Essex Finance Centre
Working Paper Series

Working Paper No 94: 01-2026

"Covariate Augmented CUSUM Bubble Monitoring Procedures"

"Sam Astill, A. M. Robert Taylor and Yang Zu"

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

Covariate Augmented CUSUM Bubble Monitoring Procedures*

Sam Astill[a], A.M. Robert Taylor[a,b] and Yang Zu[c]

[a] Essex Business School, University of Essex

[b] Aarhus Center for Econometrics, Aarhus University

[c] Department of Economics and APAEM, University of Macau

January 19, 2026

Abstract

We explore how information from covariates can be incorporated into the CUSUM-based real-time monitoring procedure for explosive asset price bubbles developed in Homm and Breitung (2012). Where dynamic covariates are present in the data generating process, the false positive rate of the basic CUSUM procedure, which is based on the assumption that prices follow a univariate data generating process, under the null of no explosivity will not, in general, be properly controlled, even asymptotically. In contrast, accounting for these relevant covariates in the construction of the CUSUM statistics leads to a procedure whose false positive rate can be controlled using the same asymptotic crossing function as employed by Homm and Breitung (2012). Doing so is also shown to have the potential to significantly increase the chance of detecting an emerging bubble episode in finite samples. We additionally allow for time varying volatility in the innovations driving the model through the use of a kernel-based variance estimator.

Keywords: rational bubbles; explosive autoregression; covariates; CUSUM; real-time monitoring; non-stationary volatility.

JEL Classification: C12, C15, C22, C32, C58, G12.

*We are very grateful to the Editor, Yixiao Sun, and three anonymous referees for their careful reading of, and very helpful and constructive comments on, earlier versions of this paper. We also thank Giuseppe Cavaliere for useful comments. Any remaining errors are our own. Taylor is grateful for research support from the Aarhus Center for Econometrics (ACE) funded by the Danish National Research Foundation grant number DNRF186. Yang Zu gratefully acknowledges financial support from the University of Macau's APAEM Seed Grant. Address correspondence to: Robert Taylor, Essex Business School, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, UK. Email: robert.taylor@essex.ac.uk.

1 Introduction and Motivation

Asset price bubbles tend to be characterised by a sudden and explosive increase in the price of an asset without a corresponding increase in the fundamental value of the asset (thereby representing a misallocation of resources), followed by a subsequent destruction of value through a price collapse. Bubbles often presage economic recessions; indeed, the 2007/08 Global Financial Crisis (GFC) was preceded by suspected price bubbles in the U.S. housing, commodity and stock markets. In the aftermath of the GFC, policymakers have considered new rules for macroprudential regulation and intervention. Crucial to the effectiveness of these is the availability of econometric methods which can monitor the behaviour of prices in asset markets in real-time, rapidly and accurately detecting emerging price bubbles.

The majority of the bubble detection literature has focused on one-shot tests for detecting the presence of historic asset price bubbles. The seminal contributions in this area were made by Phillips, Wu and Yu (2011) [PWY] and Phillips, Shi and Yu (2015) [PSY], who proposed tests for the presence of bubble episodes based on the maximum of sequences of recursive univariate augmented Dickey-Fuller [ADF] unit root statistics applied to overlapping sub-samples of the data. Other contributions based on sub-sample based methods include: Homm and Breitung (2012) [HB], Harvey *et al.* (2016), Astill *et al.* (2017), Phillips and Shi (2018) and Harvey *et al.* (2019, 2020).

Although primarily designed as one-shot tests and date-stamping procedures for historical bubbles, some of these approaches can also be implemented sequentially to provide methods to monitor for the emergence of a bubble in real-time; most notably the $BSADF$ statistic of PSY (defined as the maximum of a backward-recursive sequence of subsample ADF statistics computed over all possible subsamples ending at the last available date in the full data sample, subject to a minimum subsample length). By implementing tests sequentially, however, a critical value which diverges with the sample size (satisfying the rate condition given in equation (11) on page 1055 of PSY), needs to be used to control the *false positive rate* [FPR] of the monitoring procedure, defined as the probability of incorrectly declaring a bubble during the monitoring period; see Section 3.2 of PSY. This rate condition implies a theoretical FPR (by which we mean the FPR of the procedure in large samples) of zero. In practice, PSY (p.1066) recommend obtaining the critical value by Monte Carlo simulation, yielding a real-time monitoring procedure with a controlled, but non-zero, FPR. This procedure is, however, infeasible in the case where the innovations display time-varying volatility. To allow for possible time-varying volatility, Phillips and Shi (2020, Section 5) propose a wild bootstrap monitoring procedure, based on the $BSADF$ statistic, whose FPR can be controlled at a specified level across a monitoring period of a given length. This procedure is implemented at the end of the chosen monitoring period, and so is not run in real-time; it may, however, be possible to modify this procedure to be implemented in real-time.

A different strand of the literature, which we focus on in this paper, has developed dedicated real-time monitoring procedures for asset price bubbles, designed so that the practitioner can fix the theoretical FPR at a given (non-zero) level. These split the data into a *training sample* and a *monitoring period*. HB use a CUSUM-based detector where a sequence of CUSUM statistics, calculated from the first differences of the data in real-time over the monitoring period, are compared against a theoretical crossing function (such that the critical value becomes larger the further into the monitoring sequence one is). In a different approach, Astill *et al.* (2018) use a method based on comparing the maximum value of statistics computed in the training sample and monitoring period. Both of these procedures are designed for the case where the innovations are unconditionally homoskedastic and assume that no relevant covariates exist. To deal with the first issue, Astill *et al.* (2023a) [AHLTZ] propose standardising the CUSUM statistics used in the HB procedure by a nonparametric kernel-based spot variance estimator at each monitoring point. They show that a monitoring procedure based on these standardised CUSUM statistics has a theoretically controlled FPR even where the innovations are unconditionally heteroskedastic. As we will show, failure to account for relevant dynamic covariates in the data generating process (DGP) can lead to spurious over-rejection in both the HB and AHLTZ procedures.

It seems eminently plausible that information *additional* to the asset price series under test could usefully be deployed in bubble detection methods. Indeed, the literature suggests several potential covariates that might aid in identifying periods of explosive behaviour. For equities, dividend discount type models (Diba and Grossman, 1998; PSY) link prices to the risk-free rate of interest, whilst the capital asset pricing model (Kim and Kim, 2016) can embed time-varying volatility. Pricing equations for commodity spot prices (Tsvetanov *et* *al.*, 2016) indicate inventories (Kilian and Murphy, 2014) play a role. Finally, given bubble behaviour in real estate may precede equity (Caballero *et al.*, 2008) and commodity market bubbles (Phillips and Yu, 2011), potential housing market covariates such as interest rates, disposable income and mortgage finance (White, 2015) may be particularly useful.

Despite these considerations, the majority of contributions in the bubble testing literature, and all of those described above, are purely *univariate*, using information from the price series under consideration alone. Two notable exceptions are Shi and Phillips (2023) and Astill *et al.* (2023b) [ATKK]. In the context of detecting house price bubbles, Shi and Phillips (2023) develop $BSADF$-type statistics applied to the (cumulated) residuals from a first-stage IVX regression (see, e.g., Kostakis *et al.*, 2015) which filters out market fundamentals from an observed price-to-rent series, and use these in a monitoring procedure based on the approach of Phillips and Shi (2020), discussed above. More relevant to the present setting, ATKK adapt the covariate augmented Dickey-Fuller [CADF] unit root test proposed by Hansen (1995) to develop versions of the historical bubble testing procedures of PWY and PSY, allowing information from covariates to be exploited. Hansen (1995) shows that the inclusion of relevant (stationary) covariates in the CADF regression reduces the error variance relative to a univariate ADF regression and so can lead to more precise estimation of the model. ATKK show that the resulting covariate augmented variants of the PWY and PSY tests can in some cases display significantly higher power to detect historical asset prices bubbles than their univariate counterparts from PWY and PSY.

Given the policy need for real-time monitoring procedures that can detect emerging bubbles as rapidly as possible, the findings in ATKK suggest it is worth exploring if the incorporation of additional information from covariates can both improve the efficacy of real-time bubble monitoring procedures to detect emerging bubble episodes, while also delivering a controlled FPR under the null. Motivated by the CUSUM approach of Kramer, Ploberger and Alt (1988) [KPA], developed for detecting structural changes in dynamic models, we propose CUSUM type real-time monitoring statistics based on recursive residuals from a regression of the first differences of the price series under test on relevant covariates. Like AHLTZ, we implement the procedures using a nonparametric kernel-based spot variance estimator at each time point to allow for time-varying volatility in the innovations. We also allow for serial correlation in the innovations, something also not allowed under the assumptions in HB.

We demonstrate that the resulting CUSUM statistic retains the same (pivotal) limiting distribution under the constant parameter unit root null as HB's original CUSUM statistic attains under the regularity conditions in their paper. Consequently, a covariate augmented monitoring procedure with a theoretically controlled FPR can be constructed by appealing to large sample results from Chu *et al.* (1996). Monte Carlo simulations show that for a wide range of potential DGPs our proposed covariate augmented CUSUM monitoring procedure, implemented using a standard BIC criterion to decide whether or not to include a candidate covariate, performs well in practice. In particular, and unlike the univariate CUSUM-based monitoring procedures, the finite sample FPRs of the covariate augmented procedures are well controlled when a genuine covariate is present in the DGP. Moreover, where the covariate enters the DGP, the *true positive rate* [TPR], defined as the cumulative probability of detecting a bubble present in the monitoring period, is much superior to the univariate procedures. Additionally, the impact on finite sample performance is very small in the case where the candidate covariate does not enter the DGP.

The remainder of the paper is organised as follows. Section 2 outlines the DGP we work with and the assumptions under which we will operate. Section 3 gives a brief description of the standard CUSUM procedure of HB. Section 4 outlines our proposed covariate augmented CUSUM monitoring procedure for covariates that are allowed to have non-zero means and details its large sample behaviour. The results from our Monte Carlo simulation study are reported in Section 5. Section 6 concludes. A supplementary appendix details: the analogous procedure for the case where it is *known* that the covariates are mean zero; proofs of the technical results given in the paper; additional simulation results, and an empirical illustration using the dataset of Welch and Goyal (2008).

2 The Model and Assumptions

Let $\{y_t\}$ be generated according to the following data generating process [DGP],

$$y_t = \mu^* + u_t \tag{1}$$

$$u_t = \begin{cases} u_{t-1} + v_t & t = 1, \ldots, \lfloor \tau T \rfloor \\ (1+\delta)u_{t-1} + v_t & t = \lfloor \tau T \rfloor + 1, \ldots, \lfloor \lambda T \rfloor \end{cases} \tag{2}$$

where $1 \leq \tau \leq \lambda$, $\lambda > 1$ and $\lfloor . \rfloor$ denotes the integer part of its argument. The initial condition u_0 is assumed to be of $O_p(1)$. Under (2), u_t follows the time-varying AR(1) process

$$\Delta u_t = \delta_t u_{t-1} + v_t, \quad t = 1, \ldots, T, \ldots, \lfloor \lambda T \rfloor \tag{3}$$

where $\Delta := (1 - L)$ is the usual first difference operator in the lag operator, L. The AR coefficient δ_t can be seen to change from 0 to $\delta \geq 0$ at time $t = \lfloor \tau T \rfloor + 1$.

In the context of (1)-(2) we will be concerned with two sub-sample periods of the series y_t. The first of these is the period $t = 1, ..., T$, which will form the *training sample* in our analysis, and the second is the period $t = T + 1, ..., \lfloor \lambda T \rfloor$, which will form the *monitoring period* for our procedure. Our model imposes that y_t follows a unit root process over the training sample $t = 1, ..., T$, while over the monitoring period y_t again follows a unit root process over the sub-period $t = T + 1, ..., \lfloor \tau T \rfloor$, but crucially is subject to potentially explosive behaviour in the period $t = \lfloor \tau T \rfloor + 1, ..., \lfloor \lambda T \rfloor$ if $\delta > 0$.[1] In total, at the end of the monitoring period, there are $\lfloor \lambda T \rfloor$ observations. When $\delta > 0$, if $\tau = 1$ then the explosive regime will begin at the start of the monitoring period. In the context of monitoring for explosive autoregressive behaviour during the monitoring period, our implicit null hypothesis is given by $H_0 : \delta = 0$, with the corresponding alternative hypothesis, $H_1 : \delta > 0$.

With respect to the error process, v_t, in (2), we allow v_t to be serially correlated, heteroskedastic and (potentially) related to an $(m \times 1)$ vector of covariates, x_t. In the same spirit as Hansen (1995), we achieve this by assuming that v_t satisfies Assumption 1.

Assumption 1. *Let v_t be generated by the pth order heteroskedastic autoregressive exogenous [ARX(p)] process*

$$\alpha(L)v_t = \beta(L)'[x_t - c_x] + \varepsilon_t, \quad \varepsilon_t = \sigma_t \eta_t \tag{4}$$

where $\alpha(z) := 1 - \sum_{k=1}^{p} \alpha_k z^k$, $\beta(z) := \sum_{k=0}^{q} \beta_k z^k$, and where $x_t := (x_{1,t}, ..., x_{m,t})'$ is an m-vector of stochastic covariates with constant mean vector c_x. Let the mean-centred vector of

[1]The DGP in (1)-(2) does not consider the case where the explosive regime collapses before the monitoring period ends. It could be extended to allow either an instantaneous collapse (as in, e.g., PWY), or a stationary collapse regime (as in, e.g., Harvey *et al.*, 2016). However, when monitoring for an emerging explosive regime in real-time, the nature of any post-explosive collapse has no bearing on the detection properties of the monitoring procedure, so we specify a non-collapsing explosive regime for simplicity.

covariates be denoted $w_t := x_t - c_x =: (w_{1,t}, ..., w_{m,t})'$. *The innovations,* η_t, *form a sequence of serially uncorrelated conditionally heteroskedastic innovations with mean zero and unit (unconditional) variance, with* σ_t *a (deterministic) time-varying volatility function, such that* ε_t *has time-varying unconditional variance,* σ_t^2.

Remark 2.1. In (4), the lag polynomial $\beta(L)$ allows for, but does not require, lags of the covariate x_t to enter the DGP. Compared to Equation (5) of Hansen (1995,p.1150), $\beta(L)$, however, excludes the possibility of leads of the covariate entering (4). This is a consequence of the fact that our interest in this paper is on developing real-time monitoring procedures, whereby lead variables would be unavailable to the practitioner; see also Remark 4.1 of ATKK (p.347). Notice that the variables in x_t are not relevant covariates if $\beta(L) = 0$. ◊

Remark 2.2. Following the bulk of the econometric bubble detection literature, we model asset prices with the time-varying AR model in (1)-(2). As discussed in PWY and Breitung and Kruse (2013), *inter alia*, this is often motivated as an approximation to the rational bubble model where the observed asset price, y_t, is equal to the sum of the fundamental price, f_t, of the asset, assumed to be a martingale ($I(1)$) process, and a bubble component, B_t, which is zero other than in its bubble phase when it is a submartingale (explosive $AR(1)$ process). Under Assumption 1, the error term, v_t, in (1)-(2) is related to a set of covariates. This therefore entails the implicit assumption that the covariates would be related to both f_t and B_t in the rational bubble model. It is, however, possible that a given covariate could be related to only the error term driving one of these components. If this were the bubble component then, as noted by a referee, we would not expect any power gains from incorporating that covariate into the CUSUM bubble detection procedure. ◊

Under the null hypothesis $H_0 : \delta = 0$, we have that $\Delta y_t = v_t$ for the full sample period $t = 1, \ldots, \lfloor \lambda T \rfloor$, and so from (4) we then have that

$$\Delta y_t = \mu + \sum_{k=1}^{p} \alpha_k \Delta y_{t-k} + \sum_{k=0}^{q} \beta_k' x_{t-k} + \varepsilon_t, \tag{5}$$

where $\mu := -\sum_{k=0}^{q} \beta_k' c_x$ and where the first summation term is understood to be present only when $p > 0$. Notice that the intercept term $\mu = 0$ if either $c_x = 0$, such that the covariates have mean zero, or $\beta(L) = 0$, such that x_t are not relevant covariates.[2] This is a heteroskedastic autoregressive model in Δy_t augmented by the level and (up to) q lags of the m covariates. Defining $g_t := (1, \Delta y_{t-1}, \ldots, \Delta y_{t-p}, x_t', x_{t-1}', ..., x_{t-q}')'$ and $\varphi := (\mu, \alpha_1, \ldots, \alpha_p, \beta_0', \beta_1', \ldots, \beta_q')'$, the null model (5) can be written more compactly as

$$\Delta y_t = \varphi' g_t + \sigma_t \eta_t, \quad t = 1, \ldots, T, \ldots, \lfloor \lambda T \rfloor \tag{6}$$

For the subsequent analysis, we need to formalise our assumptions on the covariates, x_t, and the other elements comprising (4). These are now stated in Assumption 2, with some discussion of these conditions then given in Remarks 2.3-2.8.

Assumption 2. *Let the $\{(\eta_t, w_t)\}$ sequence be defined on a complete probability space, and denote the natural filtration generated by the random vector sequence $\{(\eta_t, w_{t+1})\}$ by $\{\mathcal{F}_t\}$. Assume that:*

(a) For $t = 1, \ldots, T, \ldots, \lfloor \lambda T \rfloor$, $\sigma_t = \sigma(t/T)$ where the function $\sigma(\cdot)$ is non-stochastic, has support $[0, \lambda]$, is differentiable, is uniformly bounded by a constant M, and is such that $\sigma(.) \geqslant \epsilon^*$, for some $\epsilon^* > 0$. Furthermore, the derivative of $\sigma(\cdot)$ is Lipschitz continuous over $(0, \lambda)$.

(b) Let η_t be a martingale difference sequence [MDS] with respect to the filtration $\mathcal{F}_t$, with conditional variance $h_t := E(\eta_t^2 | \mathcal{F}_{t-1}) > 0$ satisfying the condition that $E(h_t) = \operatorname{plim}_{T\to\infty}(1/\lfloor T\lambda \rfloor) \sum_{t=1}^{\lfloor T\lambda \rfloor} h_t = 1$.

(c) $\{\eta_t\}$ is a strong mixing process with mixing coefficients of size $-r/(r-2)$, for some $r > 2$, and $E|\eta_t|^{2r} < \infty$.

(d) $\alpha(z) \neq 0$ for all $|z| \leqslant 1$.

(e) For all $0 \leq \kappa \leq \lambda$: $\operatorname{plim}_{T\to\infty}(1/\lfloor T\kappa \rfloor) \sum_{s=1}^{\lfloor T\kappa \rfloor} g_s g_s'$ is positive definite with finite elements; $\operatorname{plim}_{T\to\infty}(1/\lfloor T\kappa \rfloor) \sum_{s=1}^{\lfloor T\kappa \rfloor} g_s g_s'/\sigma_s^2 = \lim_{T\to\infty} (1/\lfloor T\kappa \rfloor)\, E(\sum_{s=1}^{\lfloor T\kappa \rfloor} g_s g_s'/\sigma_s^2) =: \Theta(\kappa)$, where $\Theta(\kappa)$ is a positive definite matrix with all elements finite and continuous in κ. Furthermore, we assume that the covariances between h_t and g_t and between h_t and $g_t g_t'$ are zero, for all $t = 1, \ldots, T, \ldots, \lfloor \lambda T \rfloor$.

[2] The constant term in (5) entails that statistics based on the residuals from estimating this model will be exact invariant to a non-zero mean, should it be present, in Δy_t, and hence to a linear trend in y_t.

(f) *The vector* w_t *satisfies* $\limsup_{T\to\infty} \frac{1}{\lfloor T\lambda \rfloor} \sum_{t=1}^{\lfloor T\lambda \rfloor} E\|w_t\|^{2+\delta} < \infty$, *for some* $\delta > 0$, *where* $\|\cdot\|$ *denotes the Euclidean norm.*

Remark 2.3. The monitoring procedure of HB assumes v_t is homoskedastic, while ATKK allow for conditional heteroskedasticity, but impose unconditional homoskedasticity, in the context of their covariate augmented PSY and PWY tests. These assumptions are arguably rather strong given that time-varying volatility appears to be a common feature in many financial time series. For example, many empirical studies report strong evidence of structural breaks in the unconditional variance of asset returns; see, among others, McMillan and Wohar (2011), Calvo-Gonzalez *et al.* (2010), and Vivian and Wohar (2012). To allow for such features, Assumption 2(a), which coincides with Assumption 2 of AHLTZ, specifies the unconditional volatility function of the regression errors, σ_t, to have a flexible nonparametric structure which allows for, *inter alia*, smooth transition breaks in volatility and trending volatility. The case of constant volatility, where $\sigma_t = \sigma$, for all t, also satisfies Assumption 2(a) with $\sigma(s) = \sigma$, for all $s \in [0, \lambda]$. Although discrete jumps in volatility are not formally allowed under Assumption 2(a), this is not restrictive in practice because one can always approximate discontinuities in $\sigma(\cdot)$ arbitrarily well using smooth transition functions.[3] $\diamondsuit$

Remark 2.4. Assumption 2(b) specifies that η_t is a conditionally heteroskedastic MDS. Allowing for conditional heteroskedasticity is desirable with financial data and, hence, this represents an important relaxation of the conditions required by AHLTZ who impose conditional homoskedasticity on their equivalent of η_t in their Assumption 1. The MDS condition in Assumption 2(b) implies that the exogeneity condition $E(g_t\eta_t) = 0$ holds. Assumption 2(c) additionally imposes that η_t is strong mixing. This assumption is made because we need to restrict the amount of dependence in $\{\eta_t^2 - 1\}$ (this process no longer being a MDS when conditional heteroskedasticity is present in η_t) for the purposes of estimating the unconditional volatility function, σ_t; see, for example, Lemma A.3 in the appendix. The final condition in Assumption 2(e) rules out any correlation between the regressors in (6), g_t, and the conditional variance of η_t and also rules out correlation between

[3]Under Assumption 2(a), σ_t depends on T, and as such $\{y_t\}$ formally constitutes a triangular array of the type $\{y_{T,t} : t = 0, 1, .., \lfloor \lambda T \rfloor; T = 0, 1, ...\}$. However, because the triangular array notation is not essential, the subscript T will be suppressed in our exposition.

the elements of the design matrix $g_t g_t'$ and the conditional variance of η_t. Where η_t is conditionally homoskedastic, this condition is rendered redundant. Moreover, this condition is also not needed in the case where an intercept term is not included in (5); see section A.1 of the supplementary appendix. ♢

Remark 2.5. Assumption 2(d) rules out the presence of unit or explosive autoregressive roots in Δy_t under the null hypothesis. Assumption 2(e) allows the covariance matrix of the covariates to display very general patterns of time-variation. This condition is weaker than the conditions placed on σ_t under Assumption 2(a) because any heteroskedasticity arising from the covariates does not show up in the limiting null distribution of the CUSUM statistics that our monitoring procedure is based on and, hence, does not need to be estimated or corrected for. Notice that time-variation in the correlation between ε_t and the covariates is also permitted. ♢

Remark 2.6. Under Assumptions 2(e) and 2(f) we can make use of the weak convergence result established in Lemma A.10 in the supplementary appendix, which is an extension of Lemma 3 of KPA to our context and plays an important role in the proof of our main results. In Assumption 2(e), the condition that $\operatorname*{plim}_{T\to\infty}(1/\lfloor T\kappa\rfloor)\sum_{t=1}^{\lfloor T\kappa\rfloor} g_t g_t'$ is positive definite with finite elements rules out the possibility of asymptotic collinearity between the regressors in g_t. Taken together with the exogeneity condition implied by Assumption 2(b), this ensures least squares [LS] estimation of φ in Lemma A.1 is consistent under the null hypothesis, $H_0 : \delta = 0$. Likewise, the analogous condition on $\operatorname*{plim}_{T\to\infty}(1/\lfloor T\kappa\rfloor)\sum_{t=1}^{\lfloor T\kappa\rfloor} g_t g_t'/\sigma_t^2$ is required in the context of weighted least squares [WLS] estimation of φ; see Lemma A.6. ♢

Remark 2.7. An analogous moment condition to Assumption 2(f) is imposed for all the covariates (and the error terms) in KPA; notice that we do not need to directly impose this condition on the lagged differences Δy_{t-k}, $k = 1, ..., p$, in our regression model in (6), because Assumption 2(c) implies that the lagged differences will satisfy an equivalent moment condition, which is stronger than Assumption 2(f). The stronger moment condition in Assumption 2(c) is needed for the proof of Lemma A.3 in the appendix, which is required in connection with estimation of the (unknown) variance function, σ_t^2. ♢

Remark 2.8. Our specification for the covariates is more general than is imposed by KPA, who impose a global homoskedasticity assumption, or by Hansen (1995), Chang *et al.* (2017) [CSS] and ATKK in the context of their covariate unit root testing methods. For example, the (covariance) stationarity assumption required to hold on the covariates by Hansen (1995) is not imposed by our assumptions as we allow for unconditional heteroskedasticity. Moreover, a version of the unconditionally homoskedastic finite-order stationary vector autoregressive model specified for the covariates in CSS and ATKK, generalised to allow for the possibility of unconditional heteroskedasticity, is also permitted under our assumptions. The assumption made in Hansen (1995), CSS and ATKK that the covariates are weakly dependent is not required for our analysis, albeit the strength of dependence allowed is restricted by Assumption 2(e) which, for example, rules out covariates with (near-) unit roots. As argued in Hansen (1995), in many cases the first differences of relevant financial and/or macroeconomic time series will be natural covariates to consider. $\diamondsuit$

3 CUSUM-based Bubble Detection Procedures

Under the assumption that v_t in (2) is a mean zero, serially uncorrelated and conditionally homoskedastic process with unconditional variance σ^2, and for a training sample $t = 1, ..., T$, as in (1)-(2), HB propose testing for explosive behaviour in the monitoring period using the CUSUM statistic:

$$S_T^t := \frac{1}{\tilde{\sigma}_t} \sum_{j=T+1}^{t} \Delta y_j \tag{7}$$

where $t > T$ is the monitoring observation. In (7), $\tilde{\sigma}_t^2$ is an estimate of σ^2 which is consistent under H_0; HB use $\tilde{\sigma}_t^2 := (t-1)^{-1} \sum_{j=2}^{t} (\Delta y_j)^2$. If S_T^t is computed sequentially at dates $t = T+1, ..., \lfloor \lambda T \rfloor$, then under the null hypothesis, H_0, of no explosive behaviour, as $T \to \infty$,

$$T^{-1/2} S_T^{\lfloor Tr \rfloor} \Rightarrow W(r) - W(1), \quad 1 < r \leq \lambda \tag{8}$$

where "$\Rightarrow$" denotes weak convergence of the associated probability measures, and where $W(\cdot)$ is used generically to denote a standard Brownian motion defined on the interval $[0, \lambda]$.

