\lambda + (1+\kappa)^{-2}(1-\lambda)}, & \tau > \lambda. \end{cases}$$

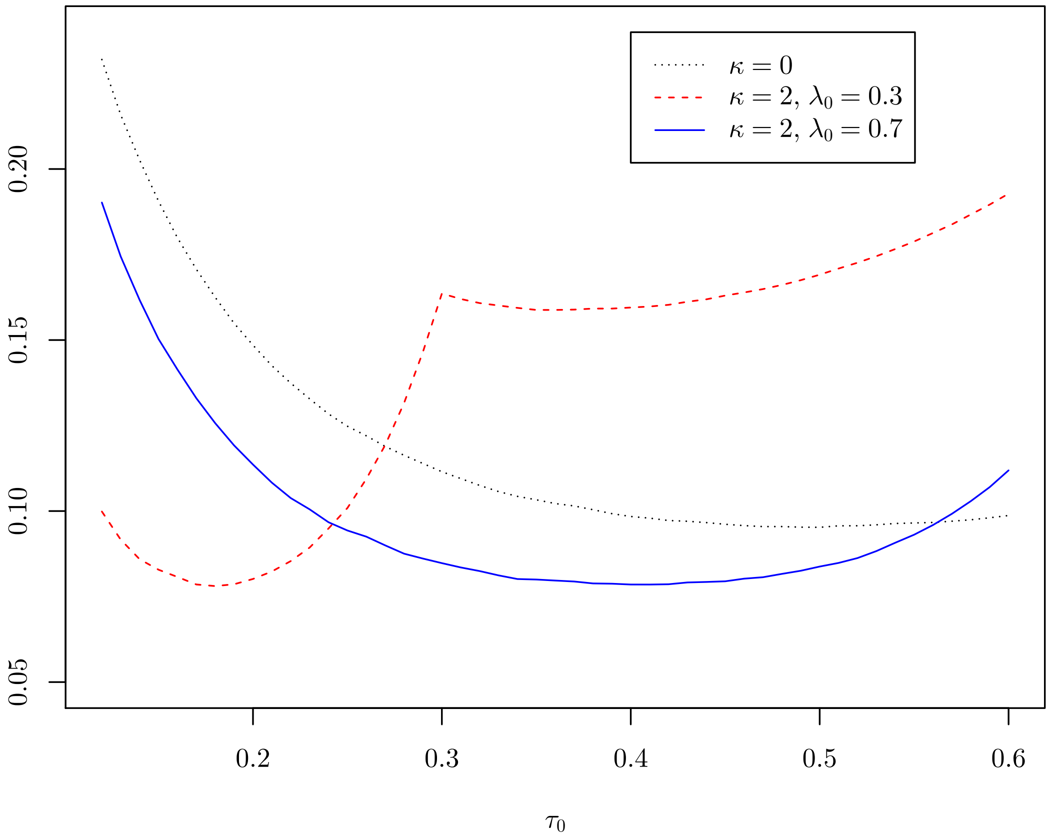
In the homoskedastic case, where no break in variance occurs, such that $\kappa = 0$, the weighted and unweighted estimators coincide. Solving here, $\chi(\tau_0) = \frac{1}{2}$ gives $\tau_0 = \frac{1}{2}$, as expected; that is, with homoskedastic data the asymptotic signal in $Q(\tau; x(\cdot), \sigma(\cdot), \delta, d)$ is maximised for a break occurring in the middle of the sample. However, in the case where $\lambda = 0.3$ and $\kappa = 2$, such that the volatility increases threefold 30% of the way into the sample, then solving $\chi(\tau_0) = \frac{1}{2}$ yields $\tau_0 \approx 0.19$. Using the weights appropriate to this form of variance step function therefore results in the largest "signal" for a break occurring at $\tau_0 \approx 0.19$. In contrast if $\lambda = 0.7$ and $\kappa = 2$, such that the volatility increases threefold 70% of the way into the sample, then solving $\chi(\tau_0) = \frac{1}{2}$ yields $\tau_0 \approx 0.40$. We therefore see, again noting that these results obtain regardless of whether or not these weightings lead to the true WLS estimator in each case, that in these two examples the weighting based on either the assumption of an early or late increase in variance results in the largest "signal" for a break occurring in the lower variance regime of the sample, as seems intuitively reasonable. □

The discussion given in Example 2 can also be illustrated numerically. A simulation experiment was carried out based on the DGP (3.10) with

$$\sigma_t = \frac{(\lambda_0(1+\kappa)^2 + (1-\lambda_0))^{1/2}}{1+\kappa}(1 + \kappa \cdot 1_{t > \lfloor \lambda_0 T \rfloor}).$$

The additional scaling of σ_t relative to that in Example 2 above is used so that $\omega^2 = 1$ for this standard deviation process for any values of λ_0 and κ, which allows meaningful comparisons to be drawn across these two parameters. In particular, simulation can be used to obtain approximations to the true value of the break fraction τ_0 that can be more accurately estimated from a finite sample data, based on different values of the standard deviation parameters. The analysis of the signal given in Example 2 above suggested that the multiplier on the deterministic signal component of $Q(\tau; x(\cdot), \sigma(\cdot), \delta, 1/2)$ would be maximised at $\tau_0 = 1/2$ for $\kappa = 0$ (homoskedasticity), at $\tau_0 \approx 0.19$ for $\kappa = 2$ and $\lambda_0 = 0.3$, and at $\tau_0 \approx 0.40$ for $\kappa = 2$ and $\lambda_0 = 0.7$. These calculations do not constitute formal proof that these values of τ_0 are those that can be most efficiently estimated under these variance patterns. However, the simulation results summarised in Figure S.1 show that they provide a good approximation in these cases at least. Figure S.1 gives plots, one for each of the three variance processes discussed here, of the simulated RMSEs of $\hat{\tau}_{FWLS}$ for estimating each of the indicated values of τ_0 the horizontal axis, based on samples of size $T = 200$ and with break size $\delta_T = 8T^{-1/2}$ (i.e. the same break size considered in Figure 6 for the purposes of comparison). The values of τ_0 that returned minimum RMSE of $\hat{\tau}_{FWLS}$ in each case were respectively $\tau_0 = 0.49$ ($\kappa = 0$), $\tau_0 = 0.18$ ($\kappa = 2, \lambda_0 = 0.3$) and $\tau_0 = 0.40$ ($\kappa = 2, \lambda_0 = 0.7$). Again we see that, even when used heuristically, the asymptotic approximation provided by $Q(\tau; x(\cdot), \sigma(\cdot), \delta, 1/2)$ in Theorem 2 provides a very useful guide to finite sample properties.

Figure S.1: Simulated RMSEs of $\hat{\tau}_{FWLS}$. $T = 200$, $\delta_T = 8T^{-1/2}$.



S.2 Limiting Distributions for Unit Root Tests and Model Selection

Theorem S.1 provides the limiting distribution under the local alternative H_c of t_τ when evaluated at the true break fraction $\tau = \tau_0$. The theorem also shows that this limit is unchanged when τ_0 is replaced by either the OLS break fraction estimate, $\hat{\tau}_{OLS}$, or the corresponding feasible WLS estimate, $\hat{\tau}_{FWLS}$.

Theorem S.1. *Let y_t be generated according to* (4.1)-(4.2) *and with e_t generated according to* (2.2)*, and let the conditions of Assumption $\mathcal{A}$ hold. Let $\delta_T = \delta T^{-d}$, $d \geq 0$. Then, under H_c:*

(i) For any $d \geq 0$, and regardless of whether $\delta = 0$ or $\delta \neq 0$,

$$t_{\tau_0} \xrightarrow{d} \frac{\frac{1}{2}(Z(1;\tau_0,c,\bar{c},\eta)^2 - 1)}{\left(\int_0^1 Z(s;\tau_0,c,\eta)^2 ds\right)^{1/2}} := \xi(\tau_0,c,\bar{c},\eta) \tag{S.1}$$

where

$$Z(s;\tau,c,\bar{c},\eta) := B_\eta^c(s) - X(s;\tau)'\left(\int_0^1 X_{\bar{c}}(s;\tau)X_{\bar{c}}(s;\tau)'ds\right)^{-1}\int_0^1 X_{\bar{c}}(s;\tau)dB_\eta^c(s;\bar{c})$$

and

$$B_\eta^c(s) := \int_0^s \exp(c(s-r))dB_\eta(r), \quad B_\eta^c(s;\bar{c}) := B_\eta^c(s) - \bar{c}\int_0^s B_\eta^c(r)dr,$$

with $B_\eta(\cdot)$ as defined in Theorem 2, and

$$X(s;\tau) := \begin{pmatrix} s \\ (s-\tau)\vee 0 \end{pmatrix}, \quad X_{\bar{c}}(s;\tau) := \begin{pmatrix} 1-\bar{c}s \\ 1-\bar{c}((s-\tau)\vee 0) \end{pmatrix}.$$

(ii) For $0 \leq d < 1/2$, and provided $\delta \neq 0$, it holds that: (a) $t_{OLS} - t_{\tau_0} \xrightarrow{p} 0$, and (b) provided the additional conditions of Theorem 3 hold, $t_{FWLS} - t_{\tau_0} \xrightarrow{p} 0$.

Remark S.1. The results in part (ii) of Theorem S.1 might appear to contradict with Proposition 3 of Kim and Perron (2009,p.12) where it is shown that for some generic break fraction estimator, $\tilde{\tau}$, the break fraction, τ_0 must be consistently estimated at some rate greater than $T^{1/2}$ in order for a DF test based on $\tilde{\tau}$, $t_{\tilde{\tau}}$ say, and t_{τ_0} to be asymptotically equivalent. However, the result in Kim and Perron (2009) relates only to the case where the trend break magnitude δ_T is fixed and non-zero (see their Assumption 1 on page 3), and therefore corresponds to the specific case of $d = 0$ and $\delta \neq 0$ in Theorem S.1. In this case we know from Theorem 1 that both $\hat{\tau}_{OLS}$ and $\hat{\tau}_{FWLS}$ are consistent at rate $O_p(T^{-1})$, which certainly satisfies the condition in Proposition 3 of Kim and Perron (2009, p.4). In the more general set-up we consider here, the trend break magnitude and convergence rate of $\hat{\tau}_{OLS}$ and $\hat{\tau}_{FWLS}$ change together; as the break magnitude slows, so commensurately does the convergence rate of $\hat{\tau}_{OLS}$, $\hat{\tau}_{FWLS}$. In particular, where the trend break magnitude is of order T^{-d}, $d \geq 0$, then, as shown in Theorems 1 and 3, respectively, $(\hat{\tau}_{OLS} - \tau_0)$ and $(\hat{\tau}_{FWLS} - \tau_0)$ are both of $O_p(T^{2d-1})$. However, this rate of consistency is still sufficiently fast for the asymptotic equivalence results in part (ii) of Theorem S.1 to hold, precisely because the magnitude of the trend break is shrinking commensurately with the reduced consistency rate. □

Remark S.2. The result in Theorem S.1 relates to the "large" break case of section 3.2 where $0 \leq d < 1/2$ in the localisation of the trend break magnitude, such that the trend break location τ_0 can be consistently estimated. Localisations which converge to zero at a faster rate, as considered in section 3.3, including the Pitman drift rate where $d = 1/2$, are excluded. Our aim here is not to provide a comprehensive treatment of the large sample properties of unit root tests in the present setting but rather to explore how weighted trend break estimators can improve the finite sample properties of unit root tests relative to standard OLS estimation. However, the results could be extended to cover the case of $d \geq 1/2$. For $d = 1/2$, results comparable to those given in section 5 of Harvey *et al.* (2012), but generalised by the non-stationary volatility allowed for under Assumption $\mathcal{A}_2$, would be obtained. For $d > 1/2$, as discussed in Case 1 in section 3.3, the magnitude of the trend break would be such that it would lead to trend break estimators which behave asymptotically the same as in the no break case. As a result, both t_{OLS} and t_{FWLS} would converge to limiting distributions of a similar form to that given in (S.1) evaluated not at τ_0 but at the random outcomes of $\hat{\tau}_{OLS}$ and $\hat{\tau}_{FWLS}$ within the search set $[\tau_L, \tau_U]$. □

Remark S.3. Theorem S.1 replicates the result given for the $t(\bar{\tau})$ statistic (which is based around the OLS break fraction estimator, $\hat{\tau}_{OLS}$) in part (ii) of Theorem 1 of Cavaliere *et al.* (2011, p.966) which pertains to the case of a trend break of fixed magnitude, $\delta \neq 0$ and $d = 0$. Theorem S.1 shows that the result also holds for t_{FWLS}, the DF statistic based on the feasible WLS break fraction estimator, and that it continues to hold (for both break fraction estimators) for breaks which are local to zero, provided the localisation rate is smaller than the Pitman rate, $d = 1/2$. □

Remark S.4. The (common) limiting null distribution of t_{OLS} and t_{FWLS}, which obtains on setting $c = 0$ in (S.1), is seen to depend on the volatility process $\sigma(\cdot)$ through the presence of the heteroskedastic Brownian motion $B_\eta(\cdot)$. This is also the case for the corresponding no-break statistic, t_0, defined in Remark 4.3, whose limiting distribution under H_c coincides with that given for the $\mathcal{MZ}_t$ statistic in Theorem 1 of Cavaliere and Taylor (2008a, pp.49-50). Consequently for pivotal inference on the unit root null hypothesis, H_0, we will need to base these tests on either the simulated critical value approach outlined in section 4.2 of Cavaliere and Taylor (2007) or a wild bootstrap approach, the latter outlined for the t_0 statistic in section 4.1 of Cavaliere and Taylor (2008a), and for the trend break case in Algorithm 1 of Cavaliere *et al.* (2011, p.971). In the no trend break case, $\delta = 0$, the asymptotic validity of the simulated critical value and wild bootstrap methods is established in Cavaliere and Taylor (2007) and Cavaliere and Taylor (2008a), respectively. In the case where a trend break occurs, $\delta \neq 0$, both approaches can be shown to deliver asymptotically pivotal inference for t_{OLS} and t_{FWLS} under the conditions of Theorem S.1. For the wild bootstrap approach, the proof of asymptotic validity follows directly from Cavaliere *et al.* (2011), noting the asymptotic equivalence of t_{OLS} and t_{FWLS}. The asymptotic validity of the simulated critical value approach follows using the same arguments as are given in section 4 of Cavaliere and Taylor (2007). □

Remark S.5. The heteroskedastic Brownian motion B_η appearing in Theorem S.1 is the same as that in Theorem 2. The results in Theorem 2 directly provide the asymptotic distribution theory for break fraction estimators under the unit root null hypothesis. Under the near-integrated local alternatives that we also allow for in Theorem S.1, the asymptotic distribution theory for the break fraction estimators can, in the usual way, be shown to obtain by replacing the heteroskedastic Brownian motion B_η in Theorem 2 with the corresponding heteroskedastic near-integrated process B_η^c defined in Theorem S.1. The general conclusions given in section 3.3 about the properties of the break fraction estimators for "small" breaks under the unit root null therefore also apply qualitatively unchanged under local alternatives. The conclusion of most practical relevance from this is that the break fraction estimator will tend to be drawn towards periods of high variance in the time series, regardless of whether the unit root null hypothesis or the near-integrated alternative hypothesis holds. $\square$

In Theorem S.2 we now establish the large sample properties of the weighted and unweighted SC-based procedures.

Theorem S.2. *Let y_t be generated according to* (4.1)-(4.2) *with e_t generated according to* (2.2)*, and let the conditions of Assumption $\mathcal{A}$ hold. Let $\delta_T = \delta T^{-d}$, $0 \leq d < 1/2$. Then, under H_c, and in each case as $T \to \infty$:*

(a) For the unweighted SC-based procedure:
(i) if $\delta \neq 0$, then $\Pr(SC_{\hat{\tau}_{OLS}} \leq SC_0) \to 1$; (ii) if $\delta = 0$, then $\Pr(SC_{\hat{\tau}_{OLS}} > SC_0) \to 1$.