Using Theorem 3.4 of Chu *et al.* (1996), HB show that under H_0, the result in (8) implies that, for any $\lambda > 1$,

$$\lim_{T \to \infty} \Pr\left(|S_T^t| > c_t \sqrt{t} \text{ for some } t \in \{T+1, ..., \lfloor \lambda T \rfloor\} \right) \leq \exp\left(-b_\alpha/2\right) \tag{9}$$

where $c_t := \sqrt{b_\alpha + log(t/T)}$. The CUSUM monitoring procedure proposed in HB then rejects H_0 if $S_T^t > c_t\sqrt{t}$ for some $t > T$, with an explosive episode signalled at the first time point t in the monitoring period for which such an exceedance occurs.[4] For such a (one-sided upper tail) test the appropriate asymptotic setting for b_α used to compute c_t that would deliver size of at most $\alpha = 0.05$ would be $b_\alpha = 4.6$ (as this value of b_α would deliver a two-sided test with size at most $\alpha = 0.10$ from the result in (9)).[5]

Astill *et al.* (2018) show that the procedure based on S_T^t does not have a controlled FPR, even in large samples, in the case where $v_t = \sigma_t \epsilon_t$ with the volatility function, σ_t, displaying time-variation of the form specified by Assumption 2(a) and ϵ_t a MDS with unit conditional variance. Based on this, AHLTZ replace S_T^t with the modified CUSUM statistic,

$$SV_T^t := \sum_{j=T+1}^{t} \frac{\Delta y_j}{\hat{\sigma}_{j,N}}, \quad t > T \tag{10}$$

where $\hat{\sigma}_{j,N}^2$ is a kernel smoothing estimator for the spot variance $\sigma_j^2 := \sigma^2(j/T)$, defined, for $j \geqslant N+1$, as

$$\hat{\sigma}_{j,N}^2 := \sum_{s=0}^{N} k_s \left(\Delta y_{j-s}\right)^2, \quad \text{with} \quad k_s := \frac{K\left(\frac{s}{N}\right)}{\sum_{s=0}^{N} K\left(\frac{s}{N}\right)} \tag{11}$$

where the kernel function, $K(\cdot)$, and bandwidth, N, satisfy the conditions stated in Assumption 3, below. AHLTZ establish that the CUSUM monitoring procedure based on SV_T^t is able to control the FPR when v_t exhibits time varying volatility of the form specified in Assumption 2(a), while retaining power close to the standard CUSUM procedure of HB when the innovations are homoskedastic.

[4]Notice that the upper tail decision rule implies that the CUSUM procedure is designed to pick up positive asset price bubbles, but will not reject against negative price bubbles. A version of the procedure designed to detect the latter could be developed by using the corresponding lower tail decision rule, while a detection procedure for either type of bubble would use the corresponding two tail decision rule.

[5] These asymptotic settings for b_α assume a monitoring period of infinite length, and monitoring procedures based on these settings for b_α can be extremely conservative in practice, particularly during the early stages of the monitoring period. HB, therefore, provide finite sample settings in their paper (Table 8, p221), reporting values of b_α that deliver a monitoring procedure with an expected FPR of $\alpha \in \{0.10, 0.05, 0.01\}$ by the end of the monitoring period for various lengths of the training and monitoring period, assuming the series y_t is an exact unit root process driven by NIID(0,1) innovations.

Henceforth, we will refer to a monitoring procedure based on the S_T^t statistic as the (standard) CUSUM monitoring procedure and that based on the SV_T^t statistic as the CUSUM^V monitoring procedure.

The validity of both CUSUM and CUSUM^V relies on the assumption that Δy_t is serially uncorrelated under H_0. This assumption is obviously violated if v_t is generated by (4) with $p > 0$, but is also, in general, violated (even if $p = 0$) when $\beta(L) \neq 0$ if, for example, either the covariates, x_t, are serially correlated, or $q > 0$, or both. The large sample results in (8) and (9) will not hold for S_T^t or SV_T^t in such cases. Consequently implementing CUSUM and CUSUM^V using the critical values from HB would result in monitoring procedures where the (theoretical) FPR would not be at the level expected by the practitioner. We next develop covariate augmented analogues of the CUSUM and CUSUM^V procedures which account for the influence of the covariates x_t, as well as any serial correlation arising from $\alpha(L)$. These will be shown to retain the large sample results in (8) and (9). Later, in section 5, we will use Monte Carlo simulation to investigate the degree of spurious detections suffered by the univariate procedures when covariates are present in the DGP, and show that these are well controlled by the covariate augmented procedures.

4 A Covariate Augmented CUSUM Monitoring Procedure

CUSUM tests for structural change in the parameters of homoskedastic weakly dependent dynamic regression models have been developed in KPA who base their approach on a statistic constructed from a standardised cumulated sum of recursive LS residuals. We will adapt this approach to our setting to develop a real-time bubble monitoring procedure which has a theoretically controlled FPR when v_t is generated according to (4). We discuss the construction of the CUSUM monitoring statistic by first considering the infeasible case where the volatility function, σ_t, is known, and then discuss the feasible version of this, based on nonparametric estimation of σ_t.

A key difference between our setting and that considered in KPA is that we allow for the presence of heteroskedasticity in both the covariates, x_t, and in disturbances, ε_t, in the null regression (5), of the form specified in Assumption 2. Except in the special case where the intercept term is excluded from the null regression (recall that this may be done where the covariates all have mean zero), which is discussed separately in Section A.1 of the supplementary appendix, the presence of unconditional heteroskedasticity necessitates constructing the CUSUM monitoring statistics from recursive WLS residuals, rather than the conventional recursive LS residuals which suffice under unconditional homoskedasticity. It is also worth clarifying at this point that the methods outlined in this section apply provided that the vector of regression variables, g_t, in the null regression model, (6), contains at least one element (even if this is just an intercept term). Where this is not the case, no regression estimation is needed and the appropriate monitoring procedure is that given in section 2.2 of AHLTZ.

Our proposed CUSUM monitoring statistic is based on recursive WLS estimation of the (null) regression in (6), which contains $1 + p + (q + 1)m$ regressors. To that end, consider the infeasible WLS transformation of (6), based on the true volatility function σ_t, given by

$$\frac{\Delta y_t}{\sigma_t} = \varphi' \frac{g_t}{\sigma_t} + \eta_t, \quad t = 1, \ldots, T, \ldots, \lfloor \lambda T \rfloor. \tag{12}$$

The (infeasible) WLS estimator for φ at time t in the monitoring sample from this regression is then given by

$$\varphi_t^W := \left(\sum_{j=\max(p+2,q+1)}^{t} \frac{g_j g_j'}{\sigma_j^2} \right)^{-1} \left(\sum_{j=\max(p+2,q+1)}^{t} \frac{g_j \Delta y_j}{\sigma_j^2} \right), \quad t = T+1, \ldots, \lfloor \lambda T \rfloor$$

with the associated (infeasible) recursive residuals based on the WLS estimate defined as

$$e_t^W := \Delta y_t - (\varphi_{t-1}^W)' g_t, \quad t = T+1, \ldots, \lfloor \lambda T \rfloor. \tag{13}$$

It is established in the proof of Theorem 1 that, under the null hypothesis, the associated infeasible sequence of CUSUM statistics $SWM_T^t := \sum_{j=T+1}^{t} e_j^W / \sigma_j$, $t = T+1, ..., \lfloor \lambda T \rfloor$, satisfies $T^{-1/2} SWM_T^{\lfloor Tr \rfloor} \Rightarrow W(r) - W(1)$, $1 < r \leq \lambda$, where it is recalled that $W(\cdot)$ generically denotes a standard Brownian motion on $[0, \lambda]$, such that we recover the usual limiting distribution in (8).

To obtain a feasible version of SWM_T^t we need to replace σ_j by a nonparametric estimate thereof. Nonparametric estimation of the variance function in time series models has been considered by, among others, Xu and Phillips (2008), Cavaliere *et al.* (2017) and Harvey *et al.* (2019), whereby a nonparametric kernel smoothing estimation procedure is applied to the squares of regression residuals from the model at hand. In the present real-time monitoring setting, however, nonparametric estimation of the variance function is nonstandard in two ways. First, because the monitoring takes place in real-time, only data up to and including each time point in the monitoring period will be available to the practitioner, and so as a consequence the smoothing is naturally performed using a one-sided kernel. Second, because new data will continue to arrive in real-time as the monitoring proceeds, the vector of regression residuals needs to be updated at each successive time point in the monitoring period.

As a consequence of the second issue discussed above, we will need to make use of the double array of ordinary least squares [OLS] residuals from estimating (5), defined as:

$$f_{i,t}^* := \Delta y_i - (\hat{\varphi}_t)' g_i, \quad i = \max(p+2, q+1), \ldots, t, \; t = T+1, \ldots, \lfloor \lambda T \rfloor \tag{14}$$

where

$$\hat{\varphi}_t := \left(\sum_{j=\max(p+2,q+1)}^{t} g_j g_j' \right)^{-1} \left(\sum_{j=\max(p+2,q+1)}^{t} g_j \Delta y_j \right), \quad t = T+1, \ldots, \lfloor \lambda T \rfloor. \tag{15}$$

Using the OLS residuals in (14), we can then define the sequence of nonparametric variance estimators across times $j = N + \max(p+1, q), ..., t$, when standing at time t, as

$$\tilde{\sigma}_{j,N,t}^2 := \sum_{s=0}^{N} k_s (f_{j-s,t}^*)^2, \quad k_s := \frac{K\left(\frac{s}{N}\right)}{\sum_{s=0}^{N} K\left(\frac{s}{N}\right)}. \tag{16}$$

in which k_s, $s = 0, ..., N$, is a sequence of weights, which are defined based on some kernel function $K(\cdot)$ and a window size N, precise conditions on which will be given in Assumption 3, below. Because of the unavailability of future data, this nonparametric variance estimator uses a left-sided, truncated kernel. Only the N most recent observations are used in the calculation of the estimator and the weights are not dependent on t.

Based on the nonparametric variance estimates in (16), we can then define the feasible WLS estimator of φ at time t as,[6]

$$\hat{\varphi}_t^W := \left(\sum_{j=N+\max(p+1,q)}^{t} \frac{g_j g_j'}{\tilde{\sigma}_{j,N,t}^2} \right)^{-1} \left(\sum_{j=N+\max(p+1,q)}^{t} \frac{g_j \Delta y_j}{\tilde{\sigma}_{j,N,t}^2} \right), \quad t = T+1, \ldots, \lfloor \lambda T \rfloor.$$

[6]The change in the lower summation indices, relative to φ_t^W, arises because the calculation of $\hat{\varphi}_t^W$ requires variance estimates which can only be computed from $j = N + \max(p+1, q)$ onwards.

Defining the feasible WLS recursive residuals as

$$\hat{e}_j^W := \Delta y_j - (\hat{\varphi}_{j-1}^W)' g_j, \quad j = T+1, \dots, \lfloor \lambda T \rfloor$$

a feasible version of the sequence of SWM_T^t statistics can then defined as,

$$SWMV_T^t := \sum_{j=T+1}^{t} \frac{\hat{e}_j^W}{\tilde{\sigma}_{j,N,j}}, \qquad t = T+1, ..., \lfloor \lambda T \rfloor \tag{17}$$

We will denote the monitoring procedure based on the sequence of $SWMV_T^t$, $t = T + 1, ..., \lfloor \lambda T \rfloor$, statistics as CUSUM^{WMV}.

In order to derive the asymptotic properties of the sequence of $SWMV_T^t$ statistics, we require the following conditions hold on the kernel function $K(\cdot)$ and the window size N. These conditions coincide with those imposed by AHLTZ (p.194) in the context of their SV_T^t statistic in (10), where a discussion of these conditions is provided.

Assumption 3. *(a) $K(\cdot)$ is strictly positive and continuously differentiable over the interval $(0,1)$, with $K(x) = 0$ for $x \leq 0$ and $x \geq 1$. Also, $\int_0^1 K(x)\mathrm{d}x > 0$, $\int_0^1 |K(x)|\mathrm{d}x < \infty$, $\int_0^1 |K(x)x|\mathrm{d}x < \infty$ and the characteristic function $\phi(t) = \int_{-\infty}^{\infty} \exp(\mathrm{i}tx)K(x)\mathrm{d}x$ of K satisfies $\int_{-\infty}^{\infty} |\phi(t)|\mathrm{d}t < \infty$. $K'(\cdot)$, the derivative of the $K(\cdot)$ function, also has a characteristic function that is absolutely integrable.*

(b) $N \to \infty$ as $T \to \infty$, such that $N/T \to 0$ and $N^{3/2}/T \to \infty$.

Remark 4.1. Implementation of $SWMV_T^t$ requires choices to be made for both the kernel and bandwidth used in constructing the nonparametric estimator $\tilde{\sigma}_{j,N,t}^2$ in (16). We found that the choices for these recommended in AHLTZ also lead to good FPR control for the procedures considered in this paper. Specifically, we therefore recommend implementation with the truncated Gaussian kernel and where the bandwidth at each point t in the monitoring period, denoted N_t^{cv}, is chosen according to the automated rule:

$$N_t^{cv*} := \mathrm{argmin}_{N \in [1,H]} CV_t^*(N), \qquad CV_t^*(N) := \frac{1}{H} \sum_{j=t-H+1}^{t} (\tilde{\sigma}_{j,N,t}^2 - (f_{j,t}^*)^2)^2. \tag{18}$$

where, for $j = t - H + 1, ..., t$,

$$\tilde{\sigma}_{j,N,t}^2 := \sum_{s=0}^{N} k_s (f_{j-s,t}^*)^2, \quad k_s := \frac{K\left(\frac{s}{N}\right)}{\sum_{s=0}^{N} K\left(\frac{s}{N}\right)}, \tag{19}$$

The estimators of the spot variances, σ_j^2, $j = t-H+1, ..., t$, each computed at time t, defined in (19) are needed to compute the time t cross-validation objective function in (18). The automated bandwidth rule minimises the estimation error of the spot variance over the most recent H observations based on the OLS residuals computed using data up to and including the current monitoring observation, t; cf. Hall and Schucany (1989). Implementation of N_t^{cv*} in (18) requires a choice of H; we follow AHLTZ and set $H = 20$. These choices for the kernel and bandwidth are used in all the numerical work in this paper. $\diamondsuit$

In Theorem 1, we establish the joint limiting null distribution of the sequence of feasible covariate augmented $SWMV_T^t$ statistics from the monitoring period.

Theorem 1. *Let the data be generated according to (1)-(4) under the null hypothesis* $H_0 : \delta = 0$*. If Assumptions 1-3 hold, then, as* $T \to \infty$*, it follows that*

$$T^{-1/2} SWMV_T^{\lfloor Tr \rfloor} \Rightarrow W(r) - W(1), \quad 1 < r \leq \lambda. \tag{20}$$

Appealing to Theorem 3.4 of Chu *et al.* (1996), Theorem 1 implies the following:

Corollary 1. *Under the conditions of Theorem 1,*

$$\lim_{T \to \infty} \Pr \left(|SWMV_T^t| > c_t \sqrt{t} \text{ for some } t \in \{T+1, ..., \lfloor \lambda T \rfloor\} \right) \leq \exp(-b_\alpha/2). \tag{21}$$

Remark 4.2. Theorem 1 and Corollary 1 imply that when the innovations v_t satisfy Assumptions 1-2, both the limiting null distribution and crossing probabilities for the covariate augmented CUSUM^{WMV} procedure are unchanged relative to those given in (8) and (9), respectively, for the original CUSUM procedure of HB in the case where v_t is conditionally homoskedastic and serially uncorrelated. Notice from (20) that the joint limiting null distribution of the $SWMV_T^t$, $t > T$, statistics does not depend on any nuisance parameters arising from time-varying behaviour in the unconditional covariance matrix of the covariates; cf. Remark 2.5. $\diamondsuit$

Next, we proceed to establish consistency results for our covariate-augmented CUSUM^{WMV} monitoring procedure. In Theorem 2 we establish consistency results for a class of mildly explosive alternatives of the form $\delta = c/T^d$ with $0 < d \leq 2/3$, for $t > \lfloor \tau T \rfloor$, and where c is a positive constant, and for fixed alternatives, $\delta = c$. We will subsequently discuss the class of mildly explosive alternatives where $2/3 < d < 1$ in Remark 4.3, and locally explosive alternatives, where $d = 1$, in Remark 4.4.

Theorem 2. *Let the data be generated according to (1)-(4) under the alternative hypothesis $H_1 : \delta = c/T^d$, for $t > \lfloor \tau T \rfloor$, with c a positive constant and $0 \leq d \leq 2/3$, and let Assumptions 1-3 hold. It then holds that,*

$$\lim_{T\to\infty} \Pr\left(|SWMV_T^t| > c_t\sqrt{t}, \text{for some } t \in \{\lfloor \tau T \rfloor + 1, \ldots, \lfloor \lambda T \rfloor\}\right) = 1. \tag{22}$$

Remark 4.3. The result in Theorem 2 immediately implies that the CUSUM^{WMV} procedure is consistent against both fixed ($d = 0$) and mildly explosive ($0 < d \leq 2/3$) alternatives of the form $\delta = c/T^d$. In both these cases T^d maintains a fixed relative relationship with N. Recall that Assumption 3(b) imposes the condition that $N^{3/2}/T \to \infty$, which implies that $N/T^{2/3} \to \infty$. Consequently, when $0 \leq d \leq 2/3$, T^d diverges at a slower rate than N and $T^d \wedge N = T^d$. However, in cases where $2/3 < d < 1$, such that the magnitude of the explosiveness parameter is very mild, this no longer holds and, as a result, $SWMV_T^t$ does not necessarily diverge at a faster rate than the boundary function $c_t\sqrt{t}$. Essentially this issue arises because the volatility estimates in (16) are constructed using the residuals from a regression model which imposes the null hypothesis. Where the null is false, this model is misspecified and for $2/3 < d < 1$ the volatility estimate diverges at such a rate that it prevents CUSUM^{WMV} from necessarily diverging at a faster rate than the boundary function $c_t\sqrt{t}$; see Lemma A.9 in the supplementary appendix. A possible solution to this is to employ a truncated volatility estimator of the form, $\tilde{\sigma}_{j,N,j} \cdot \mathbb{I}(\tilde{\sigma}_{j,N,j} \leq \mathbb{C}\ln(T)) + \mathbb{C}\ln(T) \cdot \mathbb{I}(\tilde{\sigma}_{j,N,j} > \mathbb{C}\ln(T))$, where $\mathbb{C}$ is a generic positive constant, such that $\mathbb{C}\ln(T)$ serves as a slowly varying truncation function. Under the null, the volatility estimator is consistent and the truncation level $\ln(T)$ approaches infinity, such that the truncation has no impact in the limit. However, under the alternative the truncated volatility estimator is limited to diverge at a rate no faster than $\ln(T)$, which is slower than any polynomial rate. By incorporating this truncation mechanism, we conjecture that consistency would hold over a wider range of d than $0 \leq d \leq 2/3$. However, we leave a detailed treatment of this case for future research. $\diamondsuit$

Remark 4.4. In addition to the consistency results in Theorem 2, it is also instructive to examine the behaviour of the monitoring procedure in the case of locally explosive alternatives of the form $H_{c,\tau} : \delta = c/T$, for $t > \lfloor \tau T \rfloor$, where c is a positive constant. When the volatility process is known, the asymptotic behaviour of the detector $SWMV_T^t$ can be derived along the same line of argument as the proof of Theorem 1. In particular, in the special case of $\alpha(L) = 1$ (i.e. when the fitted model has no lagged dependent variables),

$$\frac{1}{\sqrt{T}} \sum_{j=T+1}^{\lfloor Tr \rfloor} \frac{e_j^W}{\sigma_j} = \frac{1}{\sqrt{T}} \sum_{j=T+1}^{\lfloor Tr \rfloor} \eta_j - \frac{1}{\sqrt{T}} \sum_{j=T+1}^{\lfloor Tr \rfloor} \frac{(\varphi_{j-1}^W - \varphi)' g_j}{\sigma_j} + \frac{c}{T^{3/2}} \sum_{j=\lfloor \tau T \rfloor + 1}^{\lfloor Tr \rfloor} \frac{u_{j-1}}{\sigma_j}$$

As in the proof of Theorem 1, the first two terms collectively weakly converge to $W(r) - W(1)$. By the FCLT and CMT, the third term satisfies $\frac{c}{T^{3/2}} \sum_{j=\lfloor \tau T \rfloor + 1}^{\lfloor Tr \rfloor} \frac{u_{j-1}}{\sigma_j} \Rightarrow c \int_\tau^r U(s)/\sigma(s)\mathrm{d}s$, where $U(s) := \int_0^s e^{c(s-u)} \sigma(u)\mathrm{d}W(u)$. It therefore follows that the asymptotic distribution under $H_{c,\tau}$ is given by $W(r) - W(1) + c\int_\tau^r U(s)/\sigma(s)\mathrm{d}s$, from which the asymptotic probability of the CUSUM^{WMV} procedure rejecting the null when a locally explosive episode is present can be simulated. For general $\alpha(L)$, it can be shown in the same way that the asymptotic distribution is given by $W(r) - W(1) + \alpha(1)c\int_\tau^r U(s)/\sigma(s)\mathrm{d}s$. Where the volatility is estimated, we anticipate the same limit will hold under $H_{c,\tau}$ in view of the results given in Harvey *et al.* (2019) for the behaviour of the nonparametric variance estimator considered in this paper under locally explosive DGPs. ♢

Remark 4.5. Thus far we have assumed that the parameters p and q in (5), together with the composition of the m-vector of true covariates, x_t, are known. In practice, these aspects will be unknown. However, under the maintained hypothesis of no bubble in the training sample, the regression model in (5), for $t = 1, ..., T$, is an ARX model satisfying standard regularity conditions, and so an application of a consistent information criterion [IC], such as the well-known Bayesian IC [BIC], could be used to select these elements. The Monte Carlo results in section 5 will implement applying the BIC to the training sample to select p, q, and whether to include a given candidate covariate or not. ♢

We end this section with a word of caution. The CUSUM^{WMV} procedure can, in principle, reject for various forms of structural change in the null model that, while ruled by our regularity conditions, might occur in practice. As such, a rejection by CUSUM^{WMV} does not necessarily imply the presence of a bubble episode. Indeed, this is precisely our motivation for developing a procedure robust to structural changes in unconditional volatility. Another possibility is where a covariate used in the null regression displays structural change, such as an explosive episode itself or a mean shift; simulations looking at these cases are reported in Section 5.3. In practice, as with any statistical procedure, we recommend practitioners investigate the plausibility of the regularity conditions underlying CUSUM^{WMV} as part of their statistical analysis. This could, for example, include running standard tests for explosivity and mean shifts in the covariates over the training sample and then running analogous (univariate) CUSUM monitoring procedures in tandem on the covariates, removing any covariate from the analysis for which either of these reject.

5 Monte Carlo Simulations

We report results of a Monte Carlo simulation exercise evaluating the finite sample performance of the CUSUM^{WMV} monitoring procedure. Additional results are reported in the supplementary appendix and summarised in Section 5.3.

5.1 Simulation DGP and Experimental Settings

Data were generated according to (1)-(2), initialised at $u_0 = 100$ (so that bubbles in our series are generally upwardly explosive and, hence, empirically relevant), setting $\mu = 0$ without loss of generality. We set $T = 219$, so that monitoring begins at time $t = 220$, and set monitoring to end at time $\lambda T = 255$. Under the null $\delta = 0$, while under the alternative we set $\delta = 0.005$, $\tau_1 T = 220$ and $\tau_2 T = \lambda T$, such that y_t follows a unit root process during the training sample, before switching to an explosive regime starting when monitoring commences and continuing until the end of the monitoring period.

For the error term v_t and the covariate x_t, we use an unconditionally heteroskedastic extension of the simulation DGP detailed in Section 5.1 on page 143 of CSS:

$$v_t = \alpha_1 v_{t-1} + \beta x_t + \varepsilon_{1,t}, \tag{23}$$

$$x_{t+1} = \rho x_t + \varepsilon_{2,t}, \tag{24}$$

with the covariate initialised at $x_0 = 0$. The variance matrix of the innovation vector, $(\varepsilon_{1,t}, \varepsilon_{2,t})'$, was generated according to:

$$\begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} \sim NIID(0, \Sigma_t), \qquad \Sigma_t := \begin{bmatrix} \sigma_{1,t}^2 & \sigma_{12,t} \\ \sigma_{12,t} & \sigma_{2,t}^2 \end{bmatrix} \tag{25}$$

in which $\sigma^2_{1,t}$, $\sigma^2_{2,t}$ are subject to smooth upward shifts in volatility of the form:

$$\sigma_{j,t} \quad := \quad 1 + (\sqrt{4} - 1)\left[1 + \exp(-\theta(t - 219))\right]^{-1}, \qquad j = 1, 2 \tag{26}$$

with $\theta = 0.25$; that is, a logistic smooth transition in volatility from 1 to $\sqrt{4}$ centred on the end of the training sample. We report results for the following four cases for Σ_t:

(a) $\sigma^2_{1,t} = \sigma^2_{2,t} = 1$ and $\sigma_{12,t} = \sigma_{12}$, in each case for all t, such that $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ are homoskedastic with a fixed correlation of σ_{12}.

(b) $\sigma_{1,t}$ and $\sigma_{2,t}$ both satisfy (26), while $\sigma_{12,t} = \sigma_{12}\sigma_{1,t}\sigma_{2,t}$, such that the correlation between $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ remains fixed at σ_{12} for all t.

(c) $\sigma^2_{1,t}$ satisfies (26), $\sigma_{2,t} = 1$, for all t, and $\sigma_{12,t} = \sigma_{12}\sigma_{1,t}$, such that $\varepsilon_{1,t}$ exhibits time varying volatility, but with the correlation between $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ fixed at σ_{12}.

(d) $\sigma^2_{1,t}$ satisfies (26), $\sigma_{2,t} = 1$, for all t, and $\sigma_{12,t} = \sigma_{12}$, such that $\varepsilon_{1,t}$ exhibits time varying volatility with the correlation between $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ time-varying through $\sigma^2_{1,t}$.

We report rejection rates for the CUSUM^{WMV} procedure together with the standard CUSUM procedure of HB and the CUSUM^{V} procedure of AHLTZ. We also report results for a procedure, denoted CUSUM^{V*}, which is similar to the CUSUM^{WV} procedure outlined in Section A.1 but where the null regression is given by (5) but excluding the covariate regressors and the intercept. The rationale behind including this procedure is that including only lags of Δy_t should yield a procedure that is able to deal with the serial correlation in Δy_t induced by the presence of the covariate (see the discussion at the end of Section 3), but does not exploit any potential power gains available from including a relevant covariate under the alternative. It should therefore provide an FPR controlled benchmark against which to quantify the power gains (or losses) that arise from including the covariate terms.[7]

Following the discussion in Remark 4.5, in implementing the CUSUM^{WMV} procedure we use the BIC to select the null model, based on OLS estimation and using only the training sample data. The BIC is computed for (5), estimated using a common data sample ending

[7]Note that the null regression used for CUSUM^{V*} does not contain an intercept as, when excluding the covariates from the regression, an intercept is only needed if we wish to allow for a trend in y_t under the null. The statistic for this procedure is therefore computed as in Section A.1 in the Supplementary Appendix where WLS estimation is not required.

at time T, across all combinations of p and q, subject to the proviso that where $p > 1$ all of the regressors $\Delta y_{t-1}, ..., \Delta y_{t-p}$ are included in the estimated model, and similarly for $q > 0$ all of the regressors x_t, $x_{t-1}, ..., x_{t-q}$ are included in the estimated model. The maximum value allowed for p is set at $p_{\max} = 4$ and the maximum value for q is set at $q_{\max} = 2$. Based on the same set of sample observations, the BIC is also calculated for a version of (5) where the intercept and covariate regressors are excluded, again setting $p_{\max} = 4$, and with the same condition that for $p > 1$ all of the regressors $\Delta y_{t-1}, ..., \Delta y_{t-p}$ are included in the estimated model. In the case where $p = 0$ and no intercept or covariate regressors are included then no regression is performed and so the BIC is given by $\ln(\hat{\sigma}^2)$, with no penalty term, where $\hat{\sigma}^2$ is computed using the sample observations on Δy_t. If the minimum value of the BIC across all of these candidate models corresponds to a model that excludes the intercept and covariate regressors then the monitoring statistics underlying the CUSUM^{WMV} procedure coincide with those used in the CUSUM^{V*} procedure.[8] If the model with $p = 0$ and no intercept or covariate regressors is selected, the monitoring statistics underlying the CUSUM^{WMV} procedure coincide with those used in the CUSUM^{V} procedure of AHLTZ.

In implementing the CUSUM^{V*} procedure we also use the BIC applied to models estimated by OLS to select the value of p in (5) (with the intercept and covariate regressors excluded) based on the same set of sample observations from the training sample as are used in the BIC procedure for CUSUM^{WMV} outlined in the last paragraph, again setting the maximum permitted value of p to $p_{\max} = 4$, and with the same condition that for $p > 1$ all of the regressors $\Delta y_{t-1}, ..., \Delta y_{t-p}$ are included in the estimated model.[9] If the model with $p = 0$ is selected then the monitoring statistics underlying the CUSUM^{V*} procedure coincide with those underlying the CUSUM^{V} procedure of AHLTZ.

[8]We find that in all scenarios where $\beta \neq 0$ the intercept and covariate regressors are selected for inclusion in the CUSUM^{WMV} procedure in a vast majority of replications. Likewise, when $\beta = 0$, the intercept and covariate regressors are excluded by the BIC in a vast majority of replications. In the homoskedastic scenario, for instance, the intercept and covariate regressors are selected in 100% of replications when $\beta \neq 0$ and in only 1% of replications when $\beta = 0$. Additional simulations showed that this pattern is repeated in cases where the bubble begins before the start of the monitoring period.

[9]We considered allowing a larger maximum value of 12 for p in the CUSUM^{V*} procedure but found that this made no noticeable difference to the resulting FPR or TPR.

Following HB, all monitoring procedures use finite sample critical values; cf. footnote 5. We select a value of b_α such that the FPR is equal to 0.10 by time $t = 241$ when y_t is a pure unit root process driven by $NIID(0, 1)$ innovations and the covariate is an irrelevant white noise process; i.e., $\beta = \rho = \alpha_1 = 0$ and $\sigma_{12} = 0, \sigma^2_{1,t}, \sigma^2_{2,t} = 1$, for all t. For the standard CUSUM procedure this value is $b_\alpha = 0.1395$, while for CUSUM^V $b_\alpha = 0.1679$. The figures plot, in the line denoted FPR$_{\text{i.i.d.}}$, the FPR of the CUSUM^V procedure that would obtain in this baseline case under the null when the innovations are homoskedastic. CUSUM^{WMV} and CUSUM^{V*} use the same value of b_α as CUSUM^V.

5.2 Discussion of Results

The first set of results relate to the case where y_t admits a purely univariate DGP (i.e. x_t is not a relevant covariate); that is, where $\beta = \rho = \alpha_1 = 0$ and $\sigma_{12} = 0$, for all t. Here, and in any other cases where $\sigma_{12} = 0$, we omit results for the volatility shift in scenario (d) as this is identical to scenario (c) when $\sigma_{12} = 0$. These results are reported in Figure 1, with panel (a) pertaining to the baseline case where the innovations are homoskedastic.[10] For each time point e, $T + 1 \leq e \leq \lambda T$, the corresponding point on the curves in the figure represents the empirical rejection rate of the particular procedure run from time $t = T+1$ until time $t = e$.

In this baseline scenario where the covariate is irrelevant, as a point of comparison, we also report results for the (pseudo) real-time monitoring procedures proposed by PWY and PSY. The monitoring procedure of PWY is based on performing a full sample ADF test (allowing for a deterministic constant) at each point in the monitoring period using all data up to and including the current monitoring observation, and the monitoring procedure of PSY is based on performing the $BSADF$ test of PSY (again, allowing for a deterministic constant) at each point in the monitoring period using all data up to and including the current monitoring observation. The procedure of PWY compares the sequence of ADF statistics

[10]Here and in each of the remaining figures we also report the value of ϱ_2 for each simulation DGP in the case where $\sigma^2_{1,t} = \sigma^2_{2,t} = 1$, for all t. For scenarios where $\sigma^2_{1,t}$ and/or $\sigma^2_{2,t}$ are time varying, the value of ϱ^2 will also be time varying. Defining $q_t := \beta x_t + \varepsilon_{1,t}$, ϱ^2 is defined as the long run (zero frequency) squared correlation between q_t and $\varepsilon_{1,t}$, with precise details on the calculation of this quantity for this DGP provided in CSS (p.144). While Hansen (1995) and CSS show that the power of left-tailed unit root tests are inversely related to the value of ϱ^2, ATKK show that this is not necessarily the case when testing in the right-tail, and we observe that this is also the case for the CUSUM^{WMV} monitoring procedure.

with a fixed simulated critical value, with a rejection signalled if any ADF statistic in the sequence exceeds this critical value. Likewise, the procedure of PSY compares the sequence of $BSADF$ statistics with a fixed simulated critical value, with a rejection signalled if any $BSADF$ statistic in the sequence exceeds this critical value. We also include an implementation of our CUSUM^{WMV} procedure where we ignore the outcome of BIC model selection and force inclusion of the covariate (denoted CUSUM^{WMV}(Forced)). For both the PWY and PSY procedures the fixed critical value is chosen such that the FPR of the procedure is equal to 0.10 by time $t = 241$ when y_t is a pure unit root process driven by $NIID(0, 1)$ innovations, thereby mirroring the calibration process for the CUSUM procedures.[11]

We see from the results in Figure 1 that the BIC reduces the CUSUM^{WMV} and CUSUM^{V*} procedures to the CUSUM^{V} procedure in the vast majority of replications, and so the FPR and TPR of these three procedures are almost indistinguishable; indeed, forcing this irrelevant covariate to always be included is also seen to have little effect on either the FPR or TPR of CUSUM^{WMV}. As also demonstrated in AHLTZ, the standard CUSUM procedure exhibits severe FPR distortions when the innovations to y_t exhibit a smooth shift in volatility. In contrast, the CUSUM^{V}, CUSUM^{WMV} and CUSUM^{V*} procedures all control the FPR well in such cases. This shows that, like the CUSUM^{V} procedure of AHLTZ, our preferred CUSUM^{WMV} procedure has far superior FPR control to the standard CUSUM procedure in the presence of time varying volatility in a univariate setting, while only showing a modest TPR shortfall relative to the standard CUSUM procedure under the alternative when y_t is a pure unit root process driven by homoskedastic innovations. While the FPR of a monitoring procedure based on either the $SADF$ or $BSADF$ statistics is well controlled for homoskedastic innovations, these procedures, like the standard CUSUM procedure, suffer from very significant FPR distortions when the innovations exhibit time varying volatility. In the homoskedastic case, where a monitoring procedure based on $SADF$ or $BSADF$ has controlled FPR, we also observe that the TPR of the $BSADF$ and especially the $SADF$ procedures lies well below that of the CUSUM based monitoring procedures, other than where all of the procedures display very low TPRs. Due to the poor FPR

[11]For the $BSADF$ statistics, the minimum window size, r_0, was set to $0.01 + 1.8/\sqrt{t}$, as suggested in PSY, and in all ADF statistics the lag order was set to the true value of zero.

control and TPR properties they display in Figure 1 we will not consider monitoring procedures based on the $SADF$ or $BSADF$ test further in the remainder of our experiments.

We next examine the performance of the procedures for a DGP in which the covariate is relevant but the error term v_t in (23) admits no serial correlation. To that end, Figure 2 reports the FPR and TPR of the procedures for the CSS type DGP for v_t and x_t given by (23)-(24) with $\rho = \sigma_{12} = \alpha_1 = 0$ and $\beta = 0.8$ (corresponding results for $\beta = 0.5$ are given in the supplementary appendix and are qualitatively similar). As v_t is not serially correlated the BIC selects $p = 0$ in the great majority of replications so that the FPR and TPR curves for CUSUM^V and CUSUM^{V*} almost exactly coincide. Under the null, all but the standard CUSUM procedure exhibit decent FPR control in the presence of shifts in volatility. Under the alternative, CUSUM^{WMV} is seen to offer substantial power gains relative to both the CUSUM^V and CUSUM^{V*} procedures.

We next explore the properties of the monitoring procedures for DGPs that allow both v_t and x_t to be serially correlated. Figures 3-4 present the FPR and TPR of the procedures for the CSS type DGP for v_t and x_t given by (23)-(24) where, following CSS, we set $\alpha_1 = 0.2$ and $\sigma_{12} = 0.4$. We report results for $\rho = 0.8$ and $\beta \in \{-0.8, 0.8\}$, with additional figures in the supplementary appendix for the remaining combinations of β and ρ considered by CSS.

Across these figures, neither the standard CUSUM nor CUSUM^V procedures exhibit controlled FPR, with both of these procedures often displaying extreme FPR distortions relative to the baseline case where v_t is i.i.d. While the CUSUM^{WMV} and CUSUM^{V*} procedures do exhibit some slight FPR distortions relative to the case where v_t is i.i.d., these FPR distortions are very modest in comparison to those exhibited by CUSUM and CUSUM^V. Within each figure, examining the FPR performance of the procedures across panels (b)-(d) shows that the FPR performance of the CUSUM^{WMV} procedure is broadly similar in the cases where a shift in volatility occurs in $\varepsilon_{1,t}$ and where a shift in volatility occurs in both $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$, regardless of whether the correlation between these innovations remains fixed or not. This is not true for the remaining procedures which have an FPR profile that changes significantly across these three scenarios.

While we include the TPR of the standard CUSUM and CUSUM^V procedures in the figures we cannot compare these directly to the TPR of CUSUM^{WMV} and CUSUM^{V*} as the former two procedures display very significant FPR distortions under the null. Between the two FPR controlled procedures, the TPR of CUSUM^{WMV} is consistently much higher than that of CUSUM^{V*} across scenarios, showing that while CUSUM^{V*} is able to control the FPR under the null by dealing with the serial correlation induced by the presence of the covariate, it is unable to exploit the information from the covariate under the alternative, unlike CUSUM^{WMV} which displays impressive TPR properties across all scenarios considered.

5.3 Summary of Additional Results in the Supplementary Appendix

1. Results for the case where a volatility shift is present in only $\epsilon_{2,t}$ are reported in Section A.4.2. These highlight that the standard CUSUM procedure is unable to control the FPR as the volatility shift in the unmodelled covariate manifests in the values of Δy_t used to construct the CUSUM statistics. The CUSUM^V procedure is only able to control FPR when no serial correlation is present in v_t, and the FPR of the CUSUM^{V*} procedure, while better than that of CUSUM and CUSUM^V, is also quite poor. However, CUSUM^{WMV} displays good FPR control in all of the scenarios considered.

2. Results for the case where x_t is subject to measurement error are reported in Section A.4.3. These suggest that while the TPR of the CUSUM^{WMV} procedure is reduced in the presence of measurement error, increasingly so as the variance of the measurement error increases, it remains superior to the TPR exhibited by the other procedures.

3. Results where a bubble in the training sample is present in x_t are reported in Section A.4.4. The CUSUM^V procedure is unaffected, provided the bubble terminates at least H (the maximum bandwidth considered for the kernel variance estimator) periods before the start of monitoring. For CUSUM^{WMV} and CUSUM^{V*}, the residuals used in constructing the CUSUM statistics use all of the available sample data. Where the covariate is irrelevant, the FPR and TPR of the CUSUM^{WMV} and CUSUM^{V*} procedures are little altered, while a training sample bubble in a relevant covariate causes a slight inflation of the FPR of the CUSUM^{WMV} and CUSUM^{V*} procedures. This could potentially be obviated by truncation of the training sample.

4. Results for the case where an irrelevant $I(1)$ covariate, x_t, is mistakenly used in the CUSUM^{WMV} procedure are reported in Section A.4.5. These show that including x_t

causes CUSUM^{WMV} to exhibit a slightly inflated FPR and modestly lower TPR than the correctly specified univariate tests. Reassuringly, the loss in TPR is modest and is predicated on a practitioner failing to difference x_t and then forcing the inclusion of x_t in the CUSUM^{WMV} procedure, as the BIC model selection we recommend for this procedure determines x_t to be irrelevant in the vast majority of cases.

5. Results where an irrelevant covariate admits a bubble during the monitoring period are reported in Section A.4.6. We consider the case where the covariate is initially either $I(0)$ or $I(1)$ before switching to an explosive regime at the start of monitoring. Forcibly including the covariate in the CUSUM^{WMV} procedure leads to a slight inflation of the FPR under the null and a modest decrease in the TPR under the alternative, with this effect more pronounced where the covariate is initially $I(0)$. Analogous results for a relevant covariate containing a bubble at the start of monitoring are reported in Section A.4.7. These show a very slight increase in the FPR and no perceptible change in the TPR, relative to the case where no bubble is present in the covariate.

6. Results where a mean shift is present during the monitoring period in a utilised covariate are reported in Section A.4.8. These suggest this is problematic only where the covariate is relevant ($\beta \neq 0$). A mean shift in a relevant covariate which is entered in first differences, as will generally be the case with macro and financial variables (see Remark 2.8), also appears relatively benign. A mean shift in a series entered in levels is more problematic causing a large increase in the FPR of CUSUM^{WMV}. However, the approach suggested at the end of Section 4 to simultaneously monitor the covariate for structural change appears useful, in that under the no bubble null it rejects in the presence of the mean shift with significantly higher frequency than does CUSUM^{WMV}.

7. Results where a relevant but unobserved covariate, x_t, is the input to an observed local-to-unity process, z_t, with local-to-unity parameter, $c > 0$, but Δz_t (rather than $z_t - (1 - \frac{c}{T})z_{t-1}$) is incorrectly used as the covariate are reported in Section A.4.9. Relative to the correctly specified case where $c = 0$, the FPR of CUSUM^{WMV} tends to be slightly increased and the TPR slightly decreased, with these effects increasing in c. These findings echo the results reported in Hansen (1995, pp.1159-1160) for covariate

augmented unit root tests in this scenario. We note that Δz_t does not violate the regularity conditions given in Assumption 2, regardless of the value of c.

6 Conclusions

We have developed a generalisation of the univariate CUSUM-based real-time bubble monitoring procedure of HB which incorporates additional information from relevant covariates and is also robust to unconditional heteroskedasticity and serial correlation in the disturbances. We have shown that the CUSUM statistics used in this procedure follow the same limiting null distribution as those in HB, such that a monitoring procedure can be validly based on the same large sample boundary function. Monte Carlo results were presented showing that, in contrast to univariate procedures, our proposed procedure has a controlled false positive rate where a relevant dynamic covariate enters the DGP. Moreover, where an explosive episode occurs in the monitoring period, incorporating the covariate can yield significant gains in finite sample detection efficacy, relative to univariate procedures.

References

Astill, S., Harvey, D.I., Leybourne, S.J. and Taylor, A.M.R. (2017). Tests for an end-of-sample bubble in financial time series. Econometric Reviews 36, 651-666.

Astill, S., Harvey, D.I., Leybourne, S.J., Sollis, R. and Taylor, A.M.R. (2018). Real-time monitoring for explosive financial bubbles. Journal of Time Series Analysis 39, 863–891.

Astill, S., Harvey, D.I., Leybourne, S.J., Taylor, A.M.R. and Zu, Y. (2023a). CUSUM-based monitoring for explosive episodes in financial data in the presence of time-varying volatility. Journal of Financial Econometrics 21, 187-227.

Astill, S., Taylor, A.M.R., Kellard, N. and Korkos, I. (2023b). Using covariates to improve the efficacy of univariate bubble detection methods. Journal of Empirical Finance 70, 342-366.

Caballero, R.J., Farhi, E. and Gourinchas, P.-O. (2008). Financial crash, commodity prices and global imbalances. Brookings Papers on Economic Activity, 1-55.

Calvo-Gonzalez, O., Shankar, R. and Trezzi, R. (2010). Are commodity prices more volatile now? World Bank Policy Research Working Paper 5460.

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

Chang, Y., Sickles, R.C. and Song, W. (2017). Bootstrapping unit root tests with covariates. Econometric Reviews 36, 136-155.

Chu, C.-S., Stinchcombe, J.M. and White, H. (1996). Monitoring structural change. Econometrica 64, 1045-1065.

Diba, B.T. and Grossman, H.I. (1988). Explosive rational bubbles in stock prices? American Economic Review 78, 520-530.

Evans, G.W. (1991). Pitfalls in testing for explosive bubbles in asset prices. American Economic Review 81, 922-930.

Hall, P. and Schucany, W.R. (1989). A local cross-validation algorithm. Statistics and Probability Letters 8, 109-117.

Hansen, B.E. (1995). Rethinking the univariate approach to unit root testing: Using covariates to increase power. Econometric Theory 11, 1148-1171.