(b) For the weighted SC-based procedure, and provided the additional conditions of Theorem 3 hold:
(i) if $\delta \neq 0$, $\Pr(SC^*_{\hat{\tau}_{FWLS}} \leq SC^*_0) \to 1$; (ii) if $\delta = 0$, $\Pr(SC^*_{\hat{\tau}_{FWLS}} > SC^*_0) \to 1$.

Remark S.6. Theorem S.2 shows that for both the unweighted and weighted SC procedures the probability of selecting the trend break model converges to one (zero) when a trend break is (is not) present in the DGP, in each case as the sample size diverges. As a result, as the sample size diverges, both t_{SC} and t_{WSC} converge (in probability) to t_0 when no trend break is present, while when a trend break is present t_{SC} converges to t_{OLS} and t_{WSC} converges to t_{FWLS}, both of which coincide with the known τ_0 limiting distribution in (S.1). Consequently, the tests from both SC procedures will be asymptotically correctly sized when using the appropriate asymptotic critical value, obtained as outlined in section 4 using either the simulated critical value approach of Cavaliere and Taylor (2007) or the wild bootstrap approach in Algorithm 1 of Cavaliere *et al.* (2011), regardless of whether a trend break occurs or not. Moreover, the asymptotic local power of the SC tests will be identical to that of the (size-adjusted) infeasible test which assumes knowledge of whether a break has occurred or not, together with knowledge of the true break fraction, τ_0, in the former case. $\square$

Remark S.7. Observe that the unweighted SC decision rule can be equivalently expressed in terms of the (pseudo) likelihood ratio test based decision rule to include the trend break if $T(\log(s_0^2) - \log(s^2_{\hat{\tau}_{OLS}})) \geq 3\log T$ and, similarly, for the weighted SC rule if $T(\log(s_0^{*2}) - \log(s^{*2}_{\hat{\tau}_{FWLS}})) \geq 3\log T$. This is therefore seen to be analogous to testing for the presence of a trend break at the random fraction $\hat{\tau}_{OLS}$ and $\hat{\tau}_{FWLS}$ for the unweighted and weighted SC rules, respectively, and as such is related to a sup-LR type statistic in the spirit of Andrews (1993), but where the decision rule is based not on a fixed critical value but on a Schwarz-type penalty. These are then essentially pre-tests for the presence of a trend break both of which, by design, have size which shrinks to zero as the sample size diverges; the same requirement is needed on the trend break pre-tests used in the univariate testing analogue of the problem considered here in Harris *et al.* (2009) and Carrion-i-Silvestre *et al.* (2009). Because the weighted SC approach corresponds to a test based on feasible WLS estimation, is it anticipated that it will be more efficacious than the unweighted SC approach in selecting between the trend break and no trend break models in the presence of non-stationary volatility. □

S.3 Time Series Plots of U.S. and U.K. real GDP Series

These graphs appear on the next page of the supplement.

S.4 Proofs

This section contains proofs for the theorems stated in the main text. Section S.4.1 contains some preliminary results. Proofs of the results in Theorems 1-3 and Theorems S.1 and S.2 are then provided in Section S.4.2.

S.4.1 Preliminary Results

The following Lemmas are useful for the proofs of Theorems 1 and 2. The first provides a general representation of the weighted least squares criterion, and the second the heteroskedastic FCLT results required for the theorems.

The estimator of τ_0 defined in (3.1) minimises the sum of squared residuals from regressions (3.2) estimated over the range of τ, which can equivalently be expressed as regressions of $y_t^* := y_t x_t$ on $x_t 1_{t\le\lfloor\tau T\rfloor}$ and $x_t 1_{t>\lfloor\tau T\rfloor}$ for $\tau \in [\tau_L, \tau_U]$. For any τ such a regression can be represented

$$y_t^* = \hat{\mu}_{1,\tau}(x_t 1_{t\le\lfloor\tau T\rfloor}) + \hat{\mu}_{2,\tau}(x_t 1_{t>\lfloor\tau T\rfloor}) + \hat{e}^*_{\tau,t},$$

where

$$\hat{\mu}_{1,\tau} := \frac{\sum_{t=1}^{\lfloor\tau T\rfloor} y_t^* x_t}{\sum_{t=1}^{\lfloor\tau T\rfloor} x_t^2}, \quad \hat{\mu}_{2,\tau} := \frac{\sum_{t=\lfloor\tau T\rfloor+1}^{T} y_t^* x_t}{\sum_{t=\lfloor\tau T\rfloor+1}^{T} x_t^2}$$

are the OLS coefficient estimates.

Lemma S.1. *For any weights x_t, $\hat{\tau}$ defined in* (3.1) *can be represented as*

$$\hat{\tau} = \arg\max_{\tau\in[\tau_L,\tau_U]} Q_T(\tau),$$

where

$$Q_T(\tau) := \frac{\sum_{t=1}^{\lfloor\tau T\rfloor} x_t^2 \sum_{t=\lfloor\tau T\rfloor+1}^{T} x_t^2}{\sum_{t=1}^{T} x_t^2} \left(\hat{\mu}_{2,\tau} - \hat{\mu}_{1,\tau}\right)^2,$$

and where

$$\begin{aligned}\hat{\mu}_{2,\tau} - \hat{\mu}_{1,\tau} = {}& \delta_T \left(\frac{\sum_{t=1}^{\lfloor\tau_0 T\rfloor} x_t^2}{\sum_{t=1}^{\lfloor\tau T\rfloor} x_t^2} \wedge \frac{\sum_{t=\lfloor\tau_0 T\rfloor+1}^{T} x_t^2}{\sum_{t=\lfloor\tau T\rfloor+1}^{T} x_t^2} \right) \\ & - \frac{\sum_{t=1}^{T} x_t^2}{\sum_{t=1}^{\lfloor\tau T\rfloor} x_t^2 \sum_{t=\lfloor\tau T\rfloor+1}^{T} x_t^2} \left(\sum_{t=1}^{\lfloor\tau T\rfloor} e_t x_t^2 - \frac{\sum_{t=1}^{\lfloor\tau T\rfloor} x_t^2}{\sum_{t=1}^{T} x_t^2} \sum_{t=1}^{T} e_t x_t^2 \right). \qquad \text{(S.2)}\end{aligned}$$

Figure S.2: (a) Log U.S. real GDP

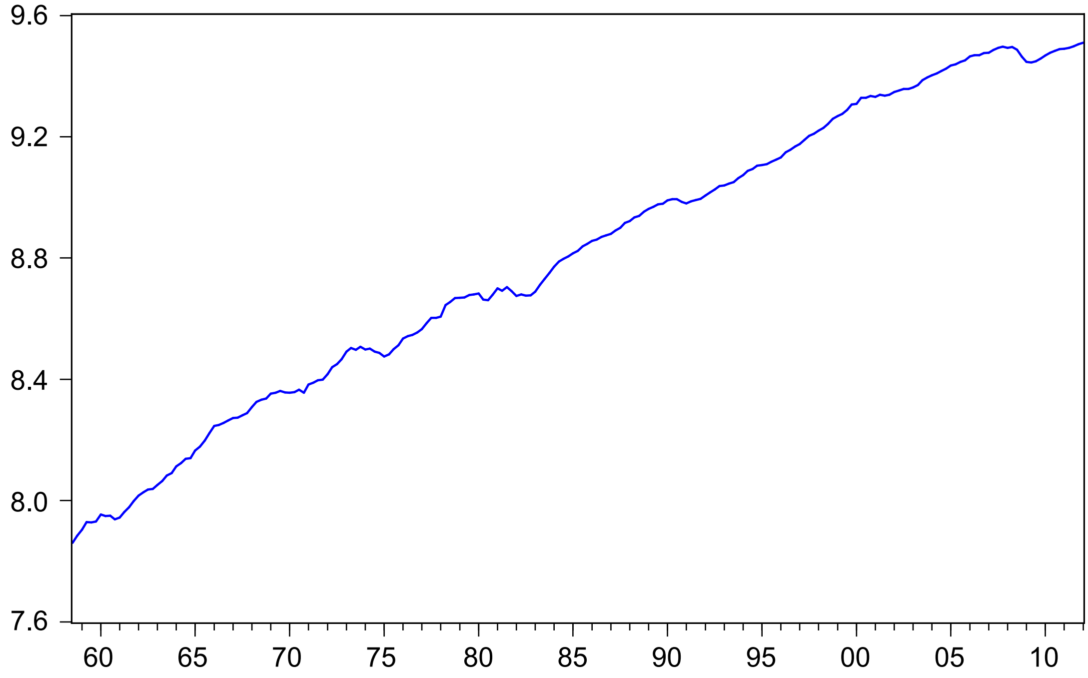
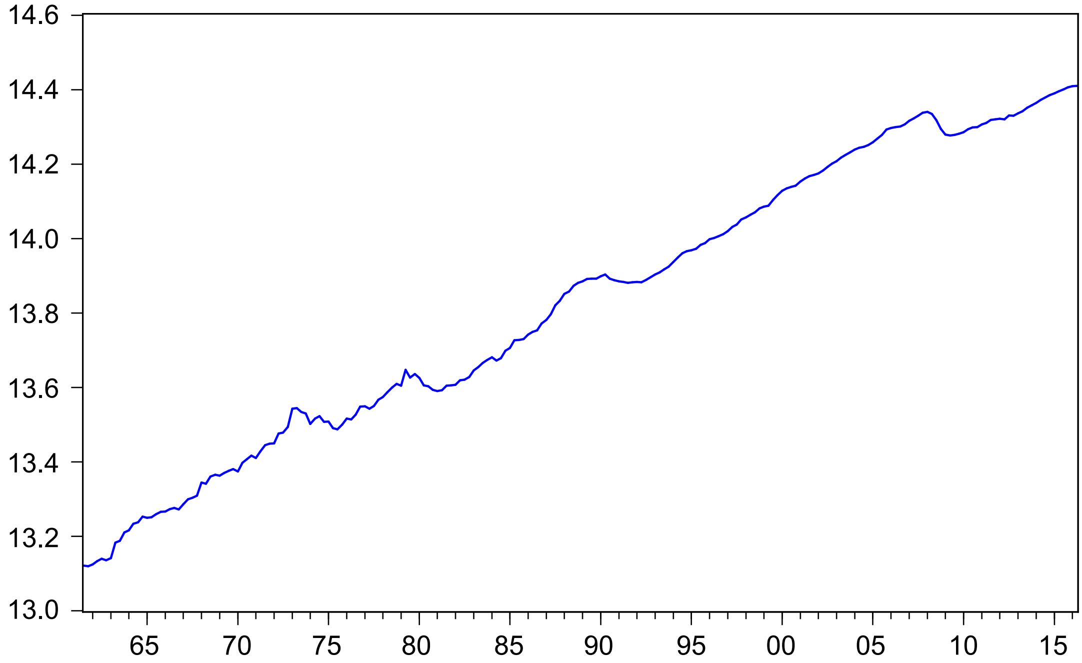


Figure S.2: (b) Log U.K. real GDP



Proof of Lemma S.1

For any τ, standard least squares algebra gives the sum of squared residuals

$$\sum_{t=1}^{T} \hat{e}_{\tau,t}^{*2} = \sum_{t=1}^{T} y_t^{*2} - \left(\hat{\mu}_{1,\tau}^2 \sum_{t=1}^{\lfloor \tau T \rfloor} x_t^2 + \hat{\mu}_{2,\tau}^2 \sum_{t=\lfloor \tau T \rfloor +1}^{T} x_t^2 \right). \tag{S.3}$$

Similar algebra for a regression of y_t^* on x_t alone gives the sum of squared residuals

$$\sum_{t=1}^{T} \hat{e}_t^{*2} = \sum_{t=1}^{T} y_t^{*2} - \hat{\mu}^2 \sum_{t=1}^{T} x_t^2 \tag{S.4}$$

where $\hat{\mu} := T^{-1} \sum_{t=1}^{T} x_t$. Subtracting (S.4) from (S.3) gives

$$\sum_{t=1}^{T} \hat{e}_{\tau,t}^{*2} = \sum_{t=1}^{T} \hat{e}_t^{*2} + \frac{(\sum_{t=1}^{T} x_t y_t^*)^2}{\sum_{t=1}^{T} x_t^2} - \frac{(\sum_{t=1}^{\lfloor \tau T \rfloor} x_t y_t^*)^2}{\sum_{t=1}^{\lfloor \tau T \rfloor} x_t^2} - \frac{(\sum_{t=\lfloor \tau T \rfloor+1}^{T} x_t y_t^*)^2}{\sum_{t=\lfloor \tau T \rfloor+1}^{T} x_t^2}.$$

Substituting $\sum_{t=1}^{T} x_t y_t^* = \sum_{t=1}^{\lfloor \tau T \rfloor} x_t y_t^* + \sum_{t=\lfloor \tau T \rfloor+1}^{T} x_t y_t^*$, expanding and rearranging gives

$$\begin{aligned} \sum_{t=1}^{T} \hat{e}_{\tau,t}^{*2} &= \sum_{t=1}^{T} \hat{e}_t^{*2} - \frac{\sum_{t=1}^{\lfloor \tau T \rfloor} x_t^2 \sum_{t=\lfloor \tau T \rfloor+1}^{T} x_t^2}{\sum_{t=1}^{T} x_t^2} (\hat{\mu}_{2,\tau} - \hat{\mu}_{1,\tau})^2 \\ &= \sum_{t=1}^{T} \hat{e}_t^{*2} - Q_T(\tau), \end{aligned}$$

so that $\arg\min_\tau \sum_{t=1}^{T} \hat{e}_{\tau,t}^{*2}$ is identical to $\arg\max_\tau Q_T(\tau)$.