Harvey, D.I., Leybourne, S.J., Sollis, R. and Taylor, A.M.R. (2016). Tests for explosive financial bubbles in the presence of non-stationary volatility. Journal of Empirical Finance 38, 548-574.

Harvey, D.I., Leybourne, S.J. and Zu, Y. (2019). Testing explosive bubbles with time-varying volatility. Econometric Reviews 38, 1131-1151.

Harvey, D.I., Leybourne, S.J. and Zu, Y. (2020). Sign-based unit root tests for explosive financial bubbles in the presence of nonstationary volatility. Econometric Theory 36, 122-169.

Harvey, D.I., Leybourne, S.J. and Zu, Y. (2023). Estimation of the variance function in structural break autoregressive models with non-stationary and explosive segments. Journal of Time Series Analysis, 44, 181-205.

Homm, U. and Breitung, J. (2012). Testing for speculative bubbles in stock markets: a comparison of alternative methods. Journal of Financial Econometrics 10, 198-231.

Kilian, L. and Murphy, D.P. (2014). The role of inventories and speculative trading in the global market for crude oil. Journal of Applied Econometrics 29, 454–478.

Kim, K.H. and Kim, T. (2016). Capital asset pricing model: A time-varying approach. Journal of Empirical Finance 37, 268–281.

Kostakis, A., Magdalinos, T. and Stamatogiannis, M.P. (2015). Robust econometric inference for stock return predictability. Review of Financial Studies 28, 1506–1553.

Kramer, W., Ploberger, W. and Alt, R. (1988). Testing for structural change in dynamic models. Econometrica 56, 1355-1369.

Kurozumi, E., Skrobotov, A. and Tsarev, A. (2023). Time-transformed test for bubbles under non-stationary volatility. Journal of Financial Econometrics, 21, 1282-1307.

McMillan, D.G. and Wohar, M.E. (2011). Structural breaks in volatility: the case of UK sector returns. Applied Financial Economics 21, 578-595.

Phillips, P.C.B. and Shi, S.-P. (2020). Real time monitoring of asset markets: Bubbles and crises. In Handbook of Statistics (Vol. 42, pp. 61-80). Elsevier.

Phillips, P.C.B. and Shi, S.-P. (2018). Financial bubble implosion and reverse regression. Econometric Theory 34, 705-753.

Phillips, P.C.B., Shi, S.-P, and Yu, J. (2015). Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P500. International Economic Review 56, 1043-1078.

Phillips, P.C.B. and Yu, J. (2011). Dating the timeline of financial bubbles during the subprime crisis. Quantitative Economics 2, 455–491.

Phillips, P.C.B., Wu, Y. and Yu, J. (2011). Explosive behavior in the 1990s Nasdaq: When did the exuberance escalate asset values? International Economic Review 52, 201-226.

Shi, S. and Phillips, P.C.B. (2023). Diagnosing housing fever with a econometric thermometer. Journal of Economic Surveys 37, 159–186.

Singleton, K.J. (2014). Investor flows and the 2008 boom/bust in oil prices. Management Science 60, 300–318.

Tsvetanov, D., Coakley, J. and Kellard, N. (2016). Bubbling over! The behaviour of oil futures along the yield curve. Journal of Empirical Finance 38, 516-533.

Vivian, A. and Wohar, M.E. (2012). Commodity volatility breaks. Journal of International Financial Markets, Institutions and Money 22, 395-422.

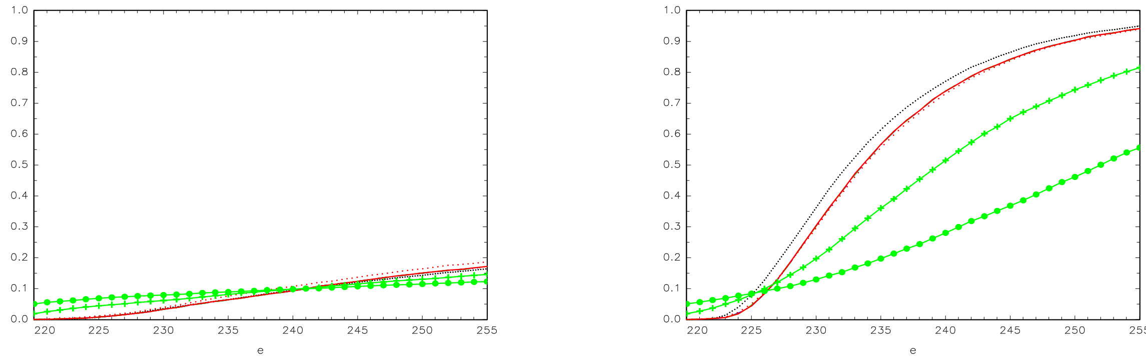
Welch, I. and Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. The Review of Financial Studies, 21(4), 1455-1508.

White, M. (2015). Cyclical and structural change in the UK housing market. Journal of European Real Estate Research 8, 85–103.

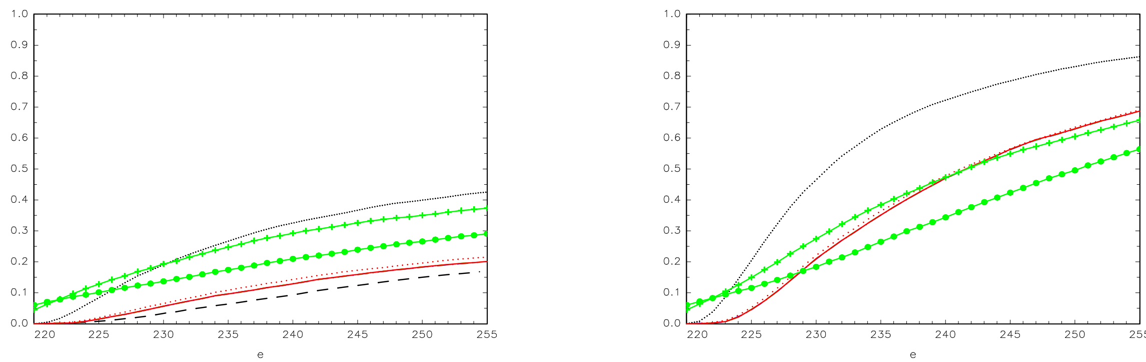
Xu, K.L. and Phillips, P.C.B. (2008). Adaptive estimation of autoregressive models with time-varying variances. Journal of Econometrics, 142(1), 265-280.

Figure 1: $\beta = \rho = \sigma_{12} = \alpha_1 = 0$ - Left Panel=FPR, Right Panel =TPR. ($\varrho^2 = 1.000$)

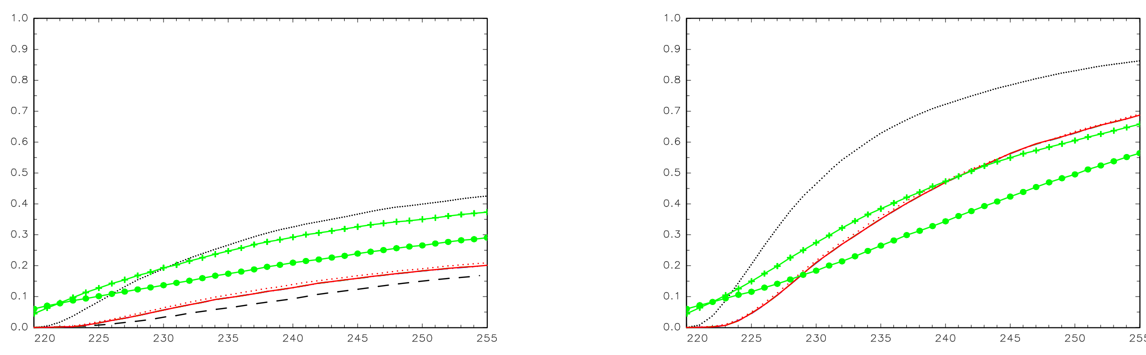
(a) Homoskedastic



(b) $\sigma_{1,t}$, $\sigma_{2,t}$ Shift.



(c) $\sigma_{1,t}$ Shift.

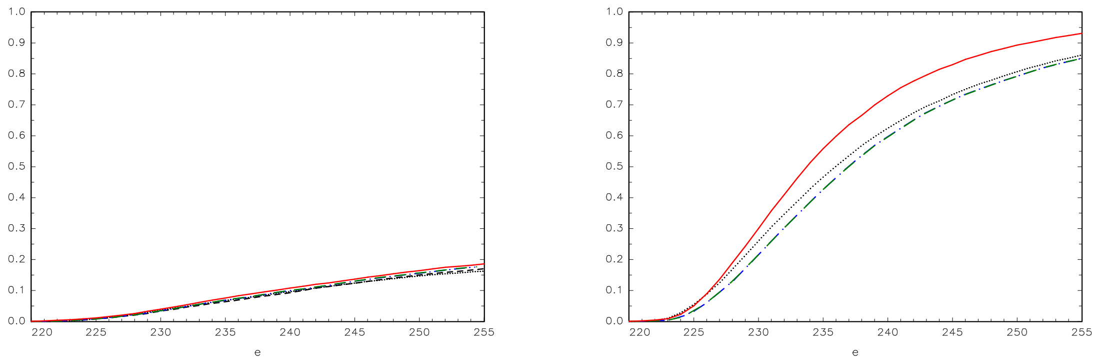


FPR$_{\text{i.i.d.}}$: - - -, CUSUM: ·····, CUSUM^V: - · -, CUSUM^{V*}: — —, CUSUM^{WMV}: ——
CUSUM^{WMV}(Forced): · · ·, $SADF$: —●—, $BSADF$: —+—

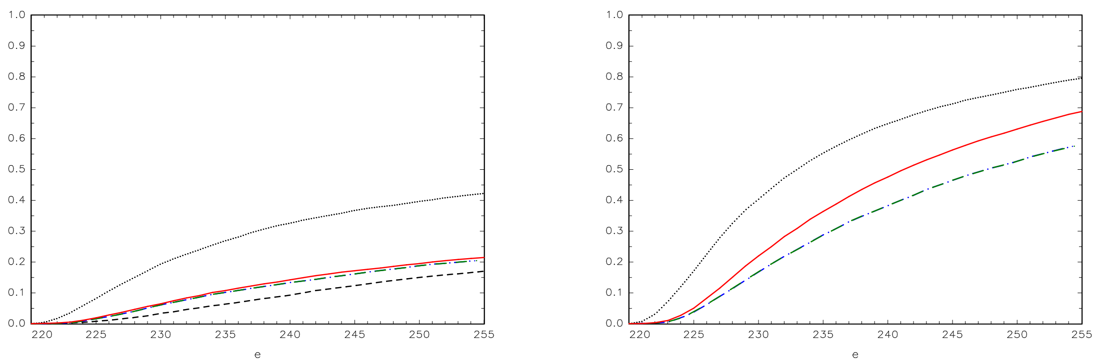
Notes: (a) Each graph in this figure, and in all subsequent figures relating to our Monte Carlo experiments, denotes the proportion of the simulation replications in which each procedure detects a bubble when run up to and including time e, for $e = 220, ..., 255$. Under the null (alternative) this therefore depicts the empirical FPR (TPR) of the procedures; (b) The red dotted line corresponds to the case where the covariate is always included in the null regression model (5) used in connection with the CUSUM^{WMV} procedure.

Figure 2: $\beta = 0.8$, $\rho = \sigma_{12} = \alpha_1 = 0$ - Left Panel=FPR, Right Panel =TPR. ($\varrho^2 = 0.610$)

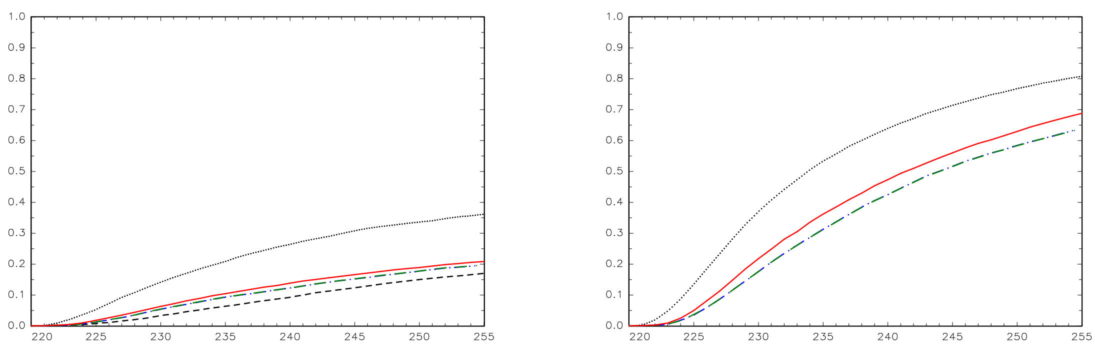
(a) Homoskedastic



(b) $\sigma_{1,t}$, $\sigma_{2,t}$ Shift.



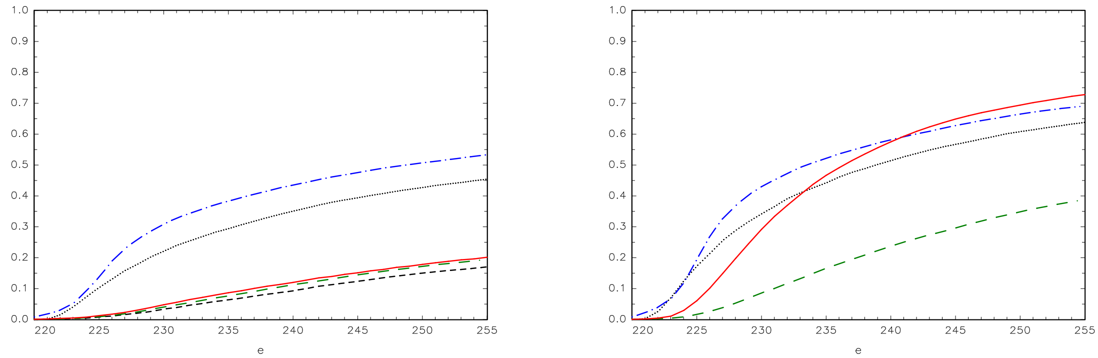
(c) $\sigma_{1,t}$ Shift.



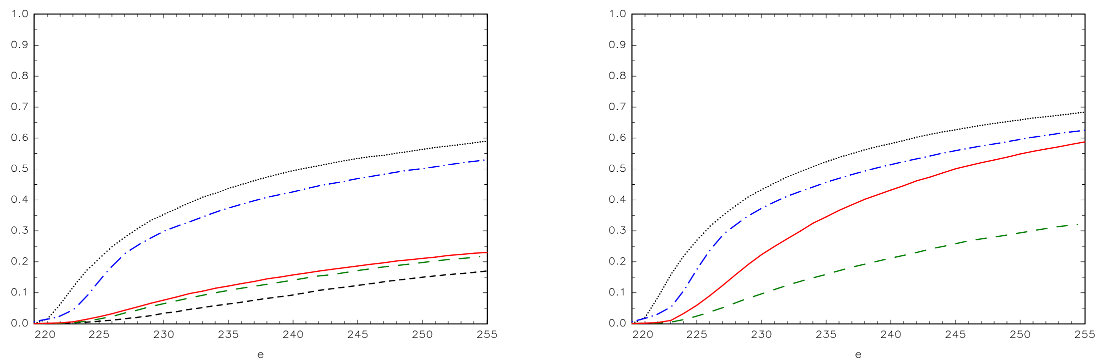
$\text{FPR}_{\text{i.i.d.}}$: - - -, CUSUM: ·····, CUSUM^V: - · · -, CUSUM^{V*}: — —, CUSUM^{WMV}: ——

Figure 3: $\beta = 0.8$, $\rho = 0.8$, $\sigma_{12} = 0.4$, $\alpha_1 = 0.2$ - Left Panel=FPR, Right Panel =TPR. ($\varrho^2 = 0.335$)

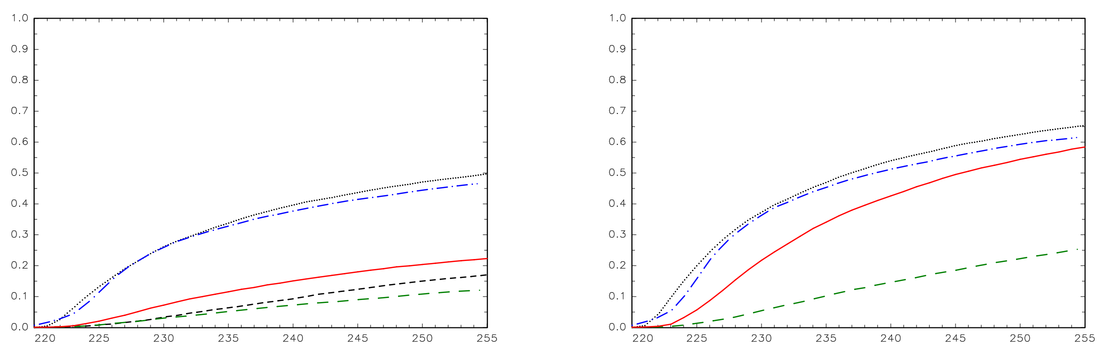
(a) Homoskedastic



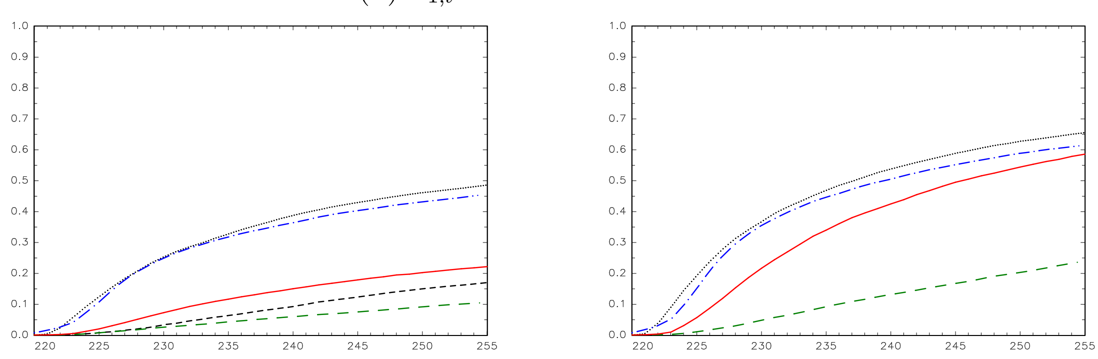
(b) $\sigma_{1,t}$, $\sigma_{2,t}$ Shift. Fixed Correlation



(c) $\sigma_{1,t}$ Shift. Fixed Correlation



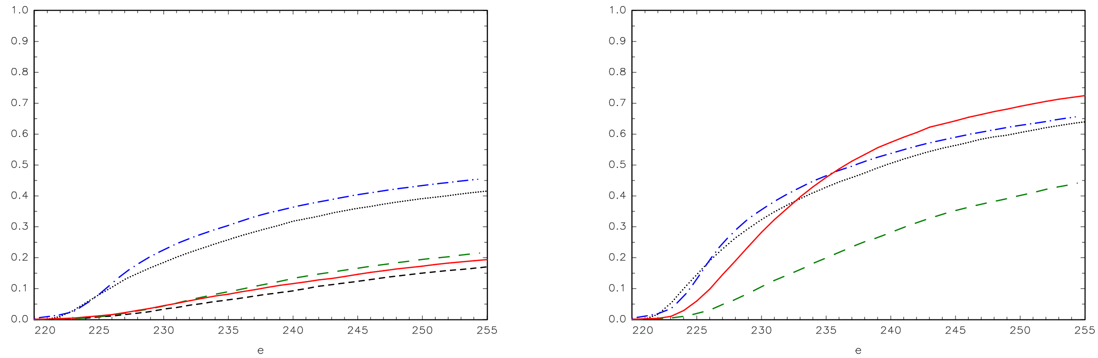
(d) $\sigma_{1,t}$ Shift. Correlation Varies



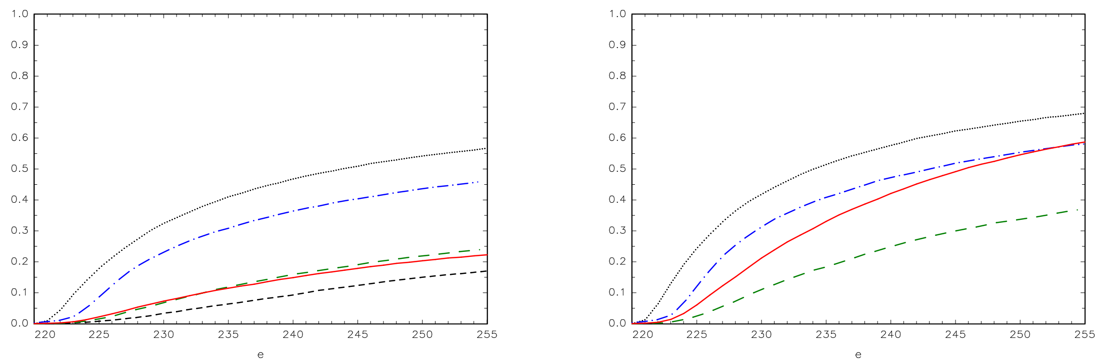
FPR$_{\text{i.i.d.}}$: - - -, CUSUM: ·····, CUSUM^V: - · -, CUSUM^{V*}: — —, CUSUM^{WMV}: ——

Figure 4: $\beta = -0.8$, $\rho = 0.8$, $\sigma_{12} = 0.4$, $\alpha_1 = 0.2$ - Left Panel=FPR, Right Panel =TPR. ($\varrho^2 = 0.026$)

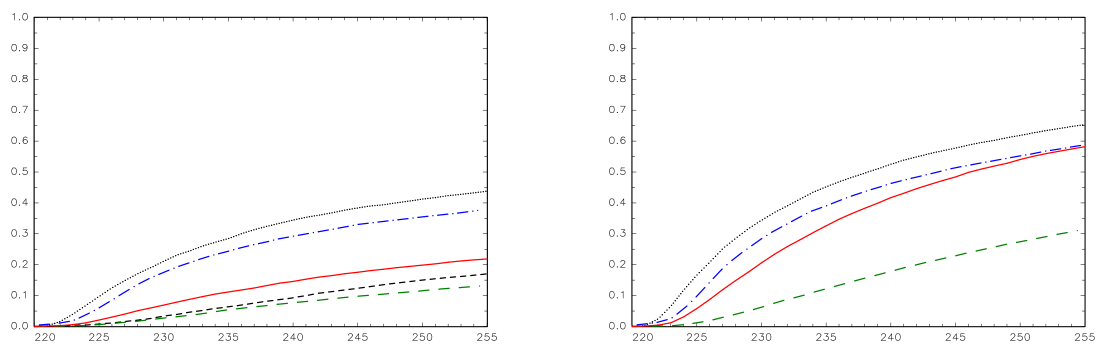
(a) Homoskedastic



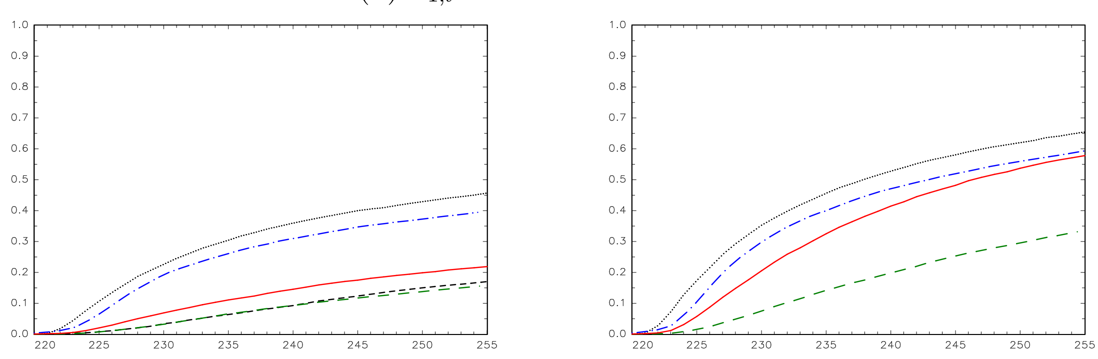
(b) $\sigma_{1,t}$, $\sigma_{2,t}$ Shift. Fixed Correlation



(c) $\sigma_{1,t}$ Shift. Fixed Correlation



(d) $\sigma_{1,t}$ Shift. Correlation Varies



$\text{FPR}_{\text{i.i.d.}}$: - - -, CUSUM: ·····, CUSUM^V: - · -, CUSUM^{V*}: — —, CUSUM^{WMV}: ——

Supplementary Appendix to "Covariate Augmented CUSUM Bubble Monitoring Procedures"

Sam Astill[a], A.M. Robert Taylor[a,b] and Yang Zu[c]

[a] Essex Business School, University of Essex

[b] Aarhus Center for Econometrics, Aarhus University

[c] Department of Economics and APAEM, University of Macau

Abstract

The contents of this supplementary appendix are as follows. Section A.1 outlines the analogue of the covariate augmented CUSUM procedure proposed in Section 4 for the case where it is *known* that all of the covariates have mean zero and the constant term is accordingly omitted from the null regression. Section A.2 provides proofs of the large sample results stated in Section 4 and Section A.1. Section A.3 reports the results from a pseudo real-time empirical exercise comparing our covariate augmented monitoring procedure with one which does not allow for covariate information, using an updated version of the dataset of Welch and Goyal (2008). Section A.4 contains the results of the additional Monte Carlo simulation experiments referred to in Section 5.

A.1 A Covariate Augmented CUSUM Monitoring Procedure - No Constant Term

In the case where the covariates, x_t, are *known* to have mean zero, such that $c_x = 0$, so that the constant term can be omitted from the null regression in (5), we have the simplified null model,

$$\Delta y_t = \sum_{k=1}^{p} \alpha_k \Delta y_{t-k} + \sum_{k=0}^{q} \beta_k' w_{t-k} + \varepsilon_t, \tag{A.1}$$

where w_t now coincides with x_t and satisfies the conditions laid out in Assumption 1, and where the first summation term is again understood to be present only when $p > 0$.

Defining $z_t := (\Delta y_{t-1}, \ldots, \Delta y_{t-p}, w_t', w_{t-1}', ..., w_{t-q}')'$ and $\phi := (\alpha_1, \ldots, \alpha_p, \beta_0', \beta_1', \ldots, \beta_q')'$, the null model (A.1) can be written more compactly as

$$\Delta y_t = \phi' z_t + \sigma_t \eta_t, \quad t = 1, \ldots, T, \ldots, \lfloor \lambda T \rfloor \tag{A.2}$$

Following KPA, our proposed CUSUM monitoring statistic is based on recursive estimation of (A.2). However, in contrast to the $SWMV_T^t$ statistics outlined in Section 4, these can be estimated by OLS, rather than WLS, and still attain the large sample results in (8) and (9); for further discussion on this point, see Remark A.1.7 below.

To that end, defining the recursive LS estimator for ϕ from (A.2) in the monitoring period as

$$\hat{\phi}_t := \left(\sum_{j=\max(p+2,q+1)}^{t} z_j z_j' \right)^{-1} \left(\sum_{j=\max(p+2,q+1)}^{t} z_j \Delta y_j \right), \quad t = T+1, \ldots, \lfloor \lambda T \rfloor \tag{A.3}$$

the (null) recursive residuals in the monitoring period can then be defined as

$$e_t := \Delta y_t - \hat{\phi}_{t-1}' z_t, \quad t = T+1, \ldots, \lfloor \lambda T \rfloor. \tag{A.4}$$

A key difference in the fitted null regression model compared to that discussed in Section 4 is that there is no constant term included in the regressors in z_t.

Consider first the infeasible case where the volatility function, σ_t, is known. Here, replacing Δy_j in the CUSUM statistic of HB in (7) by the recursive null residual, e_j, and scaling by the known volatility, σ_j, we obtain the following (infeasible) covariate augmented CUSUM statistic,

$$SW_T^t := \sum_{j=T+1}^{t} \frac{e_j}{\sigma_j}, \qquad t > T. \tag{A.5}$$

In Theorem A.1 we next establish the limiting null distribution of the sequence of infeasible covariate augmented CUSUM statistics, SW_T^t, $t > T$. In order to do so we need to replace Assumption 2(e) on the regressors in the WLS regression, (5), with a corresponding set of conditions on the OLS regression with the constant term omitted, (A.1). Analogously to Assumption 2(e), this excludes the possibility of asymptotic collinearity between the regressors in (A.1) and also needs to hold for us to be able to make use of the weak convergence result in Lemma A.10.

Assumption A.1. *For all* $0 \leq \kappa \leq \lambda$, *it holds that* $\text{plim}_{T\to\infty}(1/\lfloor T\kappa \rfloor)\sum_{s=1}^{\lfloor T\kappa \rfloor} z_s z_s' = \lim_{T\to\infty}(1/\lfloor T\kappa \rfloor)\ E(\sum_{s=1}^{\lfloor T\kappa \rfloor} z_s z_s') =: \Xi(\kappa)$, *and that* $\text{plim}_{T\to\infty}(1/\lfloor T\kappa \rfloor)\sum_{t=1}^{\lfloor T\kappa \rfloor} z_t z_t' \sigma_t^2 h_t = \lim_{T\to\infty} E(1/\lfloor T\kappa \rfloor)\sum_{t=1}^{\lfloor T\kappa \rfloor} z_t z_t' \sigma_t^2 h_t =: \Omega(\kappa)$, *with* $\Xi(\kappa)$ *and* $\Omega(\kappa)$ *both positive definite matrices with all elements finite and continuous in* κ.

Theorem A.1. *Let the data be generated according to (1)-(4) under the null hypothesis* $H_0 : \delta = 0$. *If Assumptions 1-2, excluding Assumption 2(e), and Assumption A.1 hold, then, as* $T \to \infty$, *it follows that*

$$T^{-1/2} SW_T^{\lfloor Tr \rfloor} \Rightarrow W(r) - W(1), \quad 1 < r \leq \lambda, \tag{A.6}$$

where $W(\cdot)$ *denotes a standard Brownian motion on* $[0, \lambda]$.

As in the leading case considered in Section 4 where a constant is included in the null regression, in order to develop a feasible version of this statistic we need to replace σ_j by a nonparametric estimate thereof. The nonparametric estimator for the variance function $\sigma^2(\cdot)$ we use will be based on the double array of OLS residuals

$$f_{i,j} := \Delta y_i - \hat{\phi}_j' z_i, \quad i = \max(p+2, q+1), \ldots, j,\ j = T+1, \ldots, \lfloor \lambda T \rfloor. \tag{A.7}$$

Using the OLS residuals in (A.7), we can then define the sequence of nonparametric variance estimators across $j = T+1, \ldots, \lfloor \lambda T \rfloor$, when standing at time t, as

$$\hat{\sigma}_{j,N,t}^2 := \sum_{s=0}^{N} k_s f_{j-s,t}^2, \quad k_s := \frac{K\left(\frac{s}{N}\right)}{\sum_{s=0}^{N} K\left(\frac{s}{N}\right)}, \tag{A.8}$$

where k_s, $s = 0, ..., N$, is a sequence of weights, for the kernel function $K(\cdot)$ and a window size N. An important difference, compared to the methods outlined in Section 4, is that only the $\hat{\sigma}^2_{j,N,j}$ are needed for constructing the monitoring statistic. This is because recursive LS residuals, rather than recursive WLS residuals, are used in (A.5). We will therefore use the simplified notation $\hat{\sigma}^2_{j,N} := \hat{\sigma}^2_{j,N,j}$ for $j = T+1, \ldots, \lfloor \lambda T \rfloor$, in what follows. As in the main text, due to the unavailability of future data, this nonparametric variance estimator also uses a left-sided, truncated kernel. Only the N most recent observations are used in the calculation of the estimator and the weights are not dependent on t. Notice also that for practical implementation we require that $N \leq T - \max(p+1, q)$, such that $\hat{\sigma}^2_{T+1,N}, \ldots, \hat{\sigma}^2_{\lfloor \lambda T \rfloor, N}$ can be computed.

Based on (A.8), a feasible version of the covariate augmented CUSUM statistic in (A.5) can then be defined as

$$SWV_T^t := \sum_{j=T+1}^{t} \frac{e_j}{\hat{\sigma}_{j,N}}, \qquad t > T. \tag{A.9}$$

In what follows, we will denote a monitoring procedure based on the sequence of SWV_T^t, $t = T+1, \ldots \lfloor \lambda T \rfloor$, statistics as CUSUM^{WV}.

Remark A.1.1. Notice that in the definition of SWV_T^t, the recursive residuals $\{e_j\}_{j=T+1}^t$ are used in the numerator of the statistic, while the double array of OLS residuals $\{f_{i,j}\}$ for $\max(p+2, q+1) \leqslant i \leqslant j$ are used for estimating $\hat{\sigma}_{j,N}$ in the denominator. ♢

In Theorem A.2, we establish the joint limiting null distribution of the sequence of feasible covariate augmented SWV_T^t statistics from the monitoring period. This is shown to coincide with the result given for the known volatility case in (A.6).

Theorem A.2. *Let the data be generated according to (1)-(4) under the null hypothesis $H_0 : \delta = 0$. If Assumptions 1-3, excluding Assumption 2(e), and Assumption A.1 hold, then, as $T \to \infty$, it follows that*

$$T^{-1/2} SWV_T^{\lfloor Tr \rfloor} \Rightarrow W(r) - W(1), \quad 1 < r \leq \lambda. \tag{A.10}$$

Remark A.1.2. Notice from Theorem A.2, that the joint limiting null distribution of the SWV_T^t, $t > T$, statistics does not depend on any nuisance parameters arising from time-varying behaviour in the unconditional covariance matrix of the covariates; cf. Remark 2.5. ♢

Appealing to Theorem 3.4 of Chu *et al.* (1996), the result in Theorem A.2 implies the following corollary,

Corollary A.1. *Under the conditions of Theorem A.2,*

$$\lim_{T\to\infty} \Pr\left(|SWV_T^t| > c_t\sqrt{t} \text{ for some } t \in \{T+1, ..., \lfloor \lambda T \rfloor\}\right) \leq \exp(-b_\alpha/2). \tag{A.11}$$

In Theorem A.3 we establish a similar consistency result as in Theorem 2 that the covariate augmented CUSUM^{WV} monitoring procedure is also consistent when a bubble is present in the monitoring period, rejecting the false null of no explosivity with probability one in the limit.

Theorem A.3. *Let the data be generated according to (1)-(4) under the alternative hypothesis $H_1 : \delta > 0$, and let Assumptions 1-3 and Assumption A.1 hold, excluding Assumption 2(e). It holds that,*

$$\lim_{T\to\infty} \Pr\left(|SWV_T^t| > c_t\sqrt{t}, \text{for some } t \in \{\lfloor \tau T \rfloor + 1, \ldots, \lfloor \lambda T \rfloor\}\right) = 1. \tag{A.12}$$

Remark A.1.3. The results in Theorem A.2, Corollary A.1 and Theorem A.3 imply that, where the covariates have mean zero and the constant term is correspondingly omitted from the null regression, the large sample properties of the CUSUM^{WV} procedure coincide with those given for the CUSUM^{WMV} procedure in Section 4. ◇

Remark A.1.4. As in Remark 4.4, it is also instructive to examine the behaviour of the monitoring procedure under locally explosive alternatives of the form $H_{c,\tau} : \delta_T = c/T$, for $t > \lfloor \tau T \rfloor$, where c is a positive constant. When the volatility process is known, the asymptotic behaviour of the detector SWV_T^t can be derived along the same line of argument as the proof of Theorem A.1. In particular, when $\alpha(L) = 1$,

$$\frac{1}{\sqrt{T}} \sum_{j=T+1}^{\lfloor Tr \rfloor} \frac{e_j}{\sigma_j} = \frac{1}{\sqrt{T}} \sum_{j=T+1}^{\lfloor Tr \rfloor} \eta_j - \frac{1}{\sqrt{T}} \sum_{j=T+1}^{\lfloor Tr \rfloor} \frac{(\hat{\phi}_{j-1} - \phi)' z_j}{\sigma_j} + \frac{c}{T^{3/2}} \sum_{j=\lfloor \tau T \rfloor+1}^{\lfloor Tr \rfloor} \frac{u_j}{\sigma_j}$$

As in the proof of Theorem A.1, the first term converges weakly to $W(r) - W(1)$. The second term can be shown to be of $o_p(1)$. By the FCLT and CMT, the third term satisfies $\frac{c}{T^{3/2}} \sum_{j=\lfloor \tau T \rfloor+1}^{\lfloor Tr \rfloor} \frac{u_j}{\sigma_j} \Rightarrow c \int_\tau^r e^{c(r-s)} W(s) \mathrm{d}s$. For general $\alpha(L)$, it can be shown in the same way that the asymptotic distribution under $H_{c,\tau}$ is given by $W(r) - W(1) + \alpha(1)c\int_{\tau}^{r} e^{c(r-s)}W(s)\mathrm{d}s$, from which the asymptotic probability of the CUSUM^{WMV} procedure rejecting the null of no explosivity when a locally explosive episode is present can easily be computed by numerical simulation. Where the volatility is estimated, we conjecture the same limit will hold under $H_{c,\tau}$ in view of the results given in Harvey *et al.* (2019) for the behaviour of the non-parametric variance estimator considered in this paper under locally explosive DGPs. ♢

Remark A.1.5. To implement SWV_T^t we again recommend the use of the kernel and bandwidth selection criteria and choice of the tuning parameter H, outlined in Remark 4.1. ♢

Remark A.1.6. The tests developed in this section, which omit a constant term from the null regression, do not require the use of recursive WLS residuals. However, the numerator of the CUSUM statistics developed in this Section could alternatively be constructed from the analogous recursive WLS residuals, obtained from (12) but with the constant term omitted, without altering the large sample results given in Theorem A.1, Theorem A.2, Corollary A.1 and Theorem A.3, provided a condition analogous to Assumption 2(e), omitting the constant term from g_t, held. ♢

Remark A.1.7. As Remark A.1.6 above notes, the procedures considered in this section, which omit a constant term from the null regression (in the case where it is known that the covariates have mean zero), can be based on CUSUM statistics formed either from recursive LS residuals or recursive WLS residuals, without altering their large sample properties. This is not the case, however, for the statistics considered in Section 4, where a constant term is included in the null regression. In this latter case, it can be shown that if recursive LS residuals were used then, except in the special case where $\sigma(s) = \sigma$, for all $s \in [0, \lambda]$, the resulting sequence of CUSUM statistics would not retain the nuisance parameter free null limiting distribution which appears in (8). The reason for this can be seen in the proof of Theorem 1 (see, in particular, the discussion after (A.25)) where it is shown that the limiting distribution of the CUSUM process formed from recursive LS residuals weakly converges to the difference between a standard Brownian motion process and an integral functional of a vector Gaussian process, both of which are of $O_p(1)$. It is only where recursive WLS residuals are used that the variance of the integral functional cancels exactly with its covariance with the Brownian motion term, implying that the difference between these two terms is a standard Brownian motion; see the proof of Theorem 1. As such, the claim made in Remark 10 on pages 195-196 of AHLTZ, that the limiting null distribution of CUSUM statistics which correct for the possibility of a non-zero mean in Δy_t based on recursive LS residuals will still obtain the limiting null distributional result given in (8), is incorrect except in the special case where $\sigma(s) = \sigma$, for all $s \in [0, \lambda]$. As with the procedures detailed in Section 4, recursive WLS residuals are required for this large sample result to hold. $\diamondsuit$

Remark A.1.8. If it were known that the unconditional volatility function $\sigma_t = \sigma < \infty$, for all $t = 1, ..., T, ..., \lfloor \lambda T \rfloor$, then one could consider a covariate augmented monitoring procedure based on a simplified version of the SWV_T^t statistic, given by $\tilde{\sigma}_t^{-1} \sum_{j=T+1}^t e_j$ where $\tilde{\sigma}_t^2 := (t - \max(p+1, q))^{-1} \sum_{j=\max(p+2,q+1)}^t f_{j,t}^2$. Under this restriction, it can be shown that the limit distribution of this statistic is identical to that given in Theorem A.2. In the constant unconditional volatility case one could also consider a simplified version of the $SWMV_T^t$ statistic given by $\tilde{\sigma}_t^{-1} \sum_{j=T+1}^t e_j^*$ where $\tilde{\sigma}_t^2 := (t - \max(p+1, q))^{-1} \sum_{j=\max(p+2,q+1)}^t {f_{j,t}^*}^2$, and where $e_j^* := \Delta y_j - \hat{\varphi}_{j-1}' g_j$ are recursive residuals, with $\hat{\varphi}_{j-1}$ the OLS estimator at time $j-1$ from (15). The constancy of the volatility function means that the WLS transformation is no longer needed, so that the numerator of this statistic can be based on the recursive residuals e_j^* rather than $\hat{e}_j^W$. In this case the limiting null distribution of this statistic is identical to that given in Theorem 1. Neither of these results, however, require the covariates to be homoskedastic. $\diamondsuit$

Remark A.1.9. The limiting results given in this section are based on the assumption that the covariates, x_t, are all mean zero. If that were not the case, then the limiting null distribution of the sequence of SWV_T^t, $t = T+1, ...\lfloor \lambda T \rfloor$, statistics would depend on nuisance parameters arising from c_x, the (non-zero) mean of x_t. As a consequence, the resulting CUSUM^{WV} procedure would not have a controlled FPR under the null. The safe strategy is therefore to use the CUSUM^{WMV} procedure, rather than the CUSUM^{WV} procedure, because in practice it would be unknown whether the covariates are all mean zero or not. To investigate what, if any, loss in finite sample performance is seen when using

CUSUM^{WMV} rather than CUSUM^{WV}, we repeated the simulation experiments reported in Section 5, where the covariates are all mean zero. These results show that the FPR control of the CUSUM^{WV} procedure is marginally better (that is, slightly closer to the i.i.d.-based FPR) than that of the CUSUM^{WMV} procedure, while the TPR of the two procedures is broadly similar. The safe strategy of using CUSUM^{WMV} therefore appears to be relatively costless. $\diamondsuit$

A.2 Proofs of Theorems

Throughout this section, unless otherwise stated, we use $\max_t$ or $\max_j$ as shorthand notation for $\max_{T+1\leqslant t\leqslant \lfloor \lambda T\rfloor}$or $\max_{T+1\leqslant j\leqslant \lfloor \lambda T\rfloor}$, respectively. We also denote by $(\sigma^2)'(\cdot)$ the derivative of $\sigma^2(\cdot)$. Denote the space of càdlàg functions defined over the interval $[0,\lambda]$ by $D[0,\lambda]$, and the space of continuous functions over the same interval by $C[0,\lambda]$. Notice that, due to the monitoring nature of our problem, we do not normalise the end point of the interval to 1, but instead to the fixed value, $\lambda > 1$. For positive constants a and b, $\min(a,b)$ takes the smaller constant. For two sequences $a_T, b_T \to \infty$,'$a_T \wedge b_T$' denotes taking the sequence with slower rate of divergence.

A.2.1 Preparatory Lemmata

In this section we begin by stating and proving some preparatory lemmata that will subsequently be required for the proofs of the large sample results stated in Section 4 and Section A.1.

Lemma A.1. *Let the conditions of Theorem 1 hold. Then, under H_0,*

$$\max_{T+1\leqslant t\leqslant \lfloor \lambda T\rfloor} \|\hat{\varphi}_t - \varphi\| = O_p(T^{-1/2}).$$

Proof of Lemma A.1. Observe first that

$$\max_t \|\hat{\varphi}_t - \varphi\| \;\leqslant\; \max_t \left| \left(\frac{1}{T}\sum_{j=1}^{t} g_j g_j' \right)^{-1} \left(\frac{1}{T}\sum_{j=1}^{t} g_j \varepsilon_j \right) \right|.$$

Under Assumption 1, for the same reason as noted in Remark 2.6, for large T, the minimum eigenvalue of $\frac{1}{T}\sum_{j=1}^{t} g_j g_j'$, which we denote by $\lambda_{\min}$, will be positive. Using a standard matrix norm inequality[1] we then have that

$$\max_t \left| \left(\frac{1}{T}\sum_{j=1}^{t} g_j g_j' \right)^{-1} \left(\frac{1}{T}\sum_{j=1}^{t} g_j \varepsilon_j \right) \right| \leqslant \mathbb{C}\frac{\max_t \left| \frac{1}{T}\sum_{j=1}^{t} g_j \varepsilon_j \right|}{\lambda_{\min}},$$

where $\mathbb{C}$ is a generic positive constant.

Next, we show that $\max_t \left| \frac{1}{T}\sum_{j=1}^{t} g_j \varepsilon_j \right| = O_p(T^{-1/2})$. To do so, observe that $(1/T)\sum_{j=1}^{t} g_j \varepsilon_j$ is a vector of martingales. Using the definition of the Euclidean norm for vectors, and denoting the kth element of g_j as $g_{j,k}$, to show the claimed order result, it is sufficient to establish that $\max_t \left(\frac{1}{T}\sum_{j=1}^{t} g_{j,k}\varepsilon_j \right)^2 = O_p(T^{-1})$ for any $k = 1, \ldots, K$, which follows straightforwardly from Doob's maximal inequality for martingales and the moment assumptions imposed by Assumption 2. The stated order result is therefore established. □

Lemma A.2. *Let the conditions of Theorem 1 hold. Then, under H_0, we have that*

$$\sum_{t=1}^{\lfloor \lambda T \rfloor} \|g_t\|^2 = O_p(T) \quad \text{and} \quad \sum_{t=1}^{\lfloor \lambda T \rfloor} \|\varepsilon_t g_t\|^2 = O_p(T).$$

Proof of Lemma A.2. Consider the first result. Using the definition of the Euclidean norm, $\sum_{t=1}^{\lfloor \lambda T \rfloor} \|g_t\|^2$ is the sum of squares of each element of the vector z_t all added together. Under Assumption 2, observe that even if Δy_t and w_t are nonstationary due to the presence of unconditional heteroskedasticity, this does not alter the order in probability of the sums of their squares from the case where they are unconditionally homoskedastic, and so we can apply the same approach as used in proving Lemma 3.1 (a) of Chang and Park (2002) (pp 442), to obtain that $\sum_{t=1}^{\lfloor \lambda T \rfloor} (\Delta y_{t-k})^2 = O_p(T)$ for $k = 1, \ldots, p$ and $\sum_{t=1}^{\lfloor \lambda T \rfloor} w_{i,t-k}^2 = O_p(T)$, for $i = 1, \ldots, m; k = 1, \ldots, q$. We therefore have that $\sum_{t=1}^{\lfloor \lambda T \rfloor} \|g_t\|^2 = O_p(T)$. The second result can be derived in a similar way. □

[1]That is, $\|M\| \leqslant \sqrt{r}\|M\|_2$, where r is the rank of M and $\|M\|_2$ is the 2-norm which is defined as the square root of the maximum eigenvalue of M, and apply it to the positive definite matrix $\left(\frac{1}{T}\sum_{j=1}^{t} g_j g_j' \right)^{-1}$.