Next consider $\hat{\mu}_{2,\tau} - \hat{\mu}_{1,\tau}$. Using

$$y_t^* = \mu x_t + \delta_T (x_t \cdot 1_{t \geq \lfloor \tau_0 T \rfloor}) + x_t e_t$$

in the expressions for $\hat{\mu}_{1,\tau}$ and $\hat{\mu}_{2,\tau}$ gives

$$\hat{\mu}_{1,\tau} = \mu + \delta_T \frac{\sum_{t=\lfloor \tau_0 T \rfloor+1}^{\lfloor \tau T \rfloor} x_t^2}{\sum_{t=1}^{\lfloor \tau T \rfloor} x_t^2} + \frac{\sum_{t=1}^{\lfloor \tau T \rfloor} x_t^2 e_t}{\sum_{t=1}^{\lfloor \tau T \rfloor} x_t^2}$$

(the second term being zero for $\lfloor \tau T \rfloor \leq \lfloor \tau_0 T \rfloor$) and

$$\hat{\mu}_{2,\tau} = \mu + \delta_T \frac{\sum_{t=\lfloor (\tau \vee \tau_0) T \rfloor+1}^{T} x_t^2}{\sum_{t=\lfloor \tau T \rfloor+1}^{T} x_t^2} + \frac{\sum_{t=\lfloor \tau T \rfloor+1}^{T} x_t^2 e_t}{\sum_{t=\lfloor \tau T \rfloor+1}^{T} x_t^2}.$$

To calculate $\hat{\mu}_{2,\tau} - \hat{\mu}_{1,\tau}$, note that

$$\begin{aligned} \frac{\sum_{t=\lfloor (\tau \vee \tau_0) T \rfloor+1}^{T} x_t^2}{\sum_{t=\lfloor \tau T \rfloor+1}^{T} x_t^2} - \frac{\sum_{t=\lfloor \tau_0 T \rfloor+1}^{\lfloor \tau T \rfloor} x_t^2}{\sum_{t=1}^{\lfloor \tau T \rfloor} x_t^2} &= \begin{cases} \dfrac{\sum_{t=\lfloor \tau_0 T \rfloor+1}^{T} x_t^2}{\sum_{t=\lfloor \tau T \rfloor+1}^{T} x_t^2}, & \text{if } \lfloor \tau T \rfloor \leq \lfloor \tau_0 T \rfloor \\[2ex] \dfrac{\sum_{t=1}^{\lfloor \tau_0 T \rfloor} x_t^2}{\sum_{t=1}^{\lfloor \tau T \rfloor} x_t^2}, & \text{if } \lfloor \tau T \rfloor > \lfloor \tau_0 T \rfloor \end{cases} \\ &= \frac{\sum_{t=1}^{\lfloor \tau_0 T \rfloor} x_t^2}{\sum_{t=1}^{\lfloor \tau T \rfloor} x_t^2} \wedge \frac{\sum_{t=\lfloor \tau_0 T \rfloor+1}^{T} x_t^2}{\sum_{t=\lfloor \tau T \rfloor+1}^{T} x_t^2}, \end{aligned}$$

and that straightforward rearrangements give

$$\frac{\sum_{t=\lfloor\tau T\rfloor+1}^{T} x_t^2 e_t}{\sum_{t=\lfloor\tau T\rfloor+1}^{T} x_t^2} - \frac{\sum_{t=1}^{\lfloor\tau T\rfloor} x_t^2 e_t}{\sum_{t=1}^{\lfloor\tau T\rfloor} x_t^2} = \frac{\sum_{t=1}^{T} x_t^2}{\sum_{t=1}^{\lfloor\tau T\rfloor} x_t^2 \sum_{t=\lfloor\tau T\rfloor+1}^{T} x_t^2} \left(\sum_{t=1}^{\lfloor\tau T\rfloor} e_t x_t^2 - \frac{\sum_{t=1}^{\lfloor\tau T\rfloor} x_t^2}{\sum_{t=1}^{T} x_t^2} \sum_{t=1}^{T} e_t x_t^2 \right),$$

and together these imply the result in (S.2). ■

Lemma S.2. *Let e_t be generated by (2.2) under the conditions of Assumption $\mathcal{A}$, and let $x_t = x(t/T)$ be non-stochastic weights such that the function $x(\cdot)$ satisfies the same conditions as $\sigma(\cdot)$ in Assumption $\mathcal{A}_2$. Then,*

$$\omega_{xe}^{-1} T^{-1/2} \sum_{t=1}^{\lfloor \cdot T\rfloor} x_t^2 e_t \xrightarrow{d} B_\eta(\cdot) \tag{S.5}$$

where $B_\eta(s) := B(\eta(s))$, with $B(\cdot)$ a standard Brownian motion, and where $\eta(s) := \omega_{xe}^{-2} \int_0^s x(r)^4 \sigma(r)^2 dr$ with $\omega_{xe}^2 := \int_0^1 x(r)^4 \sigma(r)^2 dr$. Moreover,

$$\sup_{\tau\in[\tau_L,\tau_U]} T^{-1/2} \sum_{t=1}^{\lfloor\tau T\rfloor} e_t x_t^2 = O_p(1) \tag{S.6}$$

Proof of Lemma S.2

Equation (S.5) follows directly from Lemma 4 of Cavaliere and Taylor (2007). The maximal inequality in equation (S.6) follows by noting that

$$\left| \sup_{\tau\in[\tau_L,\tau_U]} T^{-1/2} \sum_{t=1}^{\lfloor\tau T\rfloor} e_t x_t^2 \right| \le \max_{b\in[1,T]} T^{-1/2} \left| \sum_{t=1}^{b} e_t x_t^2 \right|,$$

and then, since $\sum_{t=1}^{b} e_t x_t^2$ is a martingale indexed by b, Doob's inequality (result 15.15 of Davidson, 1994) implies that

$$E\left(\max_{b\in[1,T]} \left(T^{-1/2} \left| \sum_{t=1}^{b} e_t x_t^2 \right| \right)^2 \right) \le 4T^{-1} E\left(\sum_{t=1}^{T} e_t x_t^2 \right)^2 = 4T^{-1} \sum_{t=1}^{T} \sigma_t^2 x_t^4 = O(1).$$

■

S.4.2 Main Proofs

Proof of Theorem 1

The argument for consistency follows Bai (1994, 1997) and especially Theorem 3.1 of Nunes *et al.* (1995). The assumptions on the weights x_t imply that

$$\frac{\sum_{t=1}^{\lfloor\tau T\rfloor} x_t^2}{\sum_{t=1}^{T} x_t^2} \to \frac{\int_0^\tau x(s)^2 ds}{\int_0^1 x(s)^2 ds} =: \chi(\tau),$$

with $\chi : [0,1] \mapsto [0,1]$ being an increasing function (exactly analogous to the variance-profile η constructed from the variance sequence σ_t^2). Thus, for example,

$$\frac{\sum_{t=1}^{\lfloor\tau T\rfloor} x_t^2}{\sum_{t=1}^{T} x_t^2} \frac{\sum_{t=\lfloor\tau T\rfloor+1}^{T} x_t^2}{\sum_{t=1}^{T} x_t^2} \to \chi(\tau)(1-\chi(\tau)).$$

and

$$\frac{\sum_{t=1}^{\lfloor \tau_0 T \rfloor} x_t^2}{\sum_{t=1}^{\lfloor \tau T \rfloor} x_t^2} \wedge \frac{\sum_{t=\lfloor \tau_0 T \rfloor+1}^{T} x_t^2}{\sum_{t=\lfloor \tau T \rfloor+1}^{T} x_t^2} \to \frac{\chi(\tau_0)}{\chi(\tau)} \wedge \frac{1-\chi(\tau_0)}{1-\chi(\tau)}.$$

Applying (S.2), it follows that for $\delta \neq 0$ and $0 \leq d < 1/2$

$$\begin{aligned} T^d(\hat{\mu}_{2,\tau} - \hat{\mu}_{1,\tau}) &= \delta \left(\frac{\sum_{t=1}^{\lfloor \tau_0 T \rfloor} x_t^2}{\sum_{t=1}^{\lfloor \tau T \rfloor} x_t^2} \wedge \frac{\sum_{t=\lfloor \tau_0 T \rfloor+1}^{T} x_t^2}{\sum_{t=\lfloor \tau T \rfloor+1}^{T} x_t^2} \right) \\ &\quad - T^{d-\frac{1}{2}} \frac{T \sum_{t=1}^{T} x_t^2}{\sum_{t=1}^{\lfloor \tau T \rfloor} x_t^2 \sum_{t=\lfloor \tau T \rfloor+1}^{T} x_t^2} \left(T^{-\frac{1}{2}} \sum_{t=1}^{\lfloor \tau T \rfloor} e_t x_t^2 - \frac{\sum_{t=1}^{\lfloor \tau T \rfloor} x_t^2}{\sum_{t=1}^{T} x_t^2} T^{-\frac{1}{2}} \sum_{t=1}^{T} e_t x_t^2 \right) \\ &\xrightarrow{p} \delta \left(\frac{\chi(\tau_0)}{\chi(\tau)} \wedge \frac{1-\chi(\tau_0)}{1-\chi(\tau)} \right), \end{aligned}$$

uniformly in τ on $[\tau_L, \tau_U]$, and, hence,

$$\frac{T^{2d}}{\sum_{t=1}^{T} x_t^2} Q_T(\tau) \xrightarrow{p} Q(\tau) := \delta^2 \chi(\tau)(1-\chi(\tau)) \left(\frac{\chi(\tau_0)}{\chi(\tau)} \wedge \frac{1-\chi(\tau_0)}{1-\chi(\tau)} \right)^2. \tag{S.7}$$

For $\tau \in [\tau_L, \tau_0]$

$$Q(\tau) = \delta^2 (1-\chi(\tau_0))^2 \frac{\chi(\tau)}{1-\chi(\tau)}$$

is an increasing function with maximum value at $\tau = \tau_0$ of $Q(\tau_0) = \chi(\tau_0)(1-\chi(\tau_0))$. For $\tau \in [\tau_0, \tau_U]$

$$Q(\tau) = \delta^2 \chi(\tau_0)^2 \frac{1-\chi(\tau)}{\chi(\tau)}$$

is a decreasing function with maximum value of $Q(\tau_0)$ at $\tau = \tau_0$. So $Q(\tau)$ has a unique maximum at $\tau = \tau_0$ and, since the convergence in (S.7) is uniform in τ, it therefore follows that

$$\hat{\tau} = \arg \max_{\tau \in [\tau_L, \tau_U]} \frac{T^{2d}}{\sum_{t=1}^{T} x_t^2} Q_T(\tau) \xrightarrow{p} \arg \max_{\tau \in [\tau_L, \tau_U]} Q(\tau) = \tau_0.$$

The asymptotic distribution of $\hat{\tau}$ follows by the same arguments as Bai (1997), except that it is necessary to check for any effect that unconditional heteroskedasticity satisfying Assumption $\mathcal{A}_2$ has on the form of the distribution. In particular, using Bai's notation in the proof of his Proposition 3, we will verify that the following two convergence results from Bai (1997) still hold under our assumptions:

(a) $v_T \sum_{t=k_0+1}^{k} z_t \epsilon_t \xrightarrow{d} B_1(s)$ and
(b) $v_T^2 \sum_{t=k_0+1}^{k} z_t z_t' \to s Q_2$,

where $B_1(s)$ is Brownian motion on $[0, \infty)$ with variance $s\Omega_2$, and k is defined as a function of s to be $k := k_0 + \lfloor s v_T^{-2} \rfloor$. Note that both limits, being simple functions of s, reflect the unconditional homoskedasticity assumed in Bai (1997). Under the form of heteroskedasticity assumed in this paper, it might have been expected that the corresponding B_1 would be a heteroskedastic Brownian Motion (as in, for example, Cavaliere and Taylor, 2007), but we will show that this is not the case. To translate the notation of Bai (1997) notation into our own, his z_t (the regressor(s) whose coefficient(s) break) is the same as our x_t, $\epsilon_t = x_t e_t$, $v_T = \delta_T$ and $k = \lfloor \tau_0 T \rfloor + \lfloor s \delta_T^{-2} \rfloor$. This specification of k is relevant for the scaled criterion function for values of τ above the true value τ_0. In our equivalent to (a), as $T \to \infty$ the term

$$\delta_T \sum_{t=\lfloor \tau_0 T \rfloor+1}^{\lfloor \tau_0 T \rfloor + \lfloor s\delta_T^{-2} \rfloor} x_t^2 e_t$$

can straightforwardly be shown to satisfy the heteroskedastic FCLT in Lemma 4 of Cavaliere and Taylor (2007), indexed by s, but as in their proof we need to explicitly derive the form of the marginal distribution for given s. In particular the variance is

$$\begin{aligned}
&\operatorname{var}\left(\delta_T \sum_{t=\lfloor \tau_0 T \rfloor+1}^{\lfloor \tau_0 T \rfloor + \lfloor s\delta_T^{-2} \rfloor} x_t^2 e_t\right)\\
&= \delta_T^2 \sum_{i=1}^{\lfloor s\delta_T^{-2} \rfloor} x^4_{\lfloor \tau_0 T \rfloor + i} \sigma^2_{\lfloor \tau_0 T \rfloor + i}\\
&= s \sum_{i=1}^{\lfloor s\delta_T^{-2} \rfloor} \int_{i/\lfloor s\delta_T^{-2} \rfloor}^{(i+1)/\lfloor s\delta_T^{-2} \rfloor} x\left(\frac{\lfloor \tau_0 T \rfloor + \lfloor r \lfloor s\delta_T^{-2} \rfloor \rfloor}{T}\right)^4 \sigma\left(\frac{\lfloor \tau_0 T \rfloor + \lfloor r \lfloor s\delta_T^{-2} \rfloor \rfloor}{T}\right)^2 dr\\
&= s \int_{1/\lfloor s\delta_T^{-2} \rfloor}^{1+1/\lfloor s\delta_T^{-2} \rfloor} x\left(\frac{\lfloor \tau_0 T \rfloor + \lfloor r \lfloor s\delta_T^{-2} \rfloor \rfloor}{T}\right)^4 \sigma\left(\frac{\lfloor \tau_0 T \rfloor + \lfloor r \lfloor s\delta_T^{-2} \rfloor \rfloor}{T}\right)^2 dr\\
&\to s \int_0^1 \bar{x}(\tau_0)^4 \bar{\sigma}(\tau_0)^2 dr = s\, \bar{x}(\tau_0)^4 \bar{\sigma}(\tau_0)^2,
\end{aligned}$$

with the convergence in the second last step following because $\delta_T^{-2}/T \to 0$ for $0 < d < 1/2$ and from the assumed càdlàg property of both the $x(\cdot)$ and $\sigma(\cdot)$ functions. The limits from above $\bar{x}(\tau_0) := \lim_{\tau \downarrow \tau_0} x(\tau)$ and $\bar{\sigma}(\tau_0) := \lim_{\tau \downarrow \tau_0} \sigma(\tau)$ may differ from $x(\tau_0)$ and $\sigma(\tau_0)$ if either function has a jump at $\tau = \tau_0$. Similarly for our equivalent to (b) above,

$$\delta_T^2 \sum_{t=\lfloor \tau_0 T \rfloor+1}^{\lfloor \tau_0 T \rfloor + \lfloor s\delta_T^{-2} \rfloor} x_t^2 \to s\, \bar{x}(\tau_0)^2.$$

Similarly for values of τ below τ_0 we find

$$\operatorname{var}\left(\delta_T \sum_{t=\lfloor \tau_0 T \rfloor - \lfloor s\delta_T^{-2} \rfloor}^{\lfloor \tau_0 T \rfloor - 1} x_t^2 e_t\right) \to s\, \underline{x}(\tau_0)^4 \underline{\sigma}(\tau_0)^2$$

and

$$\delta_T^2 \sum_{t=\lfloor \tau_0 T \rfloor - \lfloor s\delta_T^{-2} \rfloor}^{\lfloor \tau_0 T \rfloor - 1} x_t^2 \to s\, \underline{x}(\tau_0)^2.$$

where $\underline{x}(\tau_0) := \lim_{\tau \uparrow \tau_0} x(\tau)$ and $\underline{\sigma}(\tau_0) := \lim_{\tau \uparrow \tau_0} \sigma(\tau)$.

The rate of δ_T^{-2} when $0 < d < 1/2$, being less than T, is such that the effect of the weighting function $x(\tau)$ and the volatility process $\sigma(\tau)$ drop out of these limits for all values of τ other than τ_0. For a "large" break magnitude (one for which the break fraction can be consistently estimated) the asymptotic distribution of the break fraction is determined by the behaviour of the criterion function in a diminishing interval around τ_0, such that in the limit it is only the variance properties at τ_0 that enter the FCLT results.