Lemma A.3. *If Assumptions 1-3 hold, then*

$$\max_j \left| \sum_{s=0}^{N} k_s \sigma_{j-s}^2 (\eta_{j-s}^2 - 1) \right| = o_p(1).$$

Proof of Lemma A.3. The stated result can be proved using the same strategy as used in the proof of Lemma A1 of AHLTZ. The only difference relative to that case is that $(\eta_{j-s}^2 - 1)$ is no longer a martingale difference sequence. In our setting, it is a mixing sequence satisfying Assumption 2(c). Establishing the stochastic orders of the terms $E(\sum_{l=1}^{\lfloor \lambda T \rfloor} \cos(tl)\sigma_l^2(\eta_l^2 - 1))^2$ and $E(\sum_{l=1}^{\lfloor \lambda T \rfloor} \sin(tl)\sigma_l^2(\eta_l^2 - 1))^2$ will therefore need to be done differently. For the first of these, notice first that:

$$\begin{aligned} & E\left(\sum_{l=1}^{\lfloor \lambda T \rfloor} \cos(tl)\sigma_l^2(\eta_l^2 - 1)\right)^2 \\ = & \sum_{l=1}^{\lfloor \lambda T \rfloor} \cos^2(tl)\sigma_l^4 E(\eta_l^2 - 1)^2 + 2 \sum_{l > l' = 1}^{\lfloor \lambda T \rfloor} \cos(tl)\cos(tl')\sigma_l^2 \sigma_{l'}^2 E(\eta_l^2 - 1)(\eta_{l'}^2 - 1). \end{aligned}$$

Then by the uniform boundedness of the volatility function and the existence of the $(2r)$th moment of ε_l, with $r > 2$, the first term can be seen to be of $O(T)$. Because of the mixing Assumption 2(c), the second term is also of $O(T)$. It therefore follows that $E(\sum_{l=1}^{\lfloor \lambda T \rfloor} \cos(tl)\sigma_l^2(\eta_l^2 - 1))^2 = O(T)$. In similar fashion it can be established that $E(\sum_{l=1}^{\lfloor \lambda T \rfloor} \sin(tl)\sigma_l^2(\eta_l^2 - 1))^2$ is also of $O(T)$. The remainder of the proof then follows exactly the same lines as the proof of Lemma A1 of AHLTZ. □

Next, in Lemma (A.4), we establish a uniform consistency result for the sequence of nonparametric variance estimators, $\tilde{\sigma}_{j,N,t}^2$, across $T + 1 \leqslant j \leqslant t$, for $T + 1 \leqslant t \leqslant \lfloor \lambda T \rfloor$.

Lemma A.4. *Let the conditions of Theorem 1 hold. Then, under $H_0 : \delta = 0$, if $T, N \to \infty$ such that $N/T \to 0$ and $N^2/T \to \infty$, then for $T + 1 \leqslant t \leqslant \lfloor \lambda T \rfloor$,*

$$\max_t \max_{T+1 \leqslant j \leqslant t} |\tilde{\sigma}_{j,N,t}^2 - \sigma_j^2| = o_p(1).$$

Proof of Lemma A.4. First, we have the decomposition

$$\begin{aligned}
\tilde{\sigma}^2_{j,N,t} - \sigma^2_j &= \sum_{s=0}^{N} k_s(\Delta y_{j-s} - \hat{\varphi}_t g_{j-s})^2 - \sigma^2_j \\
&= \sum_{s=0}^{N} k_s \sigma^2_{j-s}(\eta^2_{j-s} - 1) + \left(\sum_{s=0}^{N} k_s \sigma^2_{j-s} - \sigma^2_j\right) \\
&\quad + \sum_{s=0}^{N} k_s((\varphi - \hat{\varphi}_t)' g_{j-s})^2 + 2\sum_{s=0}^{N} k_s \varepsilon_{j-s}((\varphi - \hat{\varphi}_t)' g_{j-s}) \\
&=: A_{1,j} + A_{2,j} + A_{3,j} + A_{4,j}, \qquad \text{(A.13)}
\end{aligned}$$

where $A_{1,j}, A_{2,j}, A_{3,j}$ and $A_{4,j}$ are defined implicitly.

By Lemma A.3, we have that $\max_j |A_{1,j}| = o_p(1)$. In view of the proof of Lemma 1 in AHLTZ, we have that $\max_j |A_{2,j}|$ is also of $o_p(1)$. For $A_{3,j}$ and $A_{4,j}$, as in the proof of Lemma A(i) in Xu and Phillips (2008), we have that:

$$\begin{aligned}
\max_j \sum_{s=0}^{N} k_s((\varphi - \hat{\varphi}_t)' g_{j-s})^2 &\leqslant \max_t \|\varphi - \hat{\varphi}_t\|^2 \max_{0 \leqslant s \leqslant N} k_s \max_j \sum_{s=0}^{N} \|g_{j-s}\|^2 \\
&\leqslant \max_t \|\varphi - \hat{\varphi}_t\|^2 \max_{0 \leqslant s \leqslant N} k_s \sum_{t=1}^{\lfloor \lambda T \rfloor} \|g_t\|^2 \\
&= O_p(T^{-1}) O(N^{-1}) O_p(T) = O_p(N^{-1}),
\end{aligned}$$

and

$$\begin{aligned}
\max_j \left| \sum_{s=0}^{N} k_s \eta_{j-s}((\varphi - \hat{\varphi}_{j-s-1})' g_{j-s}) \right| &\leqslant \max_t \|\varphi - \hat{\varphi}_t\| \sum_{s=0}^{N} k_s \|\varepsilon_{j-s} g_{j-s}\| \\
&\leqslant \max_t \|\varphi - \hat{\varphi}_t\| \left(\sum_{s=0}^{N} k_s^2\right)^{1/2} \left(\sum_{s=0}^{N} \|\varepsilon_{j-s} g_{j-s}\|^2\right)^{1/2} \\
&\leqslant \max_t \|\varphi - \hat{\varphi}_t\| \left(\sum_{s=0}^{N} k_s^2\right)^{1/2} \left(\sum_{t=1}^{\lfloor \lambda T \rfloor} \|\varepsilon_t g_t\|^2\right)^{1/2} \\
&= O_p(T^{-1/2}) O(N^{-1/2}) O_p(T^{1/2}) = O_p(N^{-1/2}),
\end{aligned}$$

where in each case we have used the order results established in Lemma A.1 and Lemma A.2. Taken together these results then establish the stated result. □

Lemma A.5. *Let the conditions of Theorem 1 hold. Then, under $H_0 : \delta = 0$, if $T, N \to \infty$ such that $N/T \to 0$ and $N^{3/2}/T \to \infty$, then for $T+1 \le t \le \lfloor \lambda T \rfloor$,*

$$\max_t \max_{T+1 \leqslant j \leqslant t} |(\tilde{\sigma}^2_{j-1,N,t} - \sigma^2_{j-1}) - (\tilde{\sigma}^2_{j,N,t} - \sigma^2_j)| = o_p(T^{-1}).$$

Proof of Lemma A.5. Again using the decomposition in (A.13), we have that

$$|(\tilde{\sigma}^2_{j-1,N,t} - \sigma^2_{j-1}) - (\tilde{\sigma}^2_{j,N,t} - \sigma^2_j)| = |(A_{1,j-1} - A_{1,j}) + (A_{2,j-1} - A_{2,j}) + (A_{3,j-1} - A_{3,j}) + (A_{4,j-1} - A_{4,j})|.$$

Let us consider the terms on the right hand side of this equation in turn. For the first term, similarly to the proof of Lemma 2 of AHLTZ we have that

$$\begin{aligned} A_{1,j} - A_{1,j-1} &= \frac{\sum_{s=1}^N \left(K\left(\frac{s}{N}\right) - K\left(\frac{s-1}{N}\right)\right)\sigma^2_{j-s}(\eta^2_{j-s} - 1)}{\sum_{s=0}^N K\left(\frac{s}{N}\right)} \\ &= \frac{1}{N} \frac{\sum_{s=1}^N K'(\tau_s)\sigma^2_{j-s}(\eta^2_{j-s} - 1)}{\sum_{s=0}^N K\left(\frac{s}{N}\right)}, \end{aligned}$$

where we have used the fact that $K(0) = K(1) = 0$, together with an application of the mean value theorem. Then using the same strategy in analysing the mixing sequence $(\eta^2_{j-s} - 1)$ as used in the proof of Lemma A.3, coupled with the absolute integrability assumption placed on the characteristic function of $K'(\cdot)$ under Assumption 3, we obtain that $\max_j \left|\sum_{s=1}^N K'(\tau_s)\sigma^2_{j-s}(\eta^2_{j-s} - 1)\right| = O_p\left(\sqrt{N}\right)$, and hence that $\max_j |A_{1,j} - A_{1,j-1}| = O_p(N^{-3/2}) = o_p(1/T)$.

For the second term, in view of the proof of Lemma 2 in AHLTZ, it holds that $\max_j |A_{2,j} - A_{2,j-1}| = o_p(1/T)$.

Turning to the third term, we have that

$$\begin{aligned} A_{3,j} - A_{3,j-1} &= \sum_{s=0}^N k_s((\varphi - \hat{\varphi}_t)' g_{j-s})^2 - \sum_{s=0}^N k_s((\varphi - \hat{\varphi}_t)' g_{j-s-1})^2 \\ &= \sum_{s=1}^N (k_s - k_{s-1})((\varphi - \hat{\varphi}_t)' g_{j-s})^2, \end{aligned}$$

and so it holds that

$$\begin{aligned} \max_j |A_{3,j} - A_{3,j-1}| &\leqslant \max_j \|\varphi - \hat{\varphi}_t\|^2 \max_{1 \leqslant s \leqslant N} |k_s - k_{s-1}| \max_j \sum_{s=0}^N \|g_{j-s}\|^2 \\ &\leqslant \max_j \|\varphi - \hat{\varphi}_t\|^2 \max_{1 \leqslant s \leqslant N} |k_s - k_{s-1}| \sum_{t=1}^{\lfloor \lambda T \rfloor} \|g_t\|^2 \\ &= O_p(T^{-1}) O(N^{-2}) O_p(T) \\ &= O_p(N^{-2}) = o_p(T^{-1}), \end{aligned}$$

where we have used the order results established in Lemma A.1 and Lemma A.2 and the fact that

$$\max_{1\leqslant s\leqslant N}|k_s-k_{s-1}|=\max_{1\leqslant s\leqslant N}\frac{\left|K\left(\frac{s}{N}\right)-K\left(\frac{s-1}{N}\right)\right|}{\sum_{s=0}^{N}K\left(\frac{s}{N}\right)}=O(N^{-2}),$$

which is a simple consequence of the mean value theorem and Assumption 3.

Similarly, for the fourth term, we have that

$$A_{4,j}-A_{4,j-1} \;=\; 2\sum_{s=1}^{N}(k_s-k_{s-1})((\varphi-\hat{\varphi}_t)'g_{j-s})\varepsilon_{j-s},$$

and so

$$\begin{aligned}\max_j|A_{4,j}-A_{4,j-1}| &\leqslant \max_j\|\varphi-\hat{\varphi}_t\|^2\left(\sum_{s=1}^{N}(k_s-k_{s-1})^2\right)^{1/2}\left(\sum_{t=1}^{\lfloor\lambda T\rfloor}\|\varepsilon_t g_t\|^2\right)^{1/2}\\ &= O_p(T^{-1/2})O(N^{-3/2})O_p(T^{1/2})=O_p(N^{-3/2})=o_p(T^{-1}),\end{aligned}$$

in view of the fact that $N^{3/2}/T\to\infty$ under Assumption 3(b).

Taken together these results therefore establish the stated result. □

Lemma A.6 gives the uniform rate of convergence for the infeasible WLS estimator. Since the proof is the same as that of Lemma A.1, we omit the proof to avoid repetition.

Lemma A.6. *Let the conditions of Theorem 1 hold. Then under H_0,*

$$\max_{T+1\leqslant t\leqslant\lfloor\lambda T\rfloor}\|\varphi_t^W-\varphi\|=O_p(T^{-1/2}).$$

Lemma A.7. *Let the conditions of Theorem 1 hold. Then under H_0,*

$$\max_{T+1\leqslant t\leqslant\lfloor\lambda T\rfloor}\|\hat{\varphi}_t^W-\varphi_t^W\|=o_p(T^{-1/2}),$$

Proof of Lemma A.7. Notice

$$\begin{aligned}\hat{\varphi}_t^W-\varphi_t^W &= (\hat{\varphi}_t^W-\varphi)-(\varphi_t^W-\varphi)\\ &= \left(\frac{1}{T}\sum_{j=1}^{t}\frac{g_jg_j'}{\tilde{\sigma}_{j,N,t}^2}\right)^{-1}\left(\frac{1}{T}\sum_{j=1}^{t}\frac{g_j\sigma_j\eta_j}{\tilde{\sigma}_{j,N,t}^2}\right)-\left(\frac{1}{T}\sum_{j=1}^{t}\frac{g_jg_j'}{\sigma_j^2}\right)^{-1}\left(\frac{1}{T}\sum_{j=1}^{t}\frac{g_j\sigma_j\eta_j}{\sigma_j^2}\right)\\ &= \left(\frac{1}{T}\sum_{j=1}^{t}\frac{g_jg_j'}{\tilde{\sigma}_{j,N,t}^2}\right)^{-1}\left(\left(\frac{1}{T}\sum_{j=1}^{t}\frac{g_j\sigma_j\eta_j}{\tilde{\sigma}_{j,N,t}^2}\right)-\left(\frac{1}{T}\sum_{j=1}^{t}\frac{g_j\sigma_j\eta_j}{\sigma_j^2}\right)\right)\\ &\quad+\left(\left(\frac{1}{T}\sum_{j=1}^{t}\frac{g_jg_j'}{\tilde{\sigma}_{j,N,t}^2}\right)^{-1}-\left(\frac{1}{T}\sum_{j=1}^{t}\frac{g_jg_j'}{\sigma_j^2}\right)^{-1}\right)\left(\frac{1}{T}\sum_{j=1}^{t}\frac{g_j\sigma_j\eta_j}{\sigma_j^2}\right)\end{aligned}$$

Using the summation by parts formula

$$\begin{aligned}
&\frac{1}{T}\sum_{j=1}^{t}\frac{g_j\sigma_j\eta_j}{\tilde{\sigma}_{j,N,t}^2}-\frac{1}{T}\sum_{j=1}^{t}\frac{g_j\sigma_j\eta_j}{\sigma_j^2}\\
=\;&\frac{1}{T}\sum_{j=1}^{t}\left(\frac{\sigma_j^2}{\tilde{\sigma}_{j,N,t}^2}\right)\frac{g_j\eta_j}{\sigma_j}-\frac{1}{T}\sum_{j=1}^{t}\frac{g_j\sigma_j\eta_j}{\sigma_j^2}\\
=\;&\left(\frac{\sigma_t^2}{\tilde{\sigma}_{t,N,t}^2}-1\right)\frac{1}{T}\sum_{j=1}^{t}\frac{g_j\eta_j}{\sigma_j}-\sum_{j=1}^{t}\left(\frac{\sigma_j^2}{\tilde{\sigma}_{j,N,t}^2}-\frac{\sigma_{j-1}^2}{\tilde{\sigma}_{j-1,N,t}^2}\right)\left(\frac{1}{T}\sum_{i=1}^{j-1}\frac{g_i\eta_i}{\sigma_i}\right).
\end{aligned}$$

Notice that Lemma A.4 implies that

$$\max_t|\tilde{\sigma}_{t,N,t}^2-\sigma_t^2|=o_p(1),$$

it follows that the above first term is $o_p(1)$. Using the proof strategy in Theorem 1 of AHLTZ, and the results of Lemma A.4 and Lemma A.5, the above second term is also $o_p(1)$. Using again Lemma A.4 and that $\frac{1}{T}\sum_{i=1}^{j-1}\frac{g_i\eta_i}{\sigma_i}=O_p(T^{-1/2})$, we have

$$\frac{1}{T}\sum_{j=1}^{t}\frac{g_j\sigma_j\eta_j}{\tilde{\sigma}_{j,N,t}^2}-\frac{1}{T}\sum_{j=1}^{t}\frac{g_j\sigma_j\eta_j}{\sigma_j^2}=o_p(T^{-1/2}).$$

Also notice that

$$\frac{1}{T}\sum_{j=1}^{t}\frac{g_jg_j'}{\tilde{\sigma}_{j,N,t}^2}=O_p(1),\quad \frac{1}{T}\sum_{j=1}^{t}\frac{g_j\sigma_j\eta_j}{\sigma_j^2}=O_p(T^{-1/2}),$$

and using similar arguments and the results of Lemma A.4, we have

$$\left(\frac{1}{T}\sum_{j=1}^{t}\frac{g_jg_j'}{\tilde{\sigma}_{j,N,t}^2}\right)^{-1}-\left(\frac{1}{T}\sum_{j=1}^{t}\frac{g_jg_j'}{\sigma_j^2}\right)^{-1}=o_p(1).$$

The result stated in the lemma then follows. □

Lemma A.8. *Let the conditions of Theorem 2 hold, and define* $\psi_T:=1+c/T^d$,

$$(T^d\wedge N)^{-1}\sum_{s=0}^{N}K(s/N)\psi_T^{-2s}=O(1),$$

Furthermore, the limit is strictly positive and nondegenerate to 0.

Proof of Lemma A.8. We first note that the following proof is valid for all $0 \leqslant d < 1$. By Assumption 3, $N^{3/2}/T \to \infty$, which implies that $N/T^{2/3} \to \infty$. Therefore, if $0 \le d \le 2/3$, the rate $T^d \wedge N$ will be T^d in the presentation of the lemma.

To prove the stated results, we first show that

$$\sum_{s=0}^{N} \psi_T^{-2s} = O(T^d \wedge N). \tag{A.14}$$

Using the formula for geometric series, we have that

$$\sum_{s=0}^{N} \psi_T^{-2s} = \psi_T^{-2N} \frac{\psi_T^{2(N+1)} - 1}{\psi_T^2 - 1}. \tag{A.15}$$

and observe that

$$\psi_T^{2(N+1)} = \left(1 + \frac{c}{T^d}\right)^{2(N+1)} = \left(1 + \frac{c}{T^d}\right)^{\frac{T^d}{c} \times 2c \times \frac{N+1}{T^d}}.$$

As $T \to \infty$, the limit of this will depend on the order of $(N+1)/T^d$. We discuss three different possibilities for this below:

(i) If $N/T^d \to \infty$, $\psi_T^{2(N+1)} \to e^{\infty} = \infty$ so $\psi_T^{2(N+1)} - 1 = \psi_T^{2(N+1)}(1 + o(1))$, and (A.15) becomes

$$\psi_T^{-2N} \frac{\psi_T^{2(N+1)}(1 + o(1))}{1 + 2\frac{c}{T^d} + o\left(\frac{c}{T^d}\right) - 1} = \psi_T^{-2N} \psi_T^{2(N+1)} T^d (1 + o(1)) = O(T^d).$$

(ii) If $N/T^d \to 0$, $\psi_T^{2(N+1)} \to e^0 = 1$, then using Taylor expansion,

$$\begin{aligned} \psi_T^{2(N+1)} - 1 &= \left(1 + \frac{c}{T^d}\right)^{2(N+1)} - 1 \\ &= \left(1 + 2(N+1)\frac{c}{T^d} + \frac{2(N+1)(2(N+1)-1)}{2!}\left(\frac{c}{T^d}\right)^2 + \cdots\right) - 1 \\ &= 2(N+1)\frac{c}{T^d}(1 + o(1)). \end{aligned}$$

It then follows that (A.15) becomes

$$\psi_T^{-2N} \frac{2(N+1)\frac{c}{T^d}(1 + o(1))}{1 + 2\frac{c}{T^d} + o\left(\frac{c}{T^d}\right) - 1} = \psi_T^{-2N}(N+1)(1 + o(1)) = O(N),$$

where we have used the fact that $\psi_T^{-2N} \to e^0 = 1$.

(iii) When $N/T^d \to \vartheta$, where ϑ is a positive constant, $\psi_T^{2(N+1)} \to e^{2c\vartheta}$, then (A.15) becomes

$$e^{-2c\vartheta} \frac{e^{2c\vartheta} - 1}{1 + a\frac{c}{T^d} + o\left(\frac{c}{T^d}\right) - 1} = O(T^d).$$

Taken together, the results in (i)-(iii) establish the stated result in (A.14).

We now turn to establishing the results stated in the lemma. Notice first that because $K(.)$ is bounded, we have

$$\sum_{s=0}^{N} K(s/N)\psi_T^{-2s} \leqslant \mathbb{C}\sum_{s=0}^{N}\psi_T^{-2s}, \tag{A.16}$$

where $\mathbb{C}$ is a generic positive constant, then the upper bound part of the result follows. To see that this is also the lower bound, notice that because the kernel is positive over the interval $(0,1)$, then there exists a closed interval $[\delta, 1-\delta]$ for an arbitrarily small $\delta > 0$, such that $K(.) \geqslant g_0 > 0$, by Assumption 3. It then follows that

$$\begin{aligned}\sum_{s=0}^{N} K(s/N)\psi_T^{-2s} &> \sum_{s=0}^{N} K(s/N)\psi_T^{-2s}\mathbb{I}(\delta \leqslant s/N \leqslant 1-\delta)\\ &\geqslant g_0\sum_{s=0}^{N}\psi_T^{-2s}\mathbb{I}(\delta \leqslant s/N \leqslant 1-\delta)\\ &= g_0\sum_{s=\delta N}^{N-\delta N}\psi_T^{-2s}.\end{aligned}$$

Letting $\delta \to 0$, it then follows straightforwardly that this term is positive and has the same order as $\sum_{s=0}^{N}\psi_T^{-2s}$. □

In the next lemma, we analyse the asymptotic behaviour of the volatility estimator under explosive alternatives. The volatility estimator is constructed with the null hypothesis imposed, by smoothing past squared residuals. When there is a structural change leading to an explosive regime, the null model becomes misspecified, and the volatility estimator may lose its consistency or even diverge. The next lemma is an intermediate result needed in the proof of Theorem 2.

Lemma A.9. *Under the conditions of Theorem 2, let ξ_T be a sequence such that $\xi_T/T \to 0$ and $\xi_T/N \to \infty$, it holds that*

$$\max_{\lfloor\tau T\rfloor+\xi_T+1\leqslant t\leqslant\lfloor\lambda T\rfloor}\left|\left(\frac{T^{2d}N}{T^d \wedge N}\right)T^{-1}\psi_T^{-2(t-1-\lfloor\tau T\rfloor)}\tilde{\sigma}^2_{t,N,t}\right| = O_p(1).$$

Moreover, $\min_{\lfloor\tau T\rfloor+\xi_T+1\leqslant t\leqslant\lfloor\lambda T\rfloor}\left|\left(\frac{T^{2d}N}{T^d\wedge N}\right)T^{-1}\psi_T^{-2(t-1-\lfloor\tau T\rfloor)}\tilde{\sigma}^2_{t,N,t}\right|$ *is, with probability 1, strictly positive.*

Proof of Lemma A.9. Note again first that the result of the lemma and the proof are valid for $0 \leq d < 1$. When $0 \leqslant d \leqslant 2/3$, the result of the lemma becomes

$$\max_{\lfloor \tau T \rfloor + \xi_T + 1 \leqslant t \leqslant \lfloor \lambda T \rfloor} \left| T^{d-1} N \psi_T^{-2(t-1-\lfloor \tau T \rfloor)} \tilde{\sigma}^2_{t,N,t} \right| = O_p(1).$$

Now we prove the lemma. We will first establish the result that

$$\max_{\lfloor \tau T \rfloor + 1 \leqslant t \leqslant \lfloor \lambda T \rfloor} |T^{-1/2} \psi_T^{-(t-\lfloor \tau T \rfloor)} u_t| = O_p(1). \tag{A.17}$$

To that end, it is convenient to use the formulation in (3). First notice that $u_t = O_p(\sqrt{T})$ during the unit root regime (i.e. before the inception of the explosive regime). Now for $\lfloor \tau T \rfloor + 1 \leqslant t \leqslant \lfloor \lambda T \rfloor$, by repeated backward substitution, we have that

$$u_t = \sum_{j=\lfloor \tau T \rfloor + 1}^{t} \psi_T^{t-j} v_j + \psi_T^{t-\lfloor \tau T \rfloor} u_{\lfloor \tau T \rfloor}, \tag{A.18}$$

where $u_{\lfloor \tau T \rfloor}$ is the last observation in the unit root regime and serves as the initial condition for the explosive regime. For the first term on the right hand side of (A.18), observe that

$$\begin{aligned} E \max_{\lfloor \tau T \rfloor + 1 \leqslant t \leqslant \lfloor \lambda T \rfloor} \psi_T^{-(t-\lfloor \tau T \rfloor)} \left| \sum_{j=\lfloor \tau T \rfloor + 1}^{t} \psi_T^{t-j} v_j \right| &\leqslant \psi_T^{\lfloor \tau T \rfloor} E \max_{\lfloor \tau T \rfloor + 1 \leqslant t \leqslant \lfloor \lambda T \rfloor} \sum_{j=\lfloor \tau T \rfloor + 1}^{t} |\psi_T^{-j} v_j| \\ &= \psi_T^{\lfloor \tau T \rfloor} E \sum_{j=\lfloor \tau T \rfloor + 1}^{\lfloor \lambda T \rfloor} |\psi_T^{-j} v_j| = O(1), \end{aligned}$$

and, hence, it follows that $\max_{\lfloor \tau T \rfloor + 1 \leqslant t \leqslant \lfloor \lambda T \rfloor} \psi_T^{-(t-\lfloor \tau T \rfloor)} \left| \sum_{j=\lfloor \tau T \rfloor + 1}^{t} \psi_T^{t-j} v_j \right| = O_p(1)$. Turning to the second term, we have that

$$\max_{\lfloor \tau T \rfloor + 1 \leqslant t \leqslant \lfloor \lambda T \rfloor} |\psi_T^{-(t-\lfloor \tau T \rfloor)} \psi_T^{t-\lfloor \tau T \rfloor} u_{\lfloor \tau T \rfloor}| = |u_{\lfloor \tau T \rfloor}| = O_p\left(\sqrt{T}\right).$$

The second term on the right hand side of (A.18) is therefore the dominant term and the result in (A.17) is established.

We are now in a position to establish the results stated in the lemma. Defining $\delta_T = c/T^d$, when $\lfloor \tau T \rfloor + \xi_T + 1 \leqslant t \leqslant \lfloor \lambda T \rfloor$, we have that

$$
\begin{aligned}
\tilde{\sigma}^2_{t,N,t} &= \sum_{s=0}^{N} k_s f^2_{t-s,t} = \sum_{s=0}^{N} k_s (\Delta y_{t-s} - \hat{\varphi}_t g_{t-s})^2 \\
&= \sum_{s=0}^{N} k_s (\delta_T u_{t-s-1} + v_{t-s} - \hat{\varphi}_t g_{t-s})^2 \\
&= \delta_T^2 \sum_{s=0}^{N} k_s u^2_{t-s-1} + \sum_{s=0}^{N} k_s v^2_{t-s} + \sum_{s=0}^{N} k_s (\hat{\varphi}_t g_{t-s})^2 + 2\delta_T \sum_{s=0}^{N} k_s u_{t-s-1} v_{t-s} \\
&\quad +2\delta_T \sum_{s=0}^{N} k_s u_{t-s-1} (\hat{\varphi}_t g_{t-s}) + 2 \sum_{s=0}^{N} k_s v_{t-s} (\hat{\varphi}_t g_{t-s}) \\
&=: D_1 + D_2 + D_3 + D_4 + D_5 + D_6,
\end{aligned}
$$

where the D_j, $j = 1, ..., 6$, terms are implicitly defined.

Let us consider the terms $D_1, \ldots, D_6$ in turn. First, notice that D_1 satisfies

$$
\begin{aligned}
& \max_{\lfloor \tau T \rfloor + \xi_T + 1 \leqslant t \leqslant \lfloor \lambda T \rfloor} T^{-1} \psi_T^{-2(t-1-\lfloor \tau T \rfloor)} D_1 \\
\leqslant \;& \delta_T^2 \max_{\lfloor \tau T \rfloor + \xi_T + 1 \leqslant t \leqslant \lfloor \lambda T \rfloor, 0 \leqslant s \leqslant N} |T^{-1} \psi_T^{-2(t-s-1-\lfloor \tau T \rfloor)} u^2_{t-s-1}| \left(\frac{\sum_{s=0}^{N} K(s/N) \psi_T^{-2s}}{\sum_{s=0}^{N} K(s/N)} \right) \\
= \;& O_p\left(\left(\frac{T^d \wedge N}{T^{2d} N} \right) \right),
\end{aligned}
$$

where we have used (A.17) and the result that $\sum_{s=0}^{N} K(s/N) \psi_T^{-2s} = O(T^d \wedge N)$ and $\sum_{s=0}^{N} K(s/N) = O(N)$. Next, notice that $\max_{\lfloor \tau T \rfloor + \xi_T + 1 \leqslant t \leqslant \lfloor \lambda T \rfloor} T^{-1} \psi_T^{-2(t-1-\lfloor \tau T \rfloor)} D_2 = o_p(1)$, because D_2 is a weighted sum of $O_p(1)$ terms, and notice that both $T^{-1} \to 0$ and $\psi_T^{-2(t-1-\lfloor \tau T \rfloor)} \to 0$ (because $t - 1 - \lfloor \tau T \rfloor > \xi_T$). Next, for D_3, observe that

$$
\begin{aligned}
& \max_{\lfloor \tau T \rfloor + \xi_T + 1 \leqslant t \leqslant \lfloor \lambda T \rfloor} T^{-1} \psi_T^{-2(t-1-\lfloor \tau T \rfloor)} D_3 \\
\leqslant \;& \max_{0 \leqslant s \leqslant N} k_s \max_{\lfloor \tau T \rfloor + \xi_T + 1 \leqslant t \leqslant \lfloor \lambda T \rfloor} T^{-1} \psi_T^{-2(t-1-\lfloor \tau T \rfloor)} \|\hat{\varphi}_t\| \sum_{s=0}^{N} \|g_{t-s}\|^2
\end{aligned}
$$

Notice that $\|g_{t-s}\|^2$ is the sum of the squares of each element of the g_{t-s} vector. The first element of g_{t-s} is 1, and the second element is Δy_{t-s-1}, which is $\delta_T u_{t-s-2} + v_{t-s-2}$. Using the same strategy as was used above in analysing D_1, we know that this element's contribution cannot be larger than D_1. Similarly the contribution of all of the third to $(p+1)$th elements is no larger than D_1, while the remaining elements are related to 1 and w_t which are clearly seen to be dominated by D_1. D_4, D_5 and D_6 are cross product terms which therefore cannot be the largest of the six terms. Thus D_1 is dominant and we have that

$$\max_{\lfloor\tau T\rfloor+\xi_T+1\leqslant t\leqslant\lfloor\lambda T\rfloor}\left|\left(\frac{T^d\wedge N}{T^{2d}N}\right)^{-1}T^{-1}\psi_T^{-2(t-1-\lfloor\tau T\rfloor)}\tilde{\sigma}^2_{t,N,t}\right| = O_p(1).$$

Notice that D_1 also satisfies

$$\begin{aligned}
&\min_{\lfloor\tau T\rfloor+\xi_T+1\leqslant t\leqslant\lfloor\lambda T\rfloor}\left(\frac{T^d\wedge N}{T^{2d}N}\right)^{-1}T^{-1}\psi_T^{-2(t-1-\lfloor\tau T\rfloor)}D_1\\
= &\min_{\lfloor\tau T\rfloor+\xi_T+1\leqslant t\leqslant\lfloor\lambda T\rfloor}\left(\frac{T^d\wedge N}{T^{2d}N}\right)^{-1}T^{-1}\psi_T^{-2(t-1-\lfloor\tau T\rfloor)}\delta_T^2\sum_{s=0}^{N}k_s u^2_{t-s-1}\\
\geqslant &\left(\frac{T^d\wedge N}{T^{2d}N}\right)^{-1}\delta_T^2\left(\sum_{s=0}^{N}k_s\psi_T^{-2s}\right)\times\min_{\lfloor\tau T\rfloor+\xi_T+1\leqslant t\leqslant\lfloor\lambda T\rfloor,0\leqslant s\leqslant N}|T^{-1}\psi_T^{-2(t-s-1-\lfloor\tau T\rfloor)}u^2_{t-s-1}|
\end{aligned}$$

First, notice that $\sum_{s=0}^{N}k_s\psi_T^{-2s} = O((T^d\wedge N)/N)$, it follows that

$$\left(\frac{T^d\wedge N}{T^{2d}N}\right)^{-1}\delta_T^2\left(\sum_{s=0}^{N}k_s\psi_T^{-2s}\right) = O(1).$$

For $\min_{\lfloor\tau T\rfloor+\xi_T+1\leqslant t\leqslant\lfloor\lambda T\rfloor,0\leqslant s\leqslant N}|T^{-1}\psi_T^{-2(t-s-1-\lfloor\tau T\rfloor)}u^2_{t-s-1}|$, notice that when $t\geqslant\lfloor\tau T\rfloor+\xi_T+1$, the index $t-s-1$ always satisfies $\lfloor\tau T\rfloor+1\leqslant t-s-1\leqslant\lfloor\lambda T\rfloor$, because $0\leqslant s\leqslant N$ and $\xi_T/N\to\infty$. It follows that

$$\min_{\lfloor\tau T\rfloor+\xi_T+1\leqslant t\leqslant\lfloor\lambda T\rfloor,0\leqslant s\leqslant N}|T^{-1}\psi_T^{-2(t-s-1-\lfloor\tau T\rfloor)}u^2_{t-s-1}| \geqslant \min_{\lfloor\tau T\rfloor+1\leqslant t\leqslant\lfloor\lambda T\rfloor}|T^{-1}\psi_T^{-2(t-\lfloor\tau T\rfloor)}u_t^2|.$$

Applying the backward substitution (A.18) for u_t again we have $\min_{\lfloor\tau T\rfloor+1\leqslant t\leqslant\lfloor\lambda T\rfloor}|T^{-1}\psi_T^{-2(t-\lfloor\tau T\rfloor)}u_t^2|$ is $O_p(1)$ due to its dominating second term related to the initial value of the explosive regime $T^{-1}u^2_{\lfloor\tau T\rfloor}$ being $O_p(1)$ but not $o_p(1)$. Consequently, the stated results are established. □

In the next Lemma A.10, we generalise Lemma 3 of KPA, which proves a weak convergence result of a partial sum of vector martingale differences to a vector Brownian motion. In our context, because we allow for the presence of unconditional heteroskedasticity in the covariates and in the regression errors, the corresponding partial sums of vector martingale differences will converge to a more general vector Gaussian process.

Lemma A.10. *Let ζ_t be a martingale difference sequence with respect to some filtration $\{\mathcal{F}_t, t \geqslant 1\}$, with $E(\zeta_t^2|\mathcal{F}_{t-1}) = 1$, and let r_t be a sequence of ℓ-dimensional random vectors measurable with respect to $\mathcal{F}_{t-1}$. Let $R(\kappa)$ be a positive definite $\ell \times \ell$ matrix for each $\kappa \in [0, \lambda]$, with each of its elements being finite and continuous in κ, then the partial sum process $G^{(T)}(\cdot) := \left(1/\sqrt{T}\right) \sum_{t=1}^{\lfloor T\cdot \rfloor} r_t\zeta_t$ converges weakly to an ℓ-dimensional Gaussian process $G(\cdot)$ with covariance structure*

$$E\left(G(\kappa_1)G(\kappa_2)'\right) = R(\min(\kappa_1, \kappa_2)),$$

for $\kappa_1, \kappa_2 \in [0, \lambda]$, where $\min(\kappa_1, \kappa_2)$ denotes the smaller of κ_1 and κ_2, provided

$$\text{plim}_{T\to\infty} \frac{1}{T} \sum_{t=1}^{\lfloor T\kappa \rfloor} r_t r_t' = \lim_{T\to\infty} E\left(\frac{1}{T}\sum_{t=1}^{\lfloor T\kappa \rfloor} r_t r_t'\right) = R(\kappa), \tag{A.19}$$

and

$$\limsup_{T\to\infty} \frac{1}{\lfloor T\lambda \rfloor} \sum_{t=1}^{\lfloor T\lambda \rfloor} E\|r_t\zeta_t\|^{2+\delta} < \infty, \tag{A.20}$$

for some $\delta > 0$.

Proof of Lemma A.10. For simplicity, we prove the stated result for the case when $\ell = 1$. As argued in Section 29.7 of Davidson (2021), the extension to the $\ell > 1$ case can be easily obtained by applying the Cramer-Wold device.

First define an element Y_T in $C[0, \lambda]$, which is an interpolated version of $G^{(T)}(\kappa) = \left(1/\sqrt{T}\right) \sum_{t=1}^{\lfloor T\kappa \rfloor} r_t\zeta_t$; that is,

$$Y^{(T)}(\kappa) := \begin{cases} G^{(T)}(\kappa) + (T\kappa - \lfloor T\kappa \rfloor)\left(1/\sqrt{T}\right) r_{\lfloor T\kappa \rfloor+1}\zeta_{\lfloor T\kappa \rfloor+1} & \kappa < \lambda \\ G^{(T)}(\lambda) & \kappa = \lambda \end{cases}.$$

Notice that

$$\sup_{\kappa\in[0,\lambda]} |Y^{(T)}(\kappa) - G^{(T)}(\kappa)| = (T\kappa - \lfloor T\kappa \rfloor)\left(1/\sqrt{T}\right) |r_{\lfloor T\kappa \rfloor+1}\zeta_{\lfloor T\kappa \rfloor+1}| \xrightarrow{p} 0,$$

so that $Y^{(T)}(\cdot)$ and $G^{(T)}(\cdot)$ have the same weak limit.

Using conditions (A.19) and (A.20), and by the multivariate central limit theorem for martingale difference sequences (see, e.g., Corollary 3.1 of Hall and Heyde, 1980), it is straightforward to show that for $0 < \kappa_1 < \cdots < \kappa_k < \lambda$, where $k > 1$ is a finite integer, we have that

$$\begin{pmatrix} Y^{(T)}(\kappa_1) \\ Y^{(T)}(\kappa_2) \\ \dots \\ Y^{(T)}(\kappa_k) \end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix} 0 \\ 0 \\ \dots \\ 0 \end{pmatrix}, \begin{pmatrix} R(\kappa_1) & R(\kappa_1) & \dots & R(\kappa_1) \\ R(\kappa_1) & R(\kappa_2) & \dots & R(\kappa_2) \\ \dots & \dots & \dots & \dots \\ R(\kappa_1) & R(\kappa_2) & \dots & R(\kappa_k) \end{pmatrix}\right).$$

Next we show $Y^{(T)}(\cdot)$ is uniformly tight. By definition, for $\lambda \geqslant \kappa_2 > \kappa_1 \geqslant 0$

$$Y^{(T)}(\kappa_2) - Y^{(T)}(\kappa_1) = G^{(T)}(\kappa_2) - G^{(T)}(\kappa_1) + R^{(T)}(\kappa_2, \kappa_1),$$

where

$$R^{(T)}(\kappa_2, \kappa_1) := (T\kappa_2 - \lfloor T\kappa_2 \rfloor)\left(1/\sqrt{T}\right) r_{\lfloor T\kappa_2 \rfloor + 1}\zeta_{\lfloor T\kappa_2 \rfloor + 1} - (T\kappa_1 - \lfloor T\kappa_1 \rfloor)\left(1/\sqrt{T}\right) r_{\lfloor T\kappa_1 \rfloor + 1}\zeta_{\lfloor T\kappa_1 \rfloor + 1}.$$

Using the maximal inequality for martingales (see, e.g., Corollary 16.20 of Davidson, 2021), we have that

$$\begin{aligned} \sup_{0<\kappa_1<\kappa_2<1} \left| \sum_{t=1}^{\lfloor T\kappa_2 \rfloor} r_t\zeta_t - \sum_{t=1}^{\lfloor T\kappa_1 \rfloor} r_t\zeta_t \right| &= \sup_{0<\kappa_1<\kappa_2<1} \left| \sum_{t=\lfloor T\kappa_1 \rfloor}^{\lfloor T\kappa_2 \rfloor} r_t\zeta_t \right| \\ &\leqslant \max_{1\leqslant k\leqslant T} \left| \sum_{t=1}^{k} r_t\zeta_t \right|, \end{aligned}$$

and so

$$P\left(\frac{1}{\sqrt{T}} \sup_{0<\kappa_1<\kappa_2<1} \left| \sum_{t=1}^{\lfloor T\kappa_2 \rfloor} r_t\zeta_t - \sum_{t=1}^{\lfloor T\kappa_1 \rfloor} r_t\zeta_t \right| > \varepsilon\right) < P\left(\frac{1}{\sqrt{T}} \max_{1\leqslant k\leqslant T} \left| \sum_{t=1}^{k} r_t\zeta_t \right| > \varepsilon\right) \leqslant \frac{E\left|\sum_{t=1}^{T} r_t\zeta_t\right|^p}{T^{p/2}\varepsilon^p},$$

which, by Burkholder's inequality (see Theorem 16.24 of Davidson, 2021), is bounded by a constant when taking $p = 2$. For the term $R^{(T)}(\kappa_1, \kappa_2)$, notice that

$$P\left(\max_{1\leqslant t\leqslant T}\left(1/\sqrt{T}\right)|r_t\zeta_t| > \varepsilon\right) \leqslant \sum_{t=1}^{T} P\left(\left(1/\sqrt{T}\right)|r_t\zeta_t| > \varepsilon\right) \leqslant T\frac{E|r_t\zeta_t|^p}{T^{p/2}\varepsilon^p},$$

which is bounded by a constant when taking $p = 4$.

It therefore follows that the stochastic continuity condition (29.58) in Theorem 29.17 of Davidson (2021) is satisfied. Moreover, condition (29.57) in Theorem 29.17 of Davidson (2021) is trivially satisfied in our context. Therefore, the tightness condition is satisfied. Applying Theorem 7.1 of Billingsley (1999), the claimed weak convergence result then follows. □

A.2.2 Proofs of Theorems

Proof of Theorem 1. Observe first that for $1 < r \leq \lambda$,

$$\frac{1}{\sqrt{T}} \sum_{j=T+1}^{\lfloor Tr \rfloor} \frac{\hat{e}_j^W}{\tilde{\sigma}_{j,N,j}} = \frac{1}{\sqrt{T}} \sum_{j=T+1}^{\lfloor Tr \rfloor} \frac{\sigma_j \eta_j}{\tilde{\sigma}_{j,N,j}} - \frac{1}{\sqrt{T}} \sum_{j=T+1}^{\lfloor Tr \rfloor} \frac{(\hat{\varphi}_{j-1}^W - \varphi)' g_j}{\tilde{\sigma}_{j,N,j}}. \tag{A.21}$$

The proof of the theorem will be constructed in two parts. In the first part, we will show the stated limiting distribution is valid in the case when the volatility function is known, by extending KPA's proof for their Theorem 1. We will then show that the result continues to hold where the volatility function is estimated.

Let us establish the first part. The proof for this part will use a weak convergence result for the vector $\left(\frac{1}{\sqrt{T}} \sum_{j=1}^{\lfloor Tr \rfloor} \frac{g_j \eta_j}{\sigma_j}, \frac{1}{\sqrt{T}} \sum_{j=1}^{\lfloor Tr \rfloor} \eta_j\right)'$. Using Lemma A.10, under the conditions of Theorem 1, we have,

$$\begin{pmatrix} \frac{1}{\sqrt{T}} \sum_{j=1}^{\lfloor Tr \rfloor} \frac{g_j \eta_j}{\sigma_j} \\ \frac{1}{\sqrt{T}} \sum_{j=1}^{\lfloor Tr \rfloor} \eta_j \end{pmatrix} \Rightarrow \mathbb{V}(r), \quad r \in [0, \lambda], \tag{A.22}$$

where $\mathbb{V}(.)$ is a $K+2$ dimensional Gaussian process,where $K := p+(q+1)m$. The elements of the covariance matrix of $\mathbb{V}(\cdot)$ can be established as follows. First, using Assumption 2(e) that h_j is uncorrelated with $g_j g_j'$, we have that

$$\begin{aligned}
\lim_{T\to\infty} E\left[\frac{1}{T}\sum_{j=1}^{\lfloor Tr \rfloor} g_j g_j' \frac{\eta_j^2}{\sigma_j^2}\right] &= \lim_{T\to\infty} E\left[\frac{1}{T}\sum_{j=1}^{\lfloor Tr \rfloor} g_j g_j' \frac{E(\eta_j^2|\mathcal{F}_{j-1})}{\sigma_j^2}\right] \\
&= \lim_{T\to\infty} E\left[\frac{1}{T}\sum_{j=1}^{\lfloor Tr \rfloor} g_j g_j' \frac{h_j}{\sigma_j^2}\right] \\
&= \lim_{T\to\infty} E\left[\frac{1}{T}\sum_{j=1}^{\lfloor Tr \rfloor} g_j g_j' \frac{1}{\sigma_j^2}\right] = r \cdot \Theta(r).
\end{aligned}$$

Next, using Assumption 2(b), we have that

$$\begin{aligned}
\lim_{T\to\infty} E\left[\frac{1}{T}\sum_{j=1}^{\lfloor Tr \rfloor} \eta_j^2\right] &= \lim_{T\to\infty} E\left[\frac{1}{T}\sum_{j=1}^{\lfloor Tr \rfloor} E(\eta_j^2|\mathcal{F}_{j-1})\right] \\
&= \lim_{T\to\infty} E\left[\frac{1}{T}\sum_{j=1}^{\lfloor Tr \rfloor} h_j\right] = r.
\end{aligned}$$

Next we have that

$$\begin{aligned}
\lim_{T\to\infty} E\left[\frac{1}{T}\sum_{j=1}^{\lfloor Tr\rfloor}\frac{g_j\eta_j^2}{\sigma_j}\right] &= \lim_{T\to\infty} E\left[\frac{1}{T}\sum_{j=1}^{\lfloor Tr\rfloor}\frac{g_jE(\eta_j^2|\mathcal{F}_{j-1})}{\sigma_j}\right] \\
&= \lim_{T\to\infty} E\left[\frac{1}{T}\sum_{j=1}^{\lfloor Tr\rfloor}\frac{g_jh_j}{\sigma_j}\right] \\
&= \lim_{T\to\infty} \frac{1}{T}\sum_{j=1}^{\lfloor Tr\rfloor}\frac{E(g_jh_j)}{\sigma_j} \\
&= \lim_{T\to\infty} \frac{1}{T}\sum_{j=1}^{\lfloor Tr\rfloor}\frac{E(g_j)E(h_j)}{\sigma_j} \qquad \text{(A.23)}
\end{aligned}$$

where the last line follows from the condition that $\text{cov}(g_t, h_t) = 0$, imposed by Assumption 2(e). Now, by the definition of the vector of regressors, under the null, we have that $E(g_j) := \gamma = (1, 0, \ldots, 0, c_x')'$, a vector whose first element is 1, its next p elements are all zero, and where c_x is the $(q+1)m$-dimensional mean vector of the covariates, x_t, as defined in Assumption 1. By Assumption 2(b), we have $E(h_j) = 1$, and so (A.23) is equal to

$$\gamma \lim_{T\to\infty}\frac{1}{T}\sum_{j=1}^{\lfloor Tr\rfloor}\frac{1}{\sigma_j} = \gamma\int_0^r \frac{1}{\sigma(x)}\mathrm{d}x.$$

Consequently, by Lemma A.10, the covariance matrix of the Gaussian process $\mathbb{V}(\cdot)$ takes the form

$$E[\mathbb{V}(r)\mathbb{V}(s)'] = \begin{pmatrix} (\min(r,s))\Theta(\min(r,s)) & \gamma\int_0^{\min(r,s)}\frac{1}{\sigma(x)}\mathrm{d}x \\ \gamma'\int_0^{\min(r,s)}\frac{1}{\sigma(x)}\mathrm{d}x & (\min(r,s)) \end{pmatrix}, \qquad \text{(A.24)}$$

for $r, s \in [0, \lambda]$.