With these expressions in hand, the proof follows each step of Bai's (1997) Proposition 3 to reach the analogous result in this theorem. The equivalences in the notation are Bai's Q_1 and Q_2 with $\underline{x}(\tau_0)^2$ and $\bar{x}(\tau_0)^2$ here, and Bai's Ω_1 and Ω_2 with $\lim_{\tau\uparrow\tau_0} \operatorname{var}(x_t^2 e_t) = \underline{x}(\tau_0)^4\underline{\sigma}(\tau_0)^2$ and $\lim_{\tau\downarrow\tau_0} \operatorname{var}(x_t^2 e_t) = \bar{x}(\tau_0)^4\bar{\sigma}(\tau_0)^2$, respectively. ■

Proof of Theorem 2

For $\delta \neq 0$ and $d = 1/2$, the representation in (S.2) can be written

$$\begin{aligned}
&T^{\frac{1}{2}}(\hat{\mu}_{2,\tau} - \hat{\mu}_{1,\tau}) \\
&= \delta\left(\frac{\sum_{t=1}^{\lfloor\tau_0 T\rfloor} x_t^2}{\sum_{t=1}^{\lfloor\tau T\rfloor} x_t^2} \wedge \frac{\sum_{t=\lfloor\tau_0 T\rfloor+1}^{T} x_t^2}{\sum_{t=\lfloor\tau T\rfloor+1}^{T} x_t^2}\right) \\
&\quad - \frac{\omega_{xe}}{T^{-1}\sum_{t=1}^{T} x_t^2} \frac{\sum_{t=1}^{T} x_t^2}{\sum_{t=1}^{\lfloor\tau T\rfloor} x_t^2} \frac{\sum_{t=1}^{T} x_t^2}{\sum_{t=\lfloor\tau T\rfloor+1}^{T} x_t^2}\left(\omega_{xe}^{-1}T^{-\frac{1}{2}}\sum_{t=1}^{\lfloor\tau T\rfloor} e_t x_t^2 - \frac{\sum_{t=1}^{\lfloor\tau T\rfloor} x_t^2}{\sum_{t=1}^{T} x_t^2}\omega_{xe}^{-1}T^{-\frac{1}{2}}\sum_{t=1}^{T} e_t x_t^2\right) \\
&\xrightarrow{d} \delta\left(\frac{\chi(\tau_0)}{\chi(\tau)} \wedge \frac{1-\chi(\tau_0)}{1-\chi(\tau)}\right) - \omega\frac{B_\eta(\tau) - \chi(\tau)B_\eta(1)}{\chi(\tau)(1-\chi(\tau))} \\
&= \frac{\omega}{(\chi(\tau)(1-\chi(\tau)))^{1/2}}\left(\frac{\delta}{\omega}(\chi(\tau_0)(1-\chi(\tau_0)))^{\frac{1}{2}}\left(\chi_1(\tau;\tau_0) \wedge \frac{1}{\chi_1(\tau;\tau_0)}\right) - \frac{B_\eta(\tau) - \chi(\tau)B_\eta(1)}{\chi(\tau)(1-\chi(\tau))}\right)
\end{aligned}$$

where

$$\omega^2 = \frac{\omega_{xe}^2}{\left(\int_0^1 x(s)^2 ds\right)^2} = \frac{\int_0^1 x(s)^4\sigma(s)^2 ds}{\left(\int_0^1 x(s)^2 ds\right)^2}.$$

Hence,

$$\begin{aligned}
\frac{T}{\omega^2\sum_{t=1}^{T} x_t^2} Q_T(\tau) &= \frac{\sum_{t=1}^{\lfloor\tau T\rfloor} x_t^2}{\omega\sum_{t=1}^{T} x_t^2} \frac{\sum_{t=\lfloor\tau T\rfloor+1}^{T} x_t^2}{\sum_{t=1}^{T} x_t^2} T\left(\hat{\mu}_{2,\tau} - \hat{\mu}_{1,\tau}\right)^2 \\
&\xrightarrow{d} \left(\frac{\delta}{\omega}(\chi(\tau_0)(1-\chi(\tau_0)))^{\frac{1}{2}}\left(\chi_1(\tau;\tau_0) \wedge \frac{1}{\chi_1(\tau;\tau_0)}\right) - \frac{B_\eta(\tau) - \chi(\tau)B_\eta(1)}{\chi(\tau)(1-\chi(\tau))}\right)^2 \\
&=: Q(\tau; x(\cdot), \sigma(\cdot), \delta, \frac{1}{2}).
\end{aligned}$$

If $d > 1/2$ then the representation in (S.2) is

$$\begin{aligned}
&T^{\frac{1}{2}}(\hat{\mu}_{2,\tau} - \hat{\mu}_{1,\tau}) \\
&= T^{1/2-d}\delta\left(\frac{\sum_{t=1}^{\lfloor\tau_0 T\rfloor} x_t^2}{\sum_{t=1}^{\lfloor\tau T\rfloor} x_t^2} \wedge \frac{\sum_{t=\lfloor\tau_0 T\rfloor+1}^{T} x_t^2}{\sum_{t=\lfloor\tau T\rfloor+1}^{T} x_t^2}\right) \\
&\quad - \frac{\omega_{xe}}{T^{-1}\sum_{t=1}^{T} x_t^2} \frac{\sum_{t=1}^{T} x_t^2}{\sum_{t=1}^{\lfloor\tau T\rfloor} x_t^2} \frac{\sum_{t=1}^{T} x_t^2}{\sum_{t=\lfloor\tau T\rfloor+1}^{T} x_t^2}\left(\omega_{xe}^{-1}T^{-\frac{1}{2}}\sum_{t=1}^{\lfloor\tau T\rfloor} e_t x_t^2 - \frac{\sum_{t=1}^{\lfloor\tau T\rfloor} x_t^2}{\sum_{t=1}^{T} x_t^2}\omega_{xe}^{-1}T^{-\frac{1}{2}}\sum_{t=1}^{T} e_t x_t^2\right) \\
&\xrightarrow{d} -\omega\frac{B_\eta(\tau) - \chi(\tau)B_\eta(1)}{\chi(\tau)(1-\chi(\tau))},
\end{aligned}$$

so that

$$\hat{\tau} = \arg\max_{\tau\in[\tau_L,\tau_U]} \frac{T}{\omega^2\sum_{t=1}^{T} x_t^2} Q_T(\tau) \xrightarrow{d} \arg\max_{\tau\in[\tau_L,\tau_U]} \frac{(B_\eta(\tau) - \chi(\tau)B_\eta(1))^2}{(\chi(\tau)(1-\chi(\tau)))}$$

which is the arg max of $Q(\tau; x(\cdot), \sigma(\cdot), \delta, d)$ with $d > 1/2$ as required. Clearly this latter results also holds when $\delta = 0$ (in which case the value of d is irrelevant).

For $d < 1/2$ (and $\delta \neq 0$), the result follows immediately as an implication of equation (S.7) established in the proof of Theorem 1.

■

Proof of Theorem 3

For clarity we write $\hat{\sigma}^2_{\hat{\tau},t} := \sum_{i=1}^{T} w_{ti}\hat{e}^2_{\hat{\tau},i}$, where $w_{ti} := K_h\left(\frac{t-i}{T}\right) / \sum_{i=1}^{T} K_h\left(\frac{t-i}{T}\right)$, instead of $\hat{\sigma}^2_t$ and $Q_T(\tau; x_t)$ instead of $Q_T(\tau)$ defined in Lemma A.1 in the rest of the proof of this Theorem. We prove the result separately for Case 1 : $d \geq 1/2$ and Case 2: $0 \leq d < 1/2$.

Proof of Theorem 3, Case 1 : $d \geq 1/2$

By Theorem 2,

$$\hat{\tau}_{WLS} = \arg \max_{\tau \in [\tau_L, \tau_U]} \frac{T}{\omega^2 \sum_{t=1}^{T} \sigma_t^{-2}} Q_T\left(\tau; \sigma_t^{-1}\right) \xrightarrow{d} \arg \max_{\tau \in [\tau_L, \tau_U]} Q\left(\tau; 1/\sigma(.), \sigma(.), \delta, d\right)$$

and

$$\hat{\tau}_{FWLS} = \arg \max_{\tau \in [\tau_L, \tau_U]} \frac{T}{\omega^2 \sum_{t=1}^{T} \sigma_t^{-2}} Q_T\left(\tau; \hat{\sigma}^{-1}_{\hat{\tau},t}\right).$$

Let

$$R_T\left(\tau; \hat{\sigma}_{\hat{\tau},t}, \sigma_t\right) := \frac{T}{\omega^2 \sum_{t=1}^{T} \sigma_t^{-2}} \left(Q_T\left(\tau; \hat{\sigma}^{-1}_{\hat{\tau},t}\right) - Q_T\left(\tau; \sigma_t^{-1}\right)\right).$$

Theorem 3 can be established by showing that

$$\begin{aligned} \hat{\tau}_{FWLS} &= \arg \max_{\tau \in [\tau_L, \tau_U]} \frac{T}{\omega^2 \sum_{t=1}^{T} \sigma_t^{-2}} Q_T\left(\tau; \sigma_t^{-1}\right) + R_T\left(\tau; \hat{\sigma}_{\hat{\tau},t}, \sigma_t\right) \\ &\xrightarrow{d} \arg \max_{\tau \in [\tau_L, \tau_U]} Q\left(\tau; 1/\sigma(.), \sigma(.), \delta, d\right), \end{aligned}$$

(by appealing to the continuous mapping theorem for argmax functionals; see Kim and Pollard, 1990, Theorem 2.7), where

$$\sup_{\tau \in [\tau_L, \tau_U]} \left|R_T\left(\tau; \hat{\sigma}_{\hat{\tau},t}, \sigma_t\right)\right| \xrightarrow{p} 0. \tag{S.8}$$

Because $T^{-1}\sum_{t=1}^{T} \sigma_t^{-2} \rightarrow \int_0^1 \sigma(s)^{-2} ds$, to show (S.8) we need to show that the following two results hold:

$$\sup_{\tau \in [\tau_L, 1]} \left|T^{-1/2} \sum_{t=1}^{\lfloor \tau T \rfloor} e_t \hat{\sigma}^{-2}_{\hat{\tau},t} - e_t \sigma_t^{-2}\right| \xrightarrow{p} 0 \tag{S.9}$$

and

$$\sup_{\tau \in [\tau_L, 1]} \left|T^{-1} \sum_{t=1}^{\lfloor \tau T \rfloor} \hat{\sigma}^{-2}_{\hat{\tau},t} - \sigma_t^{-2}\right| \xrightarrow{p} 0. \tag{S.10}$$

Similar arguments apply for the $0 \leq d < 1/2$ case (see the proof of Theorem 1), which also requires the above two results as will be done below in Case 2.

To prove (S.9), observe first that

$$\sup_{\tau \in [\tau_L, 1]} \left|T^{-1/2} \sum_{t=1}^{\lfloor \tau T \rfloor} e_t \hat{\sigma}^{-2}_{\hat{\tau},t} - e_t \sigma_t^{-2}\right| \leq T^{-1/2} \sum_{t=1}^{T} \left|e_t\left(\hat{\sigma}^{-2}_{\hat{\tau},t} - \sigma_t^{-2}\right)\right|. \tag{S.11}$$

Define $\tilde{\sigma}_t^2 := \sum_{i=1}^T w_{ti} e_i^2$. Then, following Robinson (1987), to prove that the right hand side of (S.11) is $o_p(1)$, it is sufficient, by virtue of results in the proof of Theorem 2 of Xu and Phillips (2008) (specifically their equation (20) and part (a) on page 276) and in Lemma A of Xu and Phillips (2008), for us to show that

$$\sum_{t=1}^{T} \left|\hat{\sigma}_{\hat{\tau},t}^2 - \tilde{\sigma}_t^2\right|^2 \xrightarrow{p} 0. \tag{S.12}$$

Similarly, to prove (S.10), it is sufficient, by virtue of the results in part (d) of the proof of Theorem 2 of Xu and Phillips (2008, p.277), for us to show that

$$\max_{1\le t\le T} \left|\hat{\sigma}_{\hat{\tau},t}^2 - \tilde{\sigma}_t^2\right| \xrightarrow{p} 0. \tag{S.13}$$

For simplicity we set $d = 1/2$ now, as the results for $d > 1/2$ are essentially a special case and along exactly the same lines. We will now establish the validity of the results in (S.12) and (S.13). To show (S.12), write

$$\begin{aligned}
&\sum_{t=1}^{T} \left|\hat{\sigma}_{\hat{\tau},t}^2 - \tilde{\sigma}_t^2\right|^2 \\
\le\ & \sup_{\tau\in[\tau_L,\tau_U]} \sum_{t=1}^{T} \left|\hat{\sigma}_{\tau,t}^2 - \tilde{\sigma}_t^2\right|^2 \\
\le\ & \sup_{\tau\in[\tau_L,\tau_U]} \sum_{t=1}^{T} \left(\sum_{i=1}^{\lfloor \tau_0 T\rfloor} w_{ti}\left(\hat{e}_{\tau,i}^2 - e_i^2\right) + \sum_{i=\lfloor \tau_0 T\rfloor+1}^{\lfloor \tau T\rfloor} w_{ti}\left(\hat{e}_{\tau,i}^2 - e_i^2\right) + \sum_{i=\lfloor \tau T\rfloor+1}^{T} w_{ti}\left(\hat{e}_{\tau,i}^2 - e_i^2\right)\right)^2 \\
=:\ & \sup_{\tau\in[\tau_L,\tau_U]} \sum_{t=1}^{T} \left(A_{\tau,t} + B_{\tau,t} + C_{\tau,t}\right)^2.
\end{aligned} \tag{S.14}$$

Throughout this proof, in the interest of brevity, we will only discuss the case where $\lfloor \tau T\rfloor > \lfloor \tau_0 T\rfloor$, as the results for the case where $\lfloor \tau T\rfloor \le \lfloor \tau_0 T\rfloor$ follow along exactly the same lines.

For the $A_{\tau,t}$ term in (S.14), in which $i = 1, ..., \lfloor \tau_0 T\rfloor$,

$$\hat{e}_{\tau,i} = y_i - \hat{\mu}_{1,\tau} = \mu + e_i - \lfloor \tau T\rfloor^{-1} \sum_{t=1}^{\lfloor \tau T\rfloor} \left(\mu + \delta_T \cdot 1_{t>\lfloor \tau_0 T\rfloor} + e_t\right) = e_i + \alpha_{T,\tau} + \bar{e}_\tau \tag{S.15}$$

where $\alpha_{T,\tau} := \delta T^{-d}\tau^{-1}(\tau - \tau_0)$ and $\bar{e}_\tau := \lfloor \tau T\rfloor^{-1} \sum_{t=1}^{\lfloor \tau T\rfloor} e_t$.

For the $B_{\tau,t}$ term in (S.14), with $i = \lfloor \tau_0 T\rfloor + 1, ..., \lfloor \tau T\rfloor$,

$$\hat{e}_{\tau,i} = y_i - \hat{\mu}_{1,\tau} = \mu + \delta_T + e_i - \lfloor \tau T\rfloor^{-1} \sum_{t=1}^{\lfloor \tau T\rfloor} \left(\mu + \delta_T \cdot 1_{t>\lfloor \tau_0 T\rfloor} + e_t\right) = e_i + \beta_{T,\tau} + \bar{e}_\tau \tag{S.16}$$

where $\beta_{T,\tau} := \delta T^{-d}\tau^{-1}\tau_0$.

For the $C_{\tau,t}$ term in (S.14), in which $i = \lfloor \tau T\rfloor + 1, ..., T$,

$$\hat{e}_{\tau,i} = y_i - \hat{\mu}_{2,\tau} = \mu + \delta_T + e_i - \frac{1}{T - \lfloor \tau T\rfloor} \sum_{t=\lfloor \tau T\rfloor+1}^{T} \left(\mu + \delta_T \cdot 1_{t>\lfloor \tau_0 T\rfloor} + e_t\right) = e_i - \bar{e}_{(T-\lfloor \tau T\rfloor)}$$

where $\bar{e}_{(T-\lfloor \tau T \rfloor)} := (T - \lfloor \tau T \rfloor)^{-1} \sum_{t=\lfloor \tau T \rfloor+1}^{T} e_t$.

In what follows, we will only deal with the $A_{\tau,t}$ and $B_{\tau,t}$ terms. The expressions for $\hat{e}_{\tau,i}$ above imply that the arguments when dealing with the $C_{\tau,t}$ term follow similarly because it does not feature $\alpha_{T,\tau}$ and $\beta_{T,\tau}$, and because $\bar{e}_{(T-\lfloor \tau T \rfloor)} = \bar{e}_1 - \bar{e}_\tau$.

The following results will prove useful: (i) $\sup_{\tau \in [\tau_L, \tau_U]} \left| T^{1/2} \bar{e}_\tau \right| = O_p(1)$, which holds because

$$\sup_{\tau \in [\tau_L, \tau_U]} \left| T^{1/2} \bar{e}_\tau \right| = \sup_{\tau \in [\tau_L, \tau_U]} \left| \tau^{-1} T^{-1/2} \sum_{t=1}^{\lfloor \tau T \rfloor} e_t \right| \leq \tau_L^{-1} \sup_{\tau \in [\tau_L, \tau_U]} \left| T^{-1/2} \sum_{t=1}^{\lfloor \tau T \rfloor} e_t \right| = O_p(1)$$

by Lemma A.2; (ii) $\sup_{\tau \in [\tau_L, \tau_U]} |\alpha_{T,\tau}| = \sup_{\tau \in [\tau_L, \tau_U]} \left| \delta T^{-d} \tau^{-1} (\tau - \tau_0) \right| = O(T^{-d})$, which holds because $\sup_{\tau \in [\tau_L, \tau_U]} |\alpha_{T,\tau}| \leq \delta T^{-d} \tau_L^{-1} \sup_{\tau \in [\tau_L, \tau_U]} |\tau - \tau_0| \leq c T^{-d}$ for some $c > 0$; and (iii)

$$\sup_{\tau \in [\tau_L, \tau_U]} |\beta_{T,\tau}| = \sup_{\tau \in [\tau_L, \tau_U]} \left| \delta T^{-d} \tau^{-1} \tau_0 \right| = O(T^{-d}). \tag{S.17}$$

We will now show that $\sup_{\tau \in [\tau_L, \tau_U]} \sum_{t=1}^{T} A_{\tau,t}^2 = o_p(1)$ in (S.14). Using (S.15), we have that

$$\sum_{t=1}^{T} A_{\tau,t}^2 = \sum_{t=1}^{T} \left(\sum_{i=1}^{\lfloor \tau_0 T \rfloor} w_{ti} \left(\alpha_{T,\tau}^2 + \bar{e}_\tau^2 - 2 e_i \alpha_{T,\tau} - 2 e_i \bar{e}_\tau - 2 \alpha_{T,\tau} \bar{e}_\tau \right) \right)^2. \tag{S.18}$$

Consider the first term in (S.18). We have (because $d = 1/2$)

$$\begin{aligned}
\sup_{\tau \in [\tau_L, \tau_U]} \sum_{t=1}^{T} \left(\sum_{i=1}^{\lfloor \tau_0 T \rfloor} w_{ti} \alpha_{T,\tau}^2 \right)^2 &= \sup_{\tau \in [\tau_L, \tau_U]} |\alpha_{T,\tau}|^4 \sum_{t=1}^{T} \left(\sum_{i=1}^{\lfloor \tau_0 T \rfloor} w_{ti} \right)^2 \\
&\leq \sup_{\tau \in [\tau_L, \tau_U]} |\alpha_{T,\tau}|^4 \sum_{t=1}^{T} \left(\sum_{i=1}^{T} w_{ti} \right)^2 \\
&= \sup_{\tau \in [\tau_L, \tau_U]} |\alpha_{T,\tau}|^4 T = O_p\left(T^{-1}\right).
\end{aligned}$$

For the second term in (S.18), we have

$$\begin{aligned}
\sup_{\tau \in [\tau_L, \tau_U]} \sum_{t=1}^{T} \left(\sum_{i=1}^{\lfloor \tau_0 T \rfloor} w_{ti} \bar{e}_\tau^2 \right)^2 &= \sup_{\tau \in [\tau_L, \tau_U]} \left| T^{1/2} \bar{e}_\tau \right|^4 T^{-2} \sum_{t=1}^{T} \left(\sum_{i=1}^{\lfloor \tau_0 T \rfloor} w_{ti} \right)^2 \\
&\leq \sup_{\tau \in [\tau_L, \tau_U]} \left| T^{1/2} \bar{e}_\tau \right|^4 T^{-2} \sum_{t=1}^{T} \left(\sum_{i=1}^{T} w_{ti} \right)^2 = O_p\left(T^{-1}\right).
\end{aligned}$$

Before we consider the third term, note that

$$T^{-1} \sum_{t=1}^{T} \left(\sum_{i=1}^{\lfloor \tau_0 T \rfloor} w_{ti} e_i \right)^2 = O_p\left(\frac{1}{Th} \right) \tag{S.19}$$

because

$$
\begin{aligned}
T^{-1} \sum_{t=1}^{T} E\left(\sum_{i=1}^{\lfloor \tau_0 T \rfloor} w_{ti} e_i\right)^2 &= T^{-1} \sum_{t=1}^{T} \sum_{i=1}^{\lfloor \tau_0 T \rfloor} w_{ti}^2 \sigma_i^2 \\
&\leq \bar{\sigma}^2 T^{-1} \sum_{t=1}^{T} \sum_{i=1}^{\lfloor \tau_0 T \rfloor} w_{ti}^2 \\
&\leq \bar{\sigma}^2 \left(\max_{t,i} w_{ti}\right) T^{-1} \sum_{t=1}^{T} \sum_{i=1}^{\lfloor \tau_0 T \rfloor} w_{ti} \\
&\leq \bar{\sigma}^2 \left(\max_{t,i} w_{ti}\right) T^{-1} \sum_{t=1}^{T} \sum_{i=1}^{T} w_{ti} = O\left(\frac{1}{Th}\right),
\end{aligned}
$$

since $\max_{t,i} w_{ti} = O\left(1/Th\right)$; see Lemma A(d) of Xu and Phillips, 2008. Consequently, for the third term in (S.18), we have that

$$
\sup_{\tau \in [\tau_L, \tau_U]} \sum_{t=1}^{T} \left(\sum_{i=1}^{\lfloor \tau_0 T \rfloor} w_{ti} e_i \alpha_{T,\tau}\right)^2 = T \sup_{\tau \in [\tau_L, \tau_U]} \alpha_{T,\tau}^2 \cdot T^{-1} \sum_{t=1}^{T} \left(\sum_{i=1}^{\lfloor \tau_0 T \rfloor} w_{ti} e_i\right)^2 = O_p\left(\frac{1}{Th}\right).
$$

For the fourth term in (S.18), we have that

$$
\sup_{\tau \in [\tau_L, \tau_U]} \sum_{t=1}^{T} \left(\sum_{i=1}^{\lfloor \tau_0 T \rfloor} w_{ti} e_i \bar{e}_\tau\right)^2 \leq \sup_{\tau \in [\tau_L, \tau_U]} \left|T^{1/2} \bar{e}_\tau\right|^2 T^{-1} \sum_{t=1}^{T} \left(\sum_{i=1}^{\lfloor \tau_0 T \rfloor} w_{ti} e_i\right)^2 = O_p\left(\frac{1}{Th}\right).
$$

Finally, for the fifth term in (S.18), we have that

$$
\begin{aligned}
\sup_{\tau \in [\tau_L, \tau_U]} \sum_{t=1}^{T} \left(\sum_{i=1}^{\lfloor \tau_0 T \rfloor} w_{ti} \alpha_{T,\tau} \bar{e}_\tau\right)^2 &\leq \sup_{\tau \in [\tau_L, \tau_U]} \left|T^{1/2} \bar{e}_\tau\right|^2 \sup_{\tau \in [\tau_L, \tau_U]} \left|\alpha_{T,\tau}\right|^2 T^{-1} \sum_{t=1}^{T} \left(\sum_{i=1}^{T} w_{ti}\right)^2 \\
&= O_p\left(T^{-1}\right).
\end{aligned}
$$

The cross-product terms in (S.18) are all of $o_p(1)$ uniformly in $\tau \in [\tau_L, \tau_U]$ by the application of the Cauchy-Schwarz inequality.

The term $B_{\tau,t}$ in (S.14) follows similarly since $\sup_{\tau \in [\tau_L, \tau_U]} \left|\beta_{T,\tau}\right| = O\left(T^{-1/2}\right)$; see (S.17). All of the cross-product terms are again of $o_p(1)$ by the Cauchy-Schwarz inequality. This completes the proof for (S.12).

We next prove (S.13). We have

$$
\max_{1 \leq t \leq T} \left|\hat{\sigma}_{\hat{\tau},t}^2 - \tilde{\sigma}_t^2\right| \leq \max_{1 \leq t \leq T} \sup_{\tau \in [\tau_L, \tau_U]} \left|\hat{\sigma}_{\tau,t}^2 - \tilde{\sigma}_t^2\right| \leq \max_{1 \leq t \leq T} \sup_{\tau \in [\tau_L, \tau_U]} \left|A_{\tau,t} + B_{\tau,t} + C_{\tau,t}\right|.
$$

We will only show that $\max_{1 \leq t \leq T} \sup_{\tau \in [\tau_L, \tau_U]} \left|A_{\tau,t}\right| = o_p(1)$ since the term $B_{\tau,t}$ follows similarly as just noted. We have

$$
\left|A_{\tau,t}\right| = \left|\sum_{i=1}^{\lfloor \tau_0 T \rfloor} w_{ti} \left(\alpha_{T,\tau}^2 + \bar{e}_\tau^2 - 2 e_i \alpha_{T,\tau} - 2 e_i \bar{e}_\tau - 2 \alpha_{T,\tau} \bar{e}_\tau\right)\right|. \tag{S.20}
$$

For the first term in (S.20), we have that

$$
\begin{aligned}
\max_{1\le t\le T}\sup_{\tau\in[\tau_L,\tau_U]}\left|\sum_{i=1}^{\lfloor\tau_0 T\rfloor} w_{ti}\alpha_{T,\tau}^2\right| &\le \sup_{\tau\in[\tau_L,\tau_U]}\left|\alpha_{T,\tau}\right|^2 \max_{1\le t\le T}\left|\sum_{i=1}^{\lfloor\tau_0 T\rfloor} w_{ti}\right| \\
&\le \sup_{\tau\in[\tau_L,\tau_U]}\left|\alpha_{T,\tau}\right|^2 \max_{1\le t\le T}\left|\sum_{i=1}^{T} w_{ti}\right| \\
&= \sup_{\tau\in[\tau_L,\tau_U]}\left|\alpha_{T,\tau}\right|^2 = O_p\left(T^{-1}\right).
\end{aligned}
$$

For the second term in (S.20),

$$
\begin{aligned}
\max_{1\le t\le T}\sup_{\tau\in[\tau_L,\tau_U]}\left|\sum_{i=1}^{\lfloor\tau_0 T\rfloor} w_{ti}\bar{e}_\tau^2\right| &\le \sup_{\tau\in[\tau_L,\tau_U]}\left|T^{1/2}\bar{e}_\tau\right|^2 \max_{1\le t\le T}\left|\sum_{i=1}^{\lfloor\tau_0 T\rfloor} w_{ti}\right| \\
&\le \sup_{\tau\in[\tau_L,\tau_U]}\left|T^{1/2}\bar{e}_\tau\right|^2 \max_{1\le t\le T}\left|\sum_{i=1}^{T} w_{ti}\right| = O_p\left(T^{-1}\right).
\end{aligned}
$$

Before turning to the third term, we note that $T^{-1}\sum_{i=1}^{T}|e_i| = O_p(1)$. using this result, for the third term in (S.20) we have that,

$$
\begin{aligned}
\max_{1\le t\le T}\sup_{\tau\in[\tau_L,\tau_U]}\left|\sum_{i=1}^{\lfloor\tau_0 T\rfloor} w_{ti}e_i\alpha_{T,\tau}\right| &\le \sup_{\tau\in[\tau_L,\tau_U]}\left|\alpha_{T,\tau}\right| \max_{1\le t\le T}\left|\sum_{i=1}^{\lfloor\tau_0 T\rfloor} w_{ti}e_i\right| \\
&\le T\sup_{\tau\in[\tau_L,\tau_U]}\left|\alpha_{T,\tau}\right| \max_{1\le t\le T} T^{-1}\sum_{i=1}^{\lfloor\tau_0 T\rfloor} w_{ti}\left|e_i\right| \\
&\le T\sup_{\tau\in[\tau_L,\tau_U]}\left|\alpha_{T,\tau}\right| \left(\max_{t,i} w_{ti}\right) T^{-1}\sum_{i=1}^{\lfloor\tau_0 T\rfloor}\left|e_i\right| \\
&\le T\sup_{\tau\in[\tau_L,\tau_U]}\left|\alpha_{T,\tau}\right| \left(\max_{t,i} w_{ti}\right) T^{-1}\sum_{i=1}^{T}\left|e_i\right| \\
&= O_p\left(\frac{1}{\sqrt{T}h}\right) = o_p(1).
\end{aligned}
$$

Next, for the fourth term in (S.20) we have that,

$$
\begin{aligned}
\max_{1\le t\le T}\sup_{\tau\in[\tau_L,\tau_U]}\left|\sum_{i=1}^{\lfloor\tau_0 T\rfloor} w_{ti}e_i\bar{e}_\tau\right| &\le \sup_{\tau\in[\tau_L,\tau_U]}\left|T^{1/2}\bar{e}_\tau\right| T^{-1/2}\max_{1\le t\le T}\left|\sum_{i=1}^{\lfloor\tau_0 T\rfloor} w_{ti}e_i\right| \\
&\le \sup_{\tau\in[\tau_L,\tau_U]}\left|T^{1/2}\bar{e}_\tau\right| T^{1/2}\max_{1\le t\le T} T^{-1}\sum_{i=1}^{\lfloor\tau_0 T\rfloor} w_{ti}\left|e_i\right| \\
&\le \sup_{\tau\in[\tau_L,\tau_U]}\left|T^{1/2}\bar{e}_\tau\right| T^{1/2}\left(\max_{t,i} w_{ti}\right) T^{-1}\sum_{i=1}^{\lfloor\tau_0 T\rfloor}\left|e_i\right| \\
&\le \sup_{\tau\in[\tau_L,\tau_U]}\left|T^{1/2}\bar{e}_\tau\right| T^{1/2}\left(\max_{t,i} w_{ti}\right) T^{-1}\sum_{i=1}^{T}\left|e_i\right| \\
&= O_p\left(\frac{1}{\sqrt{T}h}\right).
\end{aligned}
$$

Finally, for the fifth term in (S.20) we have that,

$$\begin{aligned}
\max_{1\le t\le T}\sup_{\tau\in[\tau_L,\tau_U]}\left|\sum_{i=1}^{\lfloor\tau_0T\rfloor}w_{ti}\alpha_{T,\tau}\bar{e}_\tau\right| &\le \sup_{\tau\in[\tau_L,\tau_U]}\left|T^{1/2}\bar{e}_\tau\right|\sup_{\tau\in[\tau_L,\tau_U]}|\alpha_{T,\tau}|\,T^{-1/2}\max_{1\le t\le T}\left|\sum_{i=1}^{\lfloor\tau_0T\rfloor}w_{ti}\right| \\
&\le \sup_{\tau\in[\tau_L,\tau_U]}\left|T^{1/2}\bar{e}_\tau\right|\sup_{\tau\in[\tau_L,\tau_U]}|\alpha_{T,\tau}|\,T^{-1/2}\max_{1\le t\le T}\left|\sum_{i=1}^{T}w_{ti}\right| \\
&= O_p\left(T^{-1}\right).
\end{aligned}$$

This completes the proof of the theorem for the $d\ge 1/2$ case.

Proof of Theorem 3, Case 2 : $0\le d<1/2$

We now consider $0\le d<1/2$ and prove that (S.12) and (S.13) also hold here. To ease notation we will simply use $\hat{\tau}$ to denote $\hat{\tau}_{OLS}$ in the remainder of this proof.

To show (S.12), we write, for any τ,

$$\sum_{t=1}^{T}\left|\hat{\sigma}^2_{\tau,t}-\tilde{\sigma}^2_t\right|^2=\sum_{t=1}^{T}\left|A_{\tau,t}+B_{\tau,t}+C_{\tau,t}\right|^2. \tag{S.21}$$

We now evaluate each term in (S.21) at $\tau=\hat{\tau}$ where $\hat{\tau}-\tau_0=O_p\left(T^{-1+2d}\right)$. To that end, we first note that

$$|\alpha_{T,\hat{\tau}}|=\left|\delta T^{-d}\hat{\tau}^{-1}(\hat{\tau}-\tau_0)\right|=O_p\left(T^{-1+d}\right)$$

because $|\alpha_{T,\hat{\tau}}|=\delta T^{-d}\left|\hat{\tau}^{-1}\right||\hat{\tau}-\tau_0|\le\delta\tau_L^{-1}T^{-d}|\hat{\tau}-\tau_0|$. Moreover, $|\beta_{T,\hat{\tau}}|=\delta T^{-d}\left|\frac{\tau_0}{\hat{\tau}}\right|=O_p\left(T^{-d}\right)$.

Let us first deal with the $A_{\tau,t}$ term in (S.21). To that end, recall first that

$$\sum_{t=1}^{T}A^2_{\tau,t}=\sum_{t=1}^{T}\left(\sum_{i=1}^{\lfloor\tau_0T\rfloor}w_{ti}\left(\alpha^2_{T,\tau}+\bar{e}^2_\tau-2e_i\alpha_{T,\tau}-2e_i\bar{e}_\tau-2\alpha_{T,\tau}\bar{e}_\tau\right)\right)^2. \tag{S.22}$$

We now evaluate every term in (S.22) at $\tau=\hat{\tau}$. For the first term,

$$\sum_{t=1}^{T}\left(\sum_{i=1}^{\lfloor\tau_0T\rfloor}w_{ti}\alpha^2_{T,\hat{\tau}}\right)^2\le|\alpha_{T,\hat{\tau}}|^4\sum_{t=1}^{T}\left(\sum_{i=1}^{T}w_{ti}\right)^2=O_p\left(T^{4d-3}\right)=o_p(1).$$

For the second term,

$$\sum_{t=1}^{T}\left(\sum_{i=1}^{\lfloor\tau_0T\rfloor}w_{ti}\bar{e}^2_{\hat{\tau}}\right)^2\le\sup_{\tau\in[\tau_L,\tau_U]}\left|T^{1/2}\bar{e}_\tau\right|^4T^{-2}\sum_{t=1}^{T}\left(\sum_{i=1}^{T}w_{ti}\right)^2=O_p\left(T^{-1}\right).$$

For the third term, recalling equation (S.19),

$$\sum_{t=1}^{T}\left(\sum_{i=1}^{\lfloor\tau_0T\rfloor}w_{ti}e_i\alpha_{T,\hat{\tau}}\right)^2=T\,|\alpha_{T,\hat{\tau}}|^2\,T^{-1}\sum_{t=1}^{T}\left(\sum_{i=1}^{\lfloor\tau_0T\rfloor}w_{ti}e_i\right)^2=O_p\left(\frac{1}{T^{2-2d}h}\right)=o_p(1).$$

For the fourth term,

$$\sum_{t=1}^{T}\left(\sum_{i=1}^{\lfloor\tau_0T\rfloor}w_{ti}e_i\bar{e}_{\hat{\tau}}\right)^2\le\sup_{\tau\in[\tau_L,\tau_U]}\left|T^{1/2}\bar{e}_\tau\right|^2T^{-1}\sum_{t=1}^{T}\left(\sum_{i=1}^{\lfloor\tau_0T\rfloor}w_{ti}e_i\right)^2=O_p\left(\frac{1}{Th}\right).$$

Finally, for the fifth term,

$$\sum_{t=1}^{T}\left(\sum_{i=1}^{\lfloor\tau_0 T\rfloor} w_{ti}\alpha_{T,\hat{\tau}}\bar{e}_{\hat{\tau}}\right)^2 \leq \sup_{\tau\in[\tau_L,\tau_U]}\left|T^{1/2}\bar{e}_\tau\right|^2 \left|\alpha_{T,\hat{\tau}}\right|^2 T^{-1}\sum_{t=1}^{T}\left(\sum_{i=1}^{T} w_{ti}\right)^2$$
$$= O_p\left(T^{2d-2}\right) = o_p(1).$$

We next deal with $B_{\tau,t}$ term in (S.21). Using (S.14) and (S.16), we have that

$$\sum_{t=1}^{T} B_{\tau,t}^2 = \sum_{t=1}^{T}\left(\sum_{i=\lfloor\tau_0 T\rfloor+1}^{\lfloor\tau T\rfloor} w_{ti}\left(\beta_{T,\tau}^2 + \bar{e}_\tau^2 - 2e_i\beta_{T,\tau} - 2e_i\bar{e}_\tau - 2\beta_{T,\tau}\bar{e}_\tau\right)\right)^2. \quad \text{(S.23)}$$

We will need to evaluate each term at $\tau = \hat{\tau}$. For the first term we have that

$$\sum_{t=1}^{T}\left(\sum_{i=\lfloor\tau_0 T\rfloor+1}^{\lfloor\hat{\tau} T\rfloor} w_{ti}\beta_{T,\hat{\tau}}^2\right)^2 = \left|\beta_{T,\hat{\tau}}\right|^4 \sum_{t=1}^{T}\left(\sum_{i=\lfloor\tau_0 T\rfloor+1}^{\lfloor\hat{\tau} T\rfloor} w_{ti}\right)^2$$
$$\leq \left|\beta_{T,\hat{\tau}}\right|^4 \left(\max_{t,i} w_{ti}\right)^2 T\left(T\left(\hat{\tau}-\tau_0\right)\right)^2 = O_p\left(\frac{1}{Th^2}\right) = o_p(1).$$

For the second term

$$\sum_{t=1}^{T}\left(\sum_{i=\lfloor\tau_0 T\rfloor+1}^{\lfloor\hat{\tau} T\rfloor} w_{ti}\bar{e}_{\hat{\tau}}^2\right)^2 = \left|\bar{e}_{\hat{\tau}}\right|^4 \sum_{t=1}^{T}\left(\sum_{i=\lfloor\tau_0 T\rfloor+1}^{\lfloor\hat{\tau} T\rfloor} w_{ti}\right)^2$$
$$\leq \sup_{\tau\in[\tau_L,\tau_U]}\left|T^{1/2}\bar{e}_\tau\right|^4 \left(\max_{t,i} w_{ti}\right)^2 T^{-1}\left(T\left(\hat{\tau}-\tau\right)\right)^2$$
$$= O_p\left(\frac{1}{T^{3-4d}h^2}\right) = o_p(1).$$

To deal with the third term in (S.23), define

$$V_{\lfloor\tau T\rfloor,t} := \sum_{i=\lfloor\tau_0 T\rfloor+1}^{\lfloor\tau T\rfloor} w_{ti}e_i$$

for $t = 1,\ldots,T$. Since $\hat{\tau}-\tau_0 = O_p\left(T^{-1}\delta_T^{-2}\right)$ we consider values of τ in the set

$$K_T(M) := \left\{\tau : \lfloor\tau_0 T\rfloor + 1 \leq \lfloor\tau T\rfloor \leq \lfloor\tau_0 T\rfloor + v\delta_T^{-2} \text{ for all } 0 \leq v \leq M\right\},$$

for $M < \infty$, which is an interval above τ_0. (The corresponding interval below τ_0 is handled identically, see the proof of Theorem 1 of Bai (1994) for the same approach.) Since $V_{\lfloor\tau T\rfloor,t}$ is a martingale indexed by $\lfloor\tau T\rfloor$, Doob's inequality implies

$$E\left(\sup_{\tau\in K_T(M)}\left|\sum_{i=1}^{\lfloor\tau T\rfloor-\lfloor\tau_0 T\rfloor} w_{t(i+\lfloor\tau_0 T\rfloor)}e_{i+\lfloor\tau_0 T\rfloor}\right|^2\right) \leq 4E\left|\sum_{i=1}^{v\delta_T^{-2}} w_{t(i+\lfloor\tau_0 T\rfloor)}e_{i+\lfloor\tau_0 T\rfloor}\right|^2$$
$$\leq 4\sum_{i=1}^{v\delta_T^{-2}} w_{t(i+\lfloor\tau_0 T\rfloor)}^2\sigma_{i+\lfloor\tau_0 T\rfloor}^2$$
$$\leq 4\bar{\sigma}^2 v\left(\max_{t,i} w_{ti}\right)^2\delta_T^{-2}$$
$$= O\left(\frac{1}{T^{2-2d}h^2}\right).$$

uniformly in t (in view of the uniform boundedness of w_{ti} over t). Hence, uniformly in t,

$$\left|V_{\lfloor\hat{\tau}T\rfloor,t}\right| \leq \sup_{\tau\in K_T(M)} \left|V_{\lfloor\tau T\rfloor,t}\right| = O_p\left(\frac{1}{T^{1-d}h}\right). \tag{S.24}$$

Therefore the third term in (S.23) is such that

$$\begin{aligned}\sum_{t=1}^{T}\left(\sum_{i=\lfloor\tau_0 T\rfloor+1}^{\lfloor\hat{\tau}T\rfloor} w_{ti}e_i\beta_{T,\hat{\tau}}\right)^2 &= T\left|\beta_{T,\hat{\tau}}\right|^2 T^{-1}\sum_{t=1}^{T}\left(\sum_{i=\lfloor\tau_0 T\rfloor+1}^{\lfloor\hat{\tau}T\rfloor} w_{ti}e_i\right)^2 \\ &= O_p\left(\frac{1}{Th^2}\right).\end{aligned}$$

For the fourth term in (S.23) we have that,

$$\begin{aligned}\sum_{t=1}^{T}\left(\sum_{i=\lfloor\tau_0 T\rfloor+1}^{\lfloor\hat{\tau}T\rfloor} w_{ti}e_i\bar{e}_{\hat{\tau}}\right)^2 &= T\left|\bar{e}_{\hat{\tau}}\right|^2 T^{-1}\sum_{t=1}^{T}\left(\sum_{i=\lfloor\tau_0 T\rfloor+1}^{\lfloor\hat{\tau}T\rfloor} w_{ti}e_i\right)^2 \\ &\leq \sup_{\tau\in[\tau_L,\tau_U]}\left|T^{1/2}\bar{e}_\tau\right|^2 T^{-1}\sum_{t=1}^{T}\left(\sum_{i=\lfloor\tau_0 T\rfloor+1}^{\lfloor\hat{\tau}T\rfloor} w_{ti}e_i\right)^2 \\ &= O_p\left(\frac{1}{T^{2-2d}h^2}\right).\end{aligned}$$

Finally, for the fifth term in (S.23) we have that,

$$\begin{aligned}&\sum_{t=1}^{T}\left(\sum_{i=\lfloor\tau_0 T\rfloor+1}^{\lfloor\hat{\tau}T\rfloor} w_{ti}\beta_{T,\hat{\tau}}\bar{e}_{\hat{\tau}}\right)^2 \\ \leq\ &\sup_{\tau\in[\tau_L,\tau_U]}\left|T^{1/2}\bar{e}_\tau\right|^2 T^{-1}\left|\beta_{T,\hat{\tau}}\right|^2\sum_{t=1}^{T}\left(\sum_{i=\lfloor\tau_0 T\rfloor+1}^{\lfloor\hat{\tau}T\rfloor} w_{ti}\right)^2 \\ \leq\ &\sup_{\tau\in[\tau_L,\tau_U]}\left|T^{1/2}\bar{e}_\tau\right|^2 T^{-1}\left|\beta_{T,\hat{\tau}}\right|^2\left(\max_{t,i} w_{ti}\right)^2 T\left(T\left(\hat{\tau}-\tau_0\right)\right)^2 \\ =\ &O_p\left(\frac{1}{T^{2-2d}h^2}\right).\end{aligned}$$

Next, to show (S.13), we write, for any τ,

$$\max_{1\leq t\leq T}\left|\hat{\sigma}^2_{\tau,t}-\tilde{\sigma}^2_t\right| \leq \max_{1\leq t\leq T}\left|A_{\tau,t}+B_{\tau,t}+C_{\tau,t}\right|. \tag{S.25}$$

For $A_{\tau,t}$, we recall that

$$\left|A_{\tau,t}\right| = \left|\sum_{i=1}^{\lfloor\tau_0 T\rfloor} w_{ti}\left(\alpha^2_{T,\tau}+\bar{e}^2_\tau-2e_i\alpha_{T,\tau}-2e_i\bar{e}_\tau-2\alpha_{T,\tau}\bar{e}_\tau\right)\right|. \tag{S.26}$$

We therefore need to evaluate each term in (S.26) at $\tau=\hat{\tau}$. For the first term, we have that

$$\begin{aligned}\max_{1\leq t\leq T}\left|\sum_{i=1}^{\lfloor\tau_0 T\rfloor} w_{ti}\alpha^2_{T,\hat{\tau}}\right| &\leq \left|\alpha_{T,\hat{\tau}}\right|^2\max_{1\leq t\leq T}\left|\sum_{i=1}^{\lfloor\tau_0 T\rfloor} w_{ti}\right| \\ &\leq \left|\alpha_{T,\hat{\tau}}\right|^2\max_{1\leq t\leq T}\sum_{i=1}^{T} w_{ti} = O_p\left(T^{-2+2d}\right) = o_p(1).\end{aligned}$$

For the second term in (S.26) we have that

$$\begin{aligned}
\max_{1\le t\le T}\left|\sum_{i=1}^{\lfloor\tau_0 T\rfloor} w_{ti}\bar{e}_{\hat{\tau}}^2\right| &\le \left|\bar{e}_{\hat{\tau}}\right|^2 \max_{1\le t\le T}\left|\sum_{i=1}^{\lfloor\tau_0 T\rfloor} w_{ti}\right| \\
&\le \sup_{\tau\in[\tau_L,\tau_U]}\left|T^{1/2}\bar{e}_\tau\right|^2 T^{-1}\max_{1\le t\le T}\left|\sum_{i=1}^{T} w_{ti}\right| = O_p\left(T^{-1}\right).
\end{aligned}$$

For the third term in (S.26) we have that

$$\begin{aligned}
\max_{1\le t\le T}\left|\sum_{i=1}^{\lfloor\tau_0 T\rfloor} w_{ti}e_i\alpha_{T,\hat{\tau}}\right| &\le \left|\alpha_{T,\hat{\tau}}\right| \max_{1\le t\le T}\left|\sum_{i=1}^{\lfloor\tau_0 T\rfloor} w_{ti}e_i\right| \\
&\le T\left|\alpha_{T,\hat{\tau}}\right| \max_{1\le t\le T} T^{-1}\sum_{i=1}^{\lfloor\tau_0 T\rfloor} w_{ti}\left|e_i\right| \\
&\le T\left|\alpha_{T,\hat{\tau}}\right| \left(\max_{t,i} w_{ti}\right) T^{-1}\sum_{i=1}^{\lfloor\tau_0 T\rfloor}\left|e_i\right| \\
&\le T\left|\alpha_{T,\hat{\tau}}\right| \left(\max_{t,i} w_{ti}\right) T^{-1}\sum_{i=1}^{T}\left|e_i\right| \\
&= O_p\left(\frac{1}{T^{1-d}h}\right) = o_p(1).
\end{aligned}$$

For the fourth term in (S.26) we have that

$$\begin{aligned}
\max_{1\le t\le T}\left|\sum_{i=1}^{\lfloor\tau_0 T\rfloor} w_{ti}e_i\bar{e}_{\hat{\tau}}\right| &\le \left|\bar{e}_{\hat{\tau}}\right| \max_{1\le t\le T}\left|\sum_{i=1}^{\lfloor\tau_0 T\rfloor} w_{ti}e_i\right| \\
&\le \sup_{\tau\in[\tau_L,\tau_U]}\left|T^{1/2}\bar{e}_\tau\right| T^{-1/2}\max_{1\le t\le T}\sum_{i=1}^{\lfloor\tau_0 T\rfloor} w_{ti}\left|e_i\right| \\
&\le \sup_{\tau\in[\tau_L,\tau_U]}\left|T^{1/2}\bar{e}_\tau\right| T^{1/2}\left(\max_{t,i} w_{ti}\right) T^{-1}\sum_{i=1}^{\lfloor\tau_0 T\rfloor}\left|e_i\right| \\
&= O_p\left(\frac{1}{\sqrt{T}h}\right).
\end{aligned}$$

Finally, for the fifth term in (S.26) we have that

$$\begin{aligned}
\max_{1\le t\le T}\left|\sum_{i=1}^{\lfloor\tau_0 T\rfloor} w_{ti}\alpha_{T,\hat{\tau}}\bar{e}_{\hat{\tau}}\right| &\le \left|\bar{e}_{\hat{\tau}}\right|\left|\alpha_{T,\hat{\tau}}\right| \max_{1\le t\le T}\left|\sum_{i=1}^{\lfloor\tau_0 T\rfloor} w_{ti}\right| \\
&\le \sup_{\tau\in[\tau_L,\tau_U]}\left|T^{1/2}\bar{e}_\tau\right| \frac{1}{\sqrt{T}}\left|\alpha_{T,\hat{\tau}}\right| \max_{1\le t\le T}\left|\sum_{i=1}^{T} w_{ti}\right| \\
&\le \left|\alpha_{T,\hat{\tau}}\right| \sup_{\tau\in[\tau_L,\tau_U]}\left|T^{1/2}\bar{e}_\tau\right| \frac{1}{T^{1/2}}\max_{1\le t\le T}\sum_{i=1}^{T} w_{ti} = O_p\left(T^{d-3/2}\right).
\end{aligned}$$

Turning next to the $B_{\tau,t}$ term in (S.25), we have that

$$|B_{\tau,t}| = \left|\sum_{i=\lfloor\tau_0 T\rfloor+1}^{\lfloor\tau T\rfloor} w_{ti}\left(\beta_{T,\tau}^2 + \bar{e}_\tau^2 - 2e_i\beta_{T,\tau} - 2e_i\bar{e}_\tau - 2\beta_{T,\tau}\bar{e}_\tau\right)\right|$$

and we will again need to evaluate each term at $\tau = \hat{\tau}$. The first term is of $o_p(1)$ because

$$\begin{aligned}
\max_{1\le t\le T}\left|\sum_{i=\lfloor\tau_0 T\rfloor+1}^{\lfloor\hat{\tau}T\rfloor} w_{ti}\beta_{T,\hat{\tau}}^2\right| &\le |\beta_{T,\hat{\tau}}|^2 \max_{1\le t\le T}\left|\sum_{i=\lfloor\tau_0 T\rfloor+1}^{\lfloor\hat{\tau}T\rfloor} w_{ti}\right| \\
&\le |\beta_{T,\hat{\tau}}|^2 \left(\max_{t,i} w_{ti}\right)|T(\hat{\tau}-\tau_0)| = O_p\left(\frac{1}{Th}\right).
\end{aligned}$$

The second term is of $o_p(1)$ because

$$\begin{aligned}
\max_{1\le t\le T}\left|\sum_{i=\lfloor\tau_0 T\rfloor+1}^{\lfloor\hat{\tau}T\rfloor} w_{ti}\bar{e}_{\hat{\tau}}^2\right| &\le |\bar{e}_{\hat{\tau}}|^2 \max_{1\le t\le T}\left|\sum_{i=\lfloor\tau_0 T\rfloor+1}^{\lfloor\hat{\tau}T\rfloor} w_{ti}\right| \\
&\le \sup_{\tau\in[\tau_L,\tau_U]}\left|T^{1/2}\bar{e}_\tau\right|^2 T^{-1}\left(\max_{t,i} w_{ti}\right)T(\hat{\tau}-\tau_0) \\
&= O_p\left(\frac{1}{T^{2-2d}h}\right).
\end{aligned}$$

For the third term, we note first that from part (d) of the proof of Theorem 2 of Xu and Phillips (2008, p.277), as $X_{t-1} = 1$ in their notation, (S.25) becomes $T^{-1}\sum_{t=1}^{T}|B_{\tau,t}|$. Using (S.24), we therefore have that

$$T^{-1}\sum_{t=1}^{T}\left|\sum_{i=\lfloor\tau_0 T\rfloor+1}^{\lfloor\hat{\tau}T\rfloor} w_{ti}e_i\beta_{T,\hat{\tau}}\right| \le |\beta_{T,\hat{\tau}}|\, T^{-1}\sum_{t=1}^{T}\left|\sum_{i=\lfloor\tau_0 T\rfloor+1}^{\lfloor\hat{\tau}T\rfloor} w_{ti}e_i\right| = O_p\left(\frac{1}{Th}\right).$$

Similarly for the fourth term,

$$\begin{aligned}
T^{-1}\sum_{t=1}^{T}\left|\sum_{i=\lfloor\tau_0 T\rfloor+1}^{\lfloor\hat{\tau}T\rfloor} w_{ti}e_i\bar{e}_{\hat{\tau}}\right| &\le |\bar{e}_{\hat{\tau}}|\, T^{-1}\sum_{t=1}^{T}\left|\sum_{i=\lfloor\tau_0 T\rfloor+1}^{\lfloor\hat{\tau}T\rfloor} w_{ti}e_i\right| \\
&\le T^{-1/2}\sup_{\tau\in[\tau_L,\tau_U]}\left|T^{1/2}\bar{e}_\tau\right| T^{-1}\sum_{t=1}^{T}\left|\sum_{i=\lfloor\tau_0 T\rfloor+1}^{\lfloor\hat{\tau}T\rfloor} w_{ti}e_i\right| \\
&= O_p\left(\frac{1}{T^{3/2-d}h}\right) = o_p(1).
\end{aligned}$$

Finally, the fifth term is also of $o_p(1)$ because

$$\begin{aligned}
\max_{1\le t\le T}\left|\sum_{i=\lfloor\tau_0 T\rfloor+1}^{\lfloor\hat{\tau}T\rfloor} w_{ti}\beta_{T,\hat{\tau}}\bar{e}_{\hat{\tau}}\right| &\le |\bar{e}_{\hat{\tau}}|\,|\beta_{T,\hat{\tau}}| \max_{1\le t\le T}\left|\sum_{i=\lfloor\tau_0 T\rfloor+1}^{\lfloor\hat{\tau}T\rfloor} w_{ti}\right| \\
&\le \sup_{\tau\in[\tau_L,\tau_U]}\left|T^{1/2}\bar{e}_\tau\right| T^{-1/2}|\beta_{T,\hat{\tau}}|\left(\max_{t,i} w_{ti}\right)T(\hat{\tau}-\tau_0) \\
&= O_p\left(\frac{1}{T^{3/2-d}h}\right).
\end{aligned}$$

This completes the proof for the $0 \le d < 1/2$ case.

■

Proof of Theorem S.1

For any τ, define $X_t(\tau) := (1, t, 1_{t>\lfloor \tau T \rfloor}(t - \lfloor \tau T \rfloor))'$. The DGP in (2.1) can then be expressed as

$$\begin{aligned} y_t &= X_t(\tau_0)'\theta + z_t, \\ &= X_t(\tau)'\theta - d_t(\tau)\delta_T + z_t \end{aligned} \tag{S.27}$$

where $\theta := (\alpha, \mu, \delta_T)'$, and $d_t(\tau) := 1_{t>\lfloor \tau T \rfloor}(t - \lfloor \tau T \rfloor) - 1_{t>\lfloor \tau_0 T \rfloor}(t - \lfloor \tau_0 T \rfloor)$. Applying QD to (S.27) for any $\bar{c}$ gives

$$y_{\bar{c},t} = X_{\bar{c},t}(\tau)'\theta - d_{\bar{c},t}\delta_T + z_{\bar{c},t}, \tag{S.28}$$

where $z_{\bar{c},t} := \Delta z_t - \bar{c}T^{-1}z_{t-1}$,

$$X_{\bar{c},t}(\tau) := \begin{pmatrix} 1_{t=1} + 1_{t>1}(-\bar{c}/T) \\ 1 - \bar{c}(t-1)/T \\ 1_{t>\lfloor \tau T \rfloor}(1 - \bar{c}(t - \lfloor \tau T \rfloor - 1)/T) \end{pmatrix},$$

and

$$d_{\bar{c},t}(\tau) := 1_{t>\lfloor \tau T \rfloor}\left(1 - \bar{c}\frac{t - \lfloor \tau T \rfloor - 1}{T}\right) - 1_{t>\lfloor \tau_0 T \rfloor}\left(1 - \bar{c}\frac{t - \lfloor \tau_0 T \rfloor - 1}{T}\right). \tag{S.29}$$

Therefore, using equation (S.28), the estimated QD coefficient vector satisfies

$$\begin{aligned} \hat{\theta}_{\bar{c}}(\tau) &= \left(\sum_{t=1}^{T} X_{\bar{c},t}(\tau)X_{\bar{c},t}(\tau)'\right)^{-1} \sum_{t=1}^{T} X_{\bar{c},t}(\tau)y_{\bar{c},t} \\ &= \theta + \left(\sum_{t=1}^{T} X_{\bar{c},t}(\tau)X_{\bar{c},t}(\tau)'\right)^{-1} \sum_{t=1}^{T} X_{\bar{c},t}(\tau)(z_{\bar{c},t} - d_{\bar{c},t}\delta_T), \end{aligned}$$

and the QD residuals are

$$\begin{aligned} \hat{z}_{\bar{c},t}(\tau) &:= y_t - X_t(\tau)'\hat{\theta}_{\bar{c}}(\tau) \\ &= z_t - d_t(\tau)\delta_T - X_t(\tau)'(\hat{\theta}_{\bar{c}}(\tau) - \theta) \\ &= z_t - X_t(\tau)'\left(\sum_{t=1}^{T} X_{\bar{c},t}(\tau)X_{\bar{c},t}(\tau)'\right)^{-1} \sum_{t=1}^{T} X_{\bar{c},t}(\tau)z_{\bar{c},t} && \text{(S.30)} \\ &\quad - \left[d_t(\tau)\delta_T - X_t(\tau)'\left(\sum_{t=1}^{T} X_{\bar{c},t}(\tau)X_{\bar{c},t}(\tau)'\right)^{-1} \sum_{t=1}^{T} X_{\bar{c},t}(\tau)d_{\bar{c},t}\delta_T\right]. && \text{(S.31)} \end{aligned}$$

The asymptotic distribution of the DF t-statistic calculated using $\hat{z}_{\bar{c},t}(\hat{\tau})$ for either $\hat{\tau} = \hat{\tau}_{OLS}$ or $\hat{\tau} = \hat{\tau}_{FWLS}$ follows from applying the FCLT results in (S.30) and showing that (S.31) (which captures the estimation effect of $\hat{\tau}$ for τ_0) is asymptotically negligible.

In (S.30) we make use of the heteroskedastic FCLT for $z_t = \phi_T z_{t-1} + e_t$ with $\phi := 1 + c/T$, such that

$$\omega_e^{-1}T^{-1/2}z_{\lfloor sT \rfloor} = \omega^{-1}T^{-1/2}\sum_{t=1}^{\lfloor sT \rfloor} e_t \xrightarrow{d} B_\eta^c(s) := \int_0^s \exp(c(s-r))dB_\eta(r)$$

where $\omega_e^2 := \int_0^1 \sigma(r)^2 dr$. It also then follows that

$$\omega_e^{-1}T^{-1/2}\sum_{t=1}^{\lfloor sT \rfloor} z_{\bar{c},t} = \omega_e^{-1}T^{-1/2}\sum_{t=1}^{\lfloor sT \rfloor}(\Delta z_t - \bar{c}T^{-1}z_{t-1}) \xrightarrow{d} B_\eta^c(s) - \bar{c}\int_0^s B_\eta^c(r)dr =: B_\eta^c(s; \bar{c}).$$

Defining $D_T := \text{diag}(1, T^{-1/2}, T^{-1/2})$, standard calculations (cf. Perron and Rodríguez, 2003, and Cavaliere *et al.*, 2011) give the following limits, in each case uniformly in $\tau \in [\tau_L, \tau_U]$,

$$D_T^{-1} T^{-1/2} X_{\lfloor sT \rfloor}(\tau) \to \begin{pmatrix} 0 \\ s \\ (s-\tau) \vee 0 \end{pmatrix}$$

$$\sum_{t=1}^{T} D_T^{-1} X_{\bar{c},t}(\tau) X_{\bar{c},t}(\tau)' D_T^{-1} \to \begin{pmatrix} 1 & 0 & 0 \\ 0 & \int_0^1 (1-\bar{c}s)^2 ds & \int_\tau^1 (1-\bar{c}s)(1-\bar{c}(s-\tau)) ds \\ 0 & \int_\tau^1 (1-\bar{c}s)(1-\bar{c}(s-\tau)) ds & \int_\tau^1 (1-\bar{c}(s-\tau))^2 ds \end{pmatrix}$$

$$\omega_e^{-1} \sum_{t=1}^{T} D_T^{-1} X_{\bar{c},t}(\tau) z_{\bar{c},t} \xrightarrow{d} \begin{pmatrix} z_{\bar{c},1} \\ \int_0^1 (1-\bar{c}s)\, dB_\eta^c(s;\bar{c}) \\ \int_\tau^1 (1-\bar{c}(s-\tau))\, dB_\eta^c(s;\bar{c}) \end{pmatrix}.$$

Defining

$$X(s;\tau) := \begin{pmatrix} s \\ (s-\tau) \vee 0 \end{pmatrix} \quad \text{and} \quad X_{\bar{c}}(s;\tau) := \begin{pmatrix} 1-\bar{c}s \\ 1-\bar{c}((s-\tau) \vee 0) \end{pmatrix}$$

the FCLT applied to (S.30) can be expressed

$$\begin{aligned} &\omega_e^{-1} T^{-1/2} z_{\lfloor sT \rfloor} - T^{-1/2} X_{\lfloor sT \rfloor}(\tau)' D_T^{-1} \left(\sum_{t=1}^{T} D_T^{-1} X_{\bar{c},t}(\tau) X_{\bar{c},t}(\tau)' D_T^{-1} \right)^{-1} \omega_e^{-1} \sum_{t=1}^{T} D_T^{-1} X_{\bar{c},t}(\tau) z_{\bar{c},t} \\ &\quad \xrightarrow{d} B_\eta^c(s) - X(s;\tau)' \left(\int_0^1 X_{\bar{c}}(s;\tau) X_{\bar{c}}(s;\tau)' ds \right)^{-1} \int_0^1 X_{\bar{c}}(s;\tau) dB_\eta^c(s;\bar{c}) \\ &\quad =: Z(s;\tau,c,\bar{c},\eta). \end{aligned} \tag{S.32}$$

If τ is evaluated at an estimator $\hat{\tau}$ that is consistent for τ_0 then the continuous mapping theorem applies to $\omega_e^{-1} T^{-1/2} \hat{z}_{\bar{c},\lfloor sT \rfloor}(\hat{\tau})$ to deduce that the limiting process is $Z(s;\tau_0,c,\bar{c},\eta)$. The derivation of the asymptotic distribution of the DF t-statistic from the derived behaviour of this partial sum process is then entirely standard.

Now consider (S.31), scaled by $T^{-1/2}$ commensurately with (S.32), written as

$$r_{\lfloor sT \rfloor}(\tau) = T^{-1/2} d_{\lfloor sT \rfloor}(\tau) \delta_T - a_{\lfloor sT \rfloor}(\tau)' b_T(\tau)$$

where

$$\begin{aligned} a_{\lfloor sT \rfloor}(\tau) &:= T^{-1/2} D_T^{-1} X_{\lfloor sT \rfloor}(\tau) \\ b_T(\tau) &:= \left(\sum_{t=1}^{T} D_T^{-1} X_{\bar{c},t}(\tau) X_{\bar{c},t}(\tau)' D_T^{-1} \right)^{-1} \sum_{t=1}^{T} D_T^{-1} X_{\bar{c},t}(\tau) d_{\bar{c},t} \delta_T. \end{aligned}$$

Then

$$\max_{s\in[0,1]} |r_{\lfloor sT \rfloor}(\tau)| \le \max_{s\in[0,1]} T^{-1/2} |\delta_T d_{\lfloor sT \rfloor}(\tau)| + \max_{s\in[0,1]} \left(a_{\lfloor sT \rfloor}(\tau)' a_{\lfloor sT \rfloor}(\tau) \right)^{1/2} \left(b_T(\tau)' b_T(\tau) \right)^{1/2} \tag{S.33}$$

The first term includes

$$d_t(\tau) = \begin{cases} 0 & \text{if } t \leq \lfloor T\tau_0 \rfloor, t \leq \lfloor T\tau \rfloor \\ -(t - \lfloor T\tau_0 \rfloor) & \text{if } \lfloor T\tau_0 \rfloor < t \leq \lfloor T\tau \rfloor \\ t - \lfloor T\tau \rfloor & \text{if } \lfloor T\tau \rfloor < t \leq \lfloor T\tau_0 \rfloor \\ -(\lfloor T\tau \rfloor - \lfloor T\tau_0 \rfloor) & \text{if } t > \lfloor T\tau_0 \rfloor, t > \lfloor T\tau \rfloor \end{cases}$$

so that $|d_t(\tau)| \leq |\lfloor T\tau \rfloor - \lfloor T\tau_0 \rfloor|$ for every t, and hence

$$\max_{s\in[0,1]} T^{-1/2}|\delta_T d_{\lfloor sT \rfloor}(\tau)| \leq T^{1/2}\delta_T \left| \frac{\lfloor \tau T \rfloor}{T} - \frac{\lfloor \tau_0 T \rfloor}{T} \right|. \tag{S.34}$$

For the second term in (S.33) we have

$$a_{\lfloor sT \rfloor}(\tau) = T^{-1/2} D_T^{-1} X_{\lfloor sT \rfloor}(\tau) = \begin{pmatrix} T^{-1/2} \\ \lfloor sT \rfloor / T \\ ((\lfloor sT \rfloor - \lfloor \tau T \rfloor) \vee 0)/T \end{pmatrix} \leq \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix},$$

so

$$\max_{s\in[0,1]} \left(a_{\lfloor sT \rfloor}(\tau)' a_{\lfloor sT \rfloor}(\tau) \right)^{1/2} \leq \sqrt{3}.$$

For the last term in (S.33), we have shown above that $\left(\sum_{t=1}^{T} D_T^{-1} X_{\bar{c},t}(\tau) X_{\bar{c},t}(\tau)' D_T^{-1} \right)^{-1}$ is a well-behaved bounded matrix, and therefore we focus on $\sum_{t=1}^{T} D_T^{-1} X_{\bar{c},t}(\tau) d_{\bar{c},t}(\tau)\delta_T$. The structure of $d_{\bar{c},t}(\tau)$ can be seen from (S.29) to be

$$d_{\bar{c},t}(\tau) = \begin{cases} 0, & \text{if } t \leq \lfloor \tau T \rfloor, t \leq \lfloor \tau_0 T \rfloor \\ \left(1 - \bar{c}\, \frac{t - \lfloor \tau T \rfloor - 1}{T}\right), & \text{if } \lfloor \tau T \rfloor < t \leq \lfloor \tau_0 T \rfloor \\ -\left(1 - \bar{c}\, \frac{t - \lfloor \tau_0 T \rfloor - 1}{T}\right), & \text{if } \lfloor \tau_0 T \rfloor < t \leq \lfloor \tau T \rfloor \\ \bar{c}\left(\frac{\lfloor \tau T \rfloor}{T} - \frac{\lfloor \tau_0 T \rfloor}{T} \right) & \text{if } t > \lfloor \tau T \rfloor, t > \lfloor \tau_0 T \rfloor. \end{cases}$$

Taking $\tau_0 \leq \tau$ (the reverse follows similarly and gives the same inequality)

$$\begin{aligned} & T^{-1} \sum_{t=1}^{T} |d_{\bar{c},t}(\tau)| \\ &= T^{-1} \sum_{t=\lfloor \tau_0 T \rfloor + 1}^{\lfloor \tau T \rfloor} \left| 1 - \bar{c}\, \frac{t - \lfloor \tau_0 T \rfloor - 1}{T} \right| + T^{-1} \sum_{t=\lfloor \tau T \rfloor + 1}^{T} |\bar{c}| \left| \frac{\lfloor \tau T \rfloor}{T} - \frac{\lfloor \tau_0 T \rfloor}{T} \right| \\ &\leq (1 + |\bar{c}|) \left| \frac{\lfloor \tau T \rfloor}{T} - \frac{\lfloor \tau_0 T \rfloor}{T} \right| + |\bar{c}| T^{-2} \sum_{t=1}^{\lfloor \tau T \rfloor - \lfloor \tau_0 T \rfloor - 1} t \\ &= (1 + |\bar{c}|) \left| \frac{\lfloor \tau T \rfloor}{T} - \frac{\lfloor \tau_0 T \rfloor}{T} \right| + \frac{|\bar{c}|}{2} \left(\frac{\lfloor \tau T \rfloor}{T} - \frac{\lfloor \tau_0 T \rfloor}{T} \right) \left(\frac{\lfloor \tau T \rfloor}{T} - \frac{\lfloor \tau_0 T \rfloor}{T} - \frac{1}{T} \right) \\ &\leq \left| \frac{\lfloor \tau T \rfloor}{T} - \frac{\lfloor \tau_0 T \rfloor}{T} \right| \left(1 + \frac{3|\bar{c}|}{2} \right). \end{aligned}$$

Considering

$$\sum_{t=1}^{T} D_T^{-1} X_{\bar{c},t}(\tau) d_{\bar{c},t}(\tau) \delta_T = \begin{pmatrix} \delta T^{-d} \sum_{t=2}^{T} \left(-\frac{\bar{c}}{T} \right) d_{\bar{c},t}(\tau) \\ \delta T^{-d-1/2} \sum_{t=1}^{T} \left(1 - \bar{c} \frac{t-1}{T} \right) d_{\bar{c},t}(\tau) \\ \delta T^{-d-1/2} \sum_{t=\lfloor \tau T \rfloor + 1}^{T} \left(1 - \bar{c} \frac{t - \lfloor \tau T \rfloor - 1}{T} \right) d_{\bar{c},t}(\tau) \end{pmatrix}. \tag{S.35}$$

it follows that

$$\left|T^{-d}\sum_{t=2}^{T}\left(-\frac{\bar{c}}{T}\right)d_{\bar{c},t}(\tau)\right| \leq |\bar{c}|T^{-d}T^{-1}\sum_{t=1}^{T}|d_{\bar{c},t}(\tau)| \leq |\bar{c}|\left(1+\frac{3|\bar{c}|}{2}\right)T^{-d}\left|\frac{\lfloor\tau T\rfloor}{T}-\frac{\lfloor\tau_0 T\rfloor}{T}\right|$$

and

$$\begin{aligned}\left|T^{-d-1/2}\sum_{t=1}^{T}\left(1-\bar{c}\frac{t-1}{T}\right)d_{\bar{c},t}(\tau)\right| &\leq (1+|\bar{c}|)T^{-d+1/2}T^{-1}\sum_{t=1}^{T}|d_{\bar{c},t}(\tau)| \\ &\leq (1+|\bar{c}|)\left(1+\frac{3|\bar{c}|}{2}\right)T^{-d+1/2}\left|\frac{\lfloor\tau T\rfloor}{T}-\frac{\lfloor\tau_0 T\rfloor}{T}\right|\end{aligned}$$

and a bound of the same order applies to the third term. Putting these together gives

$$(b_T(\tau)'b_T(\tau))^{1/2} \leq k_2 T^{-d+1/2}\left|\frac{\lfloor\tau T\rfloor}{T}-\frac{\lfloor\tau_0 T\rfloor}{T}\right|$$

for some finite constant k (depending on the various constants involving $\bar{c}$ above), and combining this with (S.34) in (S.33) gives

$$\max_{s\in[0,1]}|r_{\lfloor sT\rfloor}(\tau)| \leq kT^{-d+1/2}\left|\frac{\lfloor\tau T\rfloor}{T}-\frac{\lfloor\tau_0 T\rfloor}{T}\right|$$

for another finite constant k.

If we now consider the evaluation of this at $\tau=\hat{\tau}$, where $\hat{\tau}=\hat{\tau}_{OLS}$ or $\hat{\tau}=\hat{\tau}_{FWLS}$, the rate of consistency $\hat{\tau}-\tau_0=O_p(T^{2d-1})$ implies that

$$\left|\frac{\lfloor\hat{\tau}T\rfloor}{T}-\frac{\lfloor\tau_0 T\rfloor}{T}\right| = O_p(T^{2d-1}),$$

and, hence, that

$$\max_{s\in[0,1]}|r_{\lfloor sT\rfloor}(\hat{\tau})| \leq kT^{-d+1/2}\left|\frac{\lfloor\hat{\tau}T\rfloor}{T}-\frac{\lfloor\tau_0 T\rfloor}{T}\right| = O_p(T^{d-1/2}).$$

This is therefore of $o_p(1)$ for all $d<1/2$. ■

Proof of Theorem S.2

The two equations of the DGP

$$\begin{aligned}y_t &= \alpha+\mu t+\delta_T 1_{t>\lfloor\tau_0 T\rfloor}(t-\lfloor\tau_0 T\rfloor)+z_t \\ z_t &= \phi_T z_{t-1}+e_t\end{aligned}$$

can be combined to give

$$\begin{aligned}y_t &= X_{1,t}'\beta_{1,T}+X_{2,t}(\tau_0)'\beta_{2,T}+e_t \\ &= X_{1,t}'\beta_{1,T}+X_{2,t}(\tau)'\beta_{2,T}+e_{T,t}(\tau)\end{aligned}$$

where

$$X_{1,t} := \begin{pmatrix}1\\ t\\ y_{t-1}\end{pmatrix}, \quad \beta_{1,T} := \begin{pmatrix}\mu\phi_T+\alpha(1-\phi_T)\\ \mu(1-\phi_T)\\ \phi\end{pmatrix}$$

and

$$X_{2,t}(\tau) := \begin{pmatrix} 1_{t>\lfloor\tau_0 T\rfloor} \\ 1_{t>\lfloor\tau_0 T\rfloor}(t-\lfloor\tau_0 T\rfloor) \end{pmatrix}, \quad \beta_{2,T} := \delta_T \begin{pmatrix} \phi_T \\ 1-\phi_T \end{pmatrix}$$

and

$$e_{T,t}(\tau) := e_t - (X_{2,t}(\tau) - X_{2,t}(\tau_0))'\beta_{2,T}.$$

The compound disturbance $e_{T,t}(\tau)$ includes the DGP disturbance term e_t and also $(X_{2,t}(\tau) - X_{2,t}(\tau_0))'\beta_{2,T}$, which captures the effect of τ differing from τ_0, if applicable. When τ is replaced by a consistent estimator, e.g. $\hat{\tau}_{OLS}$ or $\hat{\tau}_{FWLS}$, the asymptotic negligibility of this latter term follows by similar arguments to those used to the same effect in the proof of Theorem 4. It is convenient to stack the observations as

$$y := \begin{pmatrix} y_2 \\ \vdots \\ y_T \end{pmatrix}, \quad X_1 := \begin{pmatrix} X'_{1,2} \\ \vdots \\ X'_{1,T} \end{pmatrix}, \quad X_{2,T} := \begin{pmatrix} X_{2,2}(\tau)' \\ \vdots \\ X_{2,T}(\tau)' \end{pmatrix},$$

and to define the orthogonal projection matrix $\bar{P}_1 := I_{T-1} - X_1(X_1'X_1)^{-1}X_1'$, so that

$$\begin{aligned} s_0^2 &= (T-1)^{-1}y'\bar{P}_1 y \\ s_\tau^2 &= (T-1)^{-1}(y'\bar{P}_1 y - y'\bar{P}_1 X_{2,\tau}(X'_{2,\tau}\bar{P}_1 X_{2,\tau})^{-1}X'_{2,\tau}\bar{P}_1 y). \end{aligned}$$

The difference between the resulting SC statistics can be written

$$\begin{aligned} SC_\tau - SC_0 &= (T-1)\log\left(\frac{s_\tau^2}{s_0^2}\right) + 2\log(T-1) \\ &= (T-1)\log\left(1 - \frac{y'\bar{P}_1 X_{2,\tau}(X'_{2,\tau}\bar{P}_1 X_{2,\tau})^{-1}X'_{2,\tau}\bar{P}_1 y}{y'\bar{P}_1 y}\right) + 2\log(T-1) \\ &= (T-1)\log\left(1 - \frac{\hat{\beta}_2(\tau)'X'_{2,\tau}\bar{P}_1 X_{2,\tau}\hat{\beta}_2(\tau)}{y'\bar{P}_1 y}\right) + 2\log(T-1). \end{aligned} \tag{S.36}$$

When a trend break is present in the DGP ($\delta \neq 0$) and $0 \leq d < 1/2$, standard but tedious least squares derivations show that with $\tau = \tau_0$

$$\hat{\beta}_2(\tau_0) = T^{-d}\delta b_{2,T} + o_p(T^{-d})$$

with $b_{2,T} := (1, -c/T)'$. This is essentially the consistency of the OLS estimator of the regression, which can be shown to hold in the presence of heteroskedasticity of the form allowed in Assumption $\mathcal{A}$. Hence

$$(T-1)\log\left(1 - \frac{\hat{\beta}_2(\tau_0)'X'_{2,\tau_0}\bar{P}_1 X_{2,\tau_0}\hat{\beta}_2(\tau_0)}{y'\bar{P}_1 y}\right) \approx -T^{1-2d}\delta^2\frac{b'_{2,T}X'_{2,\tau_0}\bar{P}_1 X_{2,\tau_0}b_{2,T}}{y'\bar{P}_1 y}$$

which diverges to $-\infty$ at rate T^{1-2d} for $0 \leq d < 1/2$, because the ratio of quadratic forms in this expression is $O_p(1)$. This in turn implies that the terms in $SC_{\tau_0} - SC_0$ have orders $-O_p(T^{1-2d}) + O(\log T)$, and hence that $SC_{\tau_0} - SC_0$ also diverges to $-\infty$. The conclusion is that the probability that the model with the break (i.e. including $X_{2,t}(\tau_0)$) is chosen converges to one when $\delta \neq 0$ and $0 \leq d < 1/2$. The same results follow similarly when τ_0 is replaced by a consistent estimator, such as $\hat{\tau}_{OLS}$ or $\hat{\tau}_{FWLS}$. It also turns out that the same rates of convergence, and hence consistent detection of a break, can also be found for sequences of τ not converging to τ_0, although in that case the expression for $b_{2,T}$ will differ and the finite sample properties would presumably be inferior to when a consistent estimator is used for τ_0.

When a trend break is absent in the DGP ($\delta = 0$), the estimator $\hat{\beta}_2(\tau_0)$ converges to zero; i.e., it is a consistent estimator of $\beta_{2,T} = 0$ in this case. Moreover it can be shown to be $O_p(T^{-1/2})$ in the usual way, so that the first term in (S.36) satisfies

$$(T-1)\log\left(1 - \frac{\hat{\beta}_2(\tau_0)' X'_{2,\tau_0} \bar{P}_1 X_{2,\tau_0} \hat{\beta}_2(\tau_0)}{y' \bar{P}_1 y}\right) \approx (T-1)\log(1 - O_p(T^{-1})) \approx O_p(1),$$

from which it follows that $SC_\tau - SC_0 \approx O_p(1) + 2\log(T-1) \to +\infty$. Thus when a break is absent, the probability that it is excluded by the SC comparison converges to 1.

The preceding arguments apply without substantial change when the regressions used for the SC calculations are weighted to allow for unconditional heteroskedasticity of the form given in $\mathcal{A}_2$. ■

Additional References

Andrews, D.W.K. (1993). Tests for parameter instability and structural change with unknown change point. Econometrica 61, 821—56 (Corrigendum, 71, 395—7).

Kim, D. and P. Perron (2009). Unit root tests allowing for a break in the trend function at an unknown time under both the null and alternative hypotheses. Journal of Econometrics 148, 1–13.

Kim, J. and D. Pollard (1990). Cube root asymptotics. The Annals of Statistics 18, 191–219.

Robinson, P.M. (1987). Asymptotically Efficient Estimation in the Presence of Heteroskedasticity of Unknown Form. Econometrica 55, 875–891.