When $\tilde{\sigma}_{j,N,j} = \sigma_j$ in (A.21), the statistic becomes

$$\begin{aligned}
\frac{1}{\sqrt{T}}\sum_{j=T+1}^{\lfloor Tr\rfloor}\frac{e_j^W}{\sigma_j} &= \frac{1}{\sqrt{T}}\sum_{j=T+1}^{\lfloor Tr\rfloor}\eta_j - \frac{1}{\sqrt{T}}\sum_{j=T+1}^{\lfloor Tr\rfloor}\frac{(\varphi_{j-1}^W-\varphi)'g_j}{\sigma_j} \\
&= \frac{1}{\sqrt{T}}\sum_{j=T+1}^{\lfloor Tr\rfloor}\eta_j - \frac{1}{T}\sum_{j=T+1}^{\lfloor Tr\rfloor}\left(\frac{g_j'}{(j/T)\sigma_j}\right)\left(\frac{1}{j}\sum_{s=1}^{j-1}\frac{g_sg_s'}{\sigma_s^2}\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{s=1}^{j-1}\frac{g_s\eta_s}{\sigma_s}\right) \\
&\Rightarrow \mathbb{V}_{K+2}(r) - \mathbb{V}_{K+2}(1) - \int_1^r \gamma'\Theta(x)^{-1}(\mathbb{V}_1(x),\ldots,\mathbb{V}_{K+1}(x))'\frac{1}{x\sigma(x)}\mathrm{d}x
\end{aligned}$$

(A.25)

where $\mathbb{V}_k(.)$ denotes the kth element of $\mathbb{V}(.)$. In the last step above deriving (A.25) we have used the weak convergence result (A.22) and an application of the continuous mapping theorem. Defining,

$$\mathbb{W}(r) := \mathbb{V}_{K+2}(r) - \int_0^r \gamma'\Theta(x)^{-1}(\mathbb{V}_1(x), \ldots, \mathbb{V}_{K+1}(x))' \frac{1}{x\sigma(x)} \mathrm{d}x$$

the weak limit in (A.25) can be written as $\mathbb{W}(r) - \mathbb{W}(1)$. Hence, if we can show the $\mathbb{W}(r)$ process is a standard Brownian motion process, then the theorem is proved.

The process $\mathbb{W}(r)$ is a Gaussian process as it is a continuous functional of the Gaussian process $\mathbb{V}(r)$. From the definition of $\mathbb{V}(r)$, it is obvious that its $(K+2)$th element $\mathbb{V}_{K+2}(r)$, which is the first term in the definition of $\mathbb{W}(r)$, is a standard Brownian motion. The process $\mathbb{W}(r)$ is then seen to be a standard Brownian motion if the second term in the definition of $\mathbb{W}(r)$ is $o_p(1)$. However, the second term in the definition of $\mathbb{W}(r)$ is a functional of the Gaussian process $\mathbb{V}(r)$ and is therefore non-degenerate. Therefore, in order to show the Gaussian process $\mathbb{W}(r)$ is a standard Brownian motion, we need to directly analyse its covariance function. If it can be shown that the variance of the second term in the definition of $\mathbb{W}(r)$ cancels exactly with its covariance with the first term, $\mathbb{V}_{K+2}(r)$, then the Gaussian process $\mathbb{W}(r)$ will be a standard Brownian motion process. We now establish that this is indeed the case. This phenomenon is also observed in the proof of KPA's Theorem 1, when they derive the weak limit of their recursive LS residual based CUSUM process.

We need to verify that $\mathrm{cov}(\mathbb{W}(z_1), \mathbb{W}(z_2)) = \min(z_1, z_2)$, which is the required condition for $\mathbb{W}(r)$ to be a standard Brownian motion process. Setting $z_1 < z_2$, without loss of generality, the covariance function is given by

$$\begin{aligned}
&\mathrm{cov}(\mathbb{W}(z_1), \mathbb{W}(z_2)) \\
&\quad = \mathrm{cov}(\mathbb{V}_{K+2}(z_1), \mathbb{V}_{K+2}(z_2)) - \\
&\quad - \mathrm{cov}\left(\mathbb{V}_{K+2}(z_1), \int_0^{z_2} \gamma'\Theta(x)^{-1}(\mathbb{V}_1(x), \ldots, \mathbb{V}_{K+1}(x))' \frac{1}{x\sigma(x)} \mathrm{d}x\right) \\
&\quad - \mathrm{cov}\left(\int_0^{z_1} \gamma'\Theta(x)^{-1}(\mathbb{V}_1(x), \ldots, \mathbb{V}_{K+1}(x))' \frac{1}{x\sigma(x)} \mathrm{d}x, \mathbb{V}_{K+2}(z_2)\right) \\
&\quad + \mathrm{cov}\left(\int_0^{z_1} \frac{\gamma'\Theta(x)^{-1}(\mathbb{V}_1(x), \ldots, \mathbb{V}_{K+1}(x))'}{x\sigma(x)} \mathrm{d}x, \int_0^{z_2} \frac{\gamma'\Theta(x)^{-1}(\mathbb{V}_1(x), \ldots, \mathbb{V}_{K+1}(x))'}{x\sigma(x)} \mathrm{d}x\right) \\
&\quad := A - B - C + D,
\end{aligned}$$

where A, B, C and D are implicitly defined.

For the first term, A, we have that

$$A = \text{cov}(\mathbb{V}_{K+2}(z_1), \mathbb{V}_{K+2}(z_2)) = z_1.$$

For the second term, B, we have that

$$\begin{aligned}
B &= \text{cov}\left(\mathbb{V}_{K+2}(z_1), \int_0^{z_2} \gamma'\Theta(x)^{-1}(\mathbb{V}_1(x), \ldots, \mathbb{V}_{K+1}(x))' \frac{1}{x\sigma(x)} \mathrm{d}x\right) \\
&= E\left(\mathbb{V}_{K+2}(z_1) \int_0^{z_2} \gamma'\Theta(x)^{-1}(\mathbb{V}_1(x), \ldots, \mathbb{V}_{K+1}(x))' \frac{1}{x\sigma(x)} \mathrm{d}x\right) \\
&= E\left(\mathbb{V}_{K+2}(z_1) \left(\int_0^{z_1} \gamma'\Theta(x)^{-1}(\mathbb{V}_1(x), \ldots, \mathbb{V}_{K+1}(x))' \frac{1}{x\sigma(x)} \mathrm{d}x + \right.\right. \\
&\quad \left.\left. \int_{z_1}^{z_2} \gamma'\Theta(x)^{-1}(\mathbb{V}_1(x), \ldots, \mathbb{V}_{K+1}(x))' \frac{1}{x\sigma(x)} \mathrm{d}x\right)\right) \\
&= \int_0^{z_1} (\gamma'\Theta(x)^{-1}\gamma) \left(\int_0^x \frac{1}{\sigma(y)} \mathrm{d}y\right) \frac{1}{x\sigma(x)} \mathrm{d}x + \int_{z_1}^{z_2} (\gamma'\Theta(x)^{-1}\gamma) \left(\int_0^{z_1} \frac{1}{\sigma(y)} \mathrm{d}y\right) \frac{1}{x\sigma(x)} \mathrm{d}x \\
&= \int_0^{z_1} (\gamma'\Theta(x)^{-1}\gamma) \left(\int_0^x \frac{1}{\sigma(y)} \mathrm{d}y\right) \frac{1}{x\sigma(x)} \mathrm{d}x + \left(\int_0^{z_1} \frac{1}{\sigma(y)} \mathrm{d}y\right) \int_{z_1}^{z_2} (\gamma'\Theta(x)^{-1}\gamma) \frac{1}{x\sigma(x)} \mathrm{d}x.
\end{aligned}$$

In the above third step, we split the integral at value z_1, such that the range of the two resulting integrals has a fixed relative magnitude with z_1. Then in the fourth step, we exchange the order of integration and expectation, and use the definition of the covariance matrix of $\mathbb{V}$ in (A.24). This proof strategy is repeatedly used below for the analysis of term C and D without being explained.

For the third term, C, we have that

$$\begin{aligned}
C &= \text{cov}\left(\int_0^{z_1} \gamma'\Theta(x)^{-1}(\mathbb{V}_1(x), \ldots, \mathbb{V}_{K+1}(x))' \frac{1}{x\sigma(x)} \mathrm{d}x, \mathbb{V}_{K+2}(z_2)\right) \\
&= E\left(\mathbb{V}_{K+2}(z_2) \int_0^{z_1} \gamma'\Theta(x)^{-1}(\mathbb{V}_1(x), \ldots, \mathbb{V}_{K+1}(x))' \frac{1}{x\sigma(x)} \mathrm{d}x\right) \\
&= \int_0^{z_1} (\gamma'\Theta(x)^{-1}\gamma) \left(\int_0^x \frac{1}{\sigma(y)} \mathrm{d}y\right) \frac{1}{x\sigma(x)} \mathrm{d}x \\
&= \int_0^{z_1} (\gamma'\Theta(x)^{-1}\gamma) \left(\int_0^x \frac{1}{\sigma(y)} \mathrm{d}y\right) \frac{1}{x\sigma(x)} \mathrm{d}x
\end{aligned}$$

Finally, for the fourth term, D, we have that

$$
\begin{aligned}
D &= \operatorname{cov}\left(\int_0^{z_1} \gamma'\Theta(x)^{-1}(\mathbb{V}_1(x),\ldots,\mathbb{V}_{K+1}(x))'\frac{1}{x\sigma(x)}\mathrm{d}x, \int_0^{z_2} \gamma'\Theta(x)^{-1}(\mathbb{V}_1(x),\ldots,\mathbb{V}_{K+1}(x))'\frac{1}{x\sigma(x)}\mathrm{d}x\right) \\
&= E\left(\left(\int_0^{z_1} \gamma'\Theta(x)^{-1}(\mathbb{V}_1(x),\ldots,\mathbb{V}_{K+1}(x))'\frac{1}{x\sigma(x)}\mathrm{d}x\right)\left(\int_0^{z_2} \gamma'\Theta(x)^{-1}(\mathbb{V}_1(x),\ldots,\mathbb{V}_{K+1}(x))'\frac{1}{x\sigma(x)}\mathrm{d}x\right)\right) \\
&= E\int_0^{z_2}\int_0^{z_1} \gamma'\Theta(x)^{-1}(\mathbb{V}_1(x),\ldots,\mathbb{V}_{K+1}(x))'\gamma'\Theta(x)^{-1}(\mathbb{V}_1(x),\ldots,\mathbb{V}_{K+1}(x))'\frac{1}{x\sigma(x)}\frac{1}{y\sigma(y)}\mathrm{d}y\mathrm{d}x \\
&= E\int_0^{z_2}\int_0^{z_1} \gamma'\Theta(x)^{-1}(\mathbb{V}_1(x),\ldots,\mathbb{V}_{K+1}(x))'(\mathbb{V}_1(y),\ldots,\mathbb{V}_{K+1}(y))\Theta(y)^{-1}\gamma\frac{1}{x\sigma(x)}\frac{1}{y\sigma(y)}\mathrm{d}y\mathrm{d}x \\
&= \int_0^{z_2}\int_0^{z_1}(\gamma'\Theta(x)^{-1}\Theta(\min(x,y))\Theta(y)^{-1}\gamma)(\min(x,y))\frac{1}{x\sigma(x)}\frac{1}{y\sigma(y)}\mathrm{d}y\mathrm{d}x \\
&= \int_0^{z_1}\int_0^{z_1}(\gamma'\Theta(x)^{-1}\Theta(\min(x,y))\Theta(y)^{-1}\gamma)(\min(x,y))\frac{1}{x\sigma(x)}\frac{1}{y\sigma(y)}\mathrm{d}y\mathrm{d}x \\
&\quad + \int_{z_1}^{z_2}\int_0^{z_1}(\gamma'\Theta(x)^{-1}\Theta(\min(x,y))\Theta(y)^{-1}\gamma)(\min(x,y))\frac{1}{x\sigma(x)}\frac{1}{y\sigma(y)}\mathrm{d}y\mathrm{d}x \\
&= 2\int_0^{z_1}\int_0^{x}(\gamma'\Theta(x)^{-1}\Theta(\min(x,y))\Theta(y)^{-1}\gamma)(\min(x,y))\frac{1}{x\sigma(x)}\frac{1}{y\sigma(y)}\mathrm{d}y\mathrm{d}x \\
&\quad + \int_{z_1}^{z_2}\int_0^{z_1}(\gamma'\Theta(x)^{-1}\Theta(\min(x,y))\Theta(y)^{-1}\gamma)(\min(x,y))\frac{1}{x\sigma(x)}\frac{1}{y\sigma(y)}\mathrm{d}y\mathrm{d}x \\
&= 2\int_0^{z_1}\int_0^{x}(\gamma'\Theta(x)^{-1}\Theta(\min(x,y))\Theta(y)^{-1}\gamma)(y)\frac{1}{x\sigma(x)}\frac{1}{y\sigma(y)}\mathrm{d}y\mathrm{d}x \\
&\quad + \int_{z_1}^{z_2}\int_0^{z_1}(\gamma'\Theta(x)^{-1}\Theta(\min(x,y))\Theta(y)^{-1}\gamma)(y)\frac{1}{x\sigma(x)}\frac{1}{y\sigma(y)}\mathrm{d}y\mathrm{d}x \\
&= 2\int_0^{z_1}\int_0^{x}(\gamma'\Theta(x)^{-1}\gamma)\frac{1}{x\sigma(x)\sigma(y)}\mathrm{d}y\mathrm{d}x + \int_{z_1}^{z_2}\int_0^{z_1}(\gamma'\Theta(x)^{-1}\gamma)\frac{1}{x\sigma(x)\sigma(y)}\mathrm{d}y\mathrm{d}x \\
&= 2\int_0^{z_1}(\gamma'\Theta(x)^{-1}\gamma)\left(\int_0^{x}\frac{1}{\sigma(y)}\mathrm{d}y\right)\frac{1}{x\sigma(x)}\mathrm{d}x + \left(\int_0^{z_1}\frac{1}{\sigma(y)}\mathrm{d}y\right)\int_{z_1}^{z_2}(\gamma'\Theta(x)^{-1}\gamma)\frac{1}{x\sigma(x)}\mathrm{d}x,
\end{aligned}
$$

where in the above seventh step, we use the symmetry of double integration with respect to x and y, to rewrite the inner integral with respect to y as 2 times the integral with a variable upper limit x, which is the variable of integration of the out layer integral.

It can be seen from the results above that $-B-C+D=0$. Consequently, $\operatorname{cov}(\mathbb{W}(z_1),\mathbb{W}(z_2)) = A-B-C+D = A = z_1$, which completes the first part of the proof. Observe from the foregoing derivations, that if recursive LS (rather than WLS) residuals were used in our context, i.e. if we use the CUSUM statistic $\frac{1}{\sqrt{T}}\sum_{j=T+1}^{\lfloor Tr\rfloor}\frac{e_j}{\sigma_j}$, then these cancellations would not happen and $\mathbb{W}(r)$ would not be a standard Brownian motion.

We now turn to the second part of the proof. We will show that

$$
\left|\frac{1}{\sqrt{T}}\sum_{j=T+1}^{\lfloor Tr\rfloor}\frac{\sigma_j\eta_j}{\tilde{\sigma}_{j,N,j}} - \frac{1}{\sqrt{T}}\sum_{j=T+1}^{\lfloor Tr\rfloor}\eta_j\right| = o_p(1), \tag{A.26}
$$

and

$$\left| \frac{1}{\sqrt{T}} \sum_{j=T+1}^{\lfloor Tr \rfloor} \frac{(\hat{\varphi}_{j-1}^{W} - \varphi)' g_j}{\tilde{\sigma}_{j,N,j}} - \frac{1}{\sqrt{T}} \sum_{j=T+1}^{\lfloor Tr \rfloor} \frac{(\varphi_{j-1}^{W} - \varphi)' g_j}{\sigma_j} \right| = o_p(1), \tag{A.27}$$

and the claimed result of the theorem follows.

(A.26) can be proved in the same way as in the proof for Theorem 1 of AHLTZ, by noting the results of Lemma A.4 and Lemma A.5.

For (A.27), notice that

$$\begin{aligned} &\left| \frac{1}{\sqrt{T}} \sum_{j=T+1}^{\lfloor Tr \rfloor} \frac{(\hat{\varphi}_{j-1}^{W} - \varphi)' g_j}{\tilde{\sigma}_{j,N,j}} - \frac{1}{\sqrt{T}} \sum_{j=T+1}^{\lfloor Tr \rfloor} \frac{(\varphi_{j-1}^{W} - \varphi)' g_j}{\sigma_j} \right| \\ \leqslant \; & \frac{1}{\sqrt{T}} \sum_{j=T+1}^{\lfloor Tr \rfloor} |(\varphi_{j-1}^{W} - \varphi)' g_j| \left| \frac{1}{\tilde{\sigma}_{j,N,j}} - \frac{1}{\sigma_j} \right| + \left| \frac{1}{\sqrt{T}} \sum_{j=T+1}^{\lfloor Tr \rfloor} \frac{(\hat{\varphi}_{j-1}^{W} - \varphi_{j-1}^{W})' g_j}{\tilde{\sigma}_{j,N,j}} \right|. \end{aligned}$$

By Lemma A.4 and Lemma A.6, the above second term is $o_p(1)$ uniformly for $r \in (0, 1]$. By Lemma A.7, the first term is $o_p(1)$, and the proof for (A.27) is complete. □

Proof of Theorem 2. Under H_1, and when $\lfloor \tau T \rfloor + 1 \leqslant t \leqslant \lfloor \lambda T \rfloor$,

$$\begin{aligned} SWV_T^t &= \frac{1}{\sqrt{T}} \sum_{j=T+1}^{t} \frac{\hat{e}_j^W}{\tilde{\sigma}_{j,N,j}} \\ &= \frac{1}{\sqrt{T}} \left(\sum_{j=T+1}^{\lfloor \tau T \rfloor} + \sum_{j=\lfloor \tau T \rfloor + 1}^{\lfloor \tau T \rfloor + \xi_T + 1} + \sum_{j=\lfloor \tau T \rfloor + \xi_T + 1}^{t} \right) \frac{\Delta y_j - (\hat{\varphi}_{j-1}^{W})' g_j}{\tilde{\sigma}_{j,N,j}} \\ &=: A_T + B_{\xi_T} + C_t \end{aligned}$$

where A_T, B_{ξ_T} and C_t are implicitly defined.

Consider first C_t. When $\lfloor \tau T \rfloor + \xi_T + 1 \leqslant t \leqslant \lfloor \lambda T \rfloor$, we have that

$$\Delta y_j = \delta_T u_{j-1} + v_j.$$

Substituting this into the expression for C_t we have that

$$\begin{aligned} C_t &= \delta_T \frac{1}{\sqrt{T}} \sum_{j=\lfloor \tau T \rfloor + \xi_T + 1}^{t} \left(\frac{u_{j-1}}{\tilde{\sigma}_{j,N,j}} + \frac{v_j}{\tilde{\sigma}_{j,N,j}} - \frac{(\hat{\varphi}_{j-1}^{W})' g_j}{\tilde{\sigma}_{j,N,j}} \right) \\ &=: C_{t1} + C_{t2} - C_{t3} \end{aligned}$$

where C_{t1}, C_{t2} and C_{t3} are the summation of three terms in the brackets implicitly defined.

When $\lfloor \tau T \rfloor + \xi_T + 1 \leqslant t \leqslant \lfloor \lambda T \rfloor$, we will first show that $|C_{t2}/C_{t1}| = o_p(1)$ and the order of C_{t3} is no larger than that of C_{t1}, so that the rate of $C_{t1} + C_{t2} + C_{t3}$ is determined by the rate of divergence of C_{t1}; then we derive a lower bound for the divergence rate of C_{t1}.

We will use Lemma A.9 in the following analysis. Notice that given $0 \leqslant d \leqslant 2/3$, $T^d \wedge N = T^d$. Consider first C_{t1}. This satisfies

$$
\begin{aligned}
& |C_{t1}| \\
= \;& \delta_T \frac{1}{\sqrt{T}} (T^d N)^{1/2} \left| \sum_{j=\lfloor \tau T \rfloor + \xi_T + 1}^{t} \frac{T^{-1/2} \psi_T^{-(j-1-\lfloor \tau T \rfloor)} u_{j-1}}{(T^{d-1} N)^{1/2} \psi_T^{-(j-1-\lfloor \tau T \rfloor)} \tilde{\sigma}_{j,N,j}} \right| \\
\leqslant \;& \max_{\lfloor \tau T \rfloor + \xi_T + 1 \leqslant j \leqslant t} \left| \frac{\delta_T (T^d N)^{1/2}}{(T^{d-1} N)^{1/2} \psi_T^{-(j-1-\lfloor \tau T \rfloor)} \tilde{\sigma}_{j,N,j}} \right| \frac{1}{\sqrt{T}} \sum_{j=\lfloor \tau T \rfloor + \xi_T + 1}^{t} |T^{-1/2} \psi_T^{-(j-1-\lfloor \tau T \rfloor)} u_{j-1}| \\
= \;& \frac{\delta_T (T^d N)^{1/2}}{\min_{\lfloor \tau T \rfloor + \xi_T + 1 \leqslant j \leqslant t} |(T^{d-1} N)^{1/2} \psi_T^{-(j-1-\lfloor \tau T \rfloor)} \tilde{\sigma}_{j,N,j}|} \frac{1}{\sqrt{T}} \sum_{j=\lfloor \tau T \rfloor + \xi_T + 1}^{t} |T^{-1/2} \psi_T^{-(j-1-\lfloor \tau T \rfloor)} u_{j-1}| \\
= \;& \frac{\delta_T (T^d N)^{1/2}}{\min_{\lfloor \tau T \rfloor + \xi_T + 1 \leqslant j \leqslant t} |(T^{d-1} N)^{1/2} \psi_T^{-(j-1-\lfloor \tau T \rfloor)} \tilde{\sigma}_{j,N,j}|} \frac{1}{\sqrt{T}} (t - 1 - \lfloor \tau T \rfloor - \xi_T) |T^{-1/2} u_{\lfloor \tau T \rfloor}| (1 + o_p(1)).
\end{aligned}
$$

Using Lemma A.9 we have that $\min_{\lfloor \tau T \rfloor + \xi_T + 1 \leqslant j \leqslant t} |(T^{d-1} N)^{1/2} \psi^{-(j-1-\lfloor \tau T \rfloor)} \tilde{\sigma}_{j,N,j}|$ is, with probability 1, strictly positive. Moreover, $|T^{-1/2} u_{\lfloor \tau T \rfloor}| = O_p(1)$, and so for any $t \geqslant \lfloor \tau^* T \rfloor$ with $\tau^* > \tau$, $(t - 1 - \lfloor \tau T \rfloor - \xi_T) = O(T)$, and it follows that C_{t1} is $O_p(T^{1/2-d}(T^d N)^{1/2}) = O_p\left(T^{\frac{1}{2}(1-d)} N^{\frac{1}{2}}\right)$. Using the same argument, we can show that C_{t2} is stochastically dominated by C_{t1}. Next consider C_{t3}. Notice that the g_j vector contains the Δy_{j-k}'s for $k = 1, 2, \ldots, p$, together with the covariates. In the explosive regime, $\Delta y_{j-k} = \delta_T u_{j-k-1} + v_{j-k}$, which is explosive, while the covariate terms are, by definition, non-explosive. Therefore, C_{t3} can be studied in the same way as C_{t1} and it can be shown that it is no larger than C_{t1}, and it also cannot be the same as C_{t1} (so there is no possibility of cancellation between the two terms), due to its dependence on u_j terms up to $j = t - 2$, while C_{t1} is defined by explosive u_j's up to $j = t - 1$. In summary, the order of C_t is determined by that of C_{t1}. We next derive a lower bound for the divergence rate of C_{t1}. To that end, observe that $|C_{t1}|$ also satisfies

$$
\begin{aligned}
&|C_{t1}| \\
\geqslant\ & \min_{\lfloor\tau T\rfloor+\xi_T+1\leqslant j\leqslant t}\left|\frac{\delta_T(T^dN)^{1/2}}{(T^{d-1}N)^{1/2}\psi_T^{-(j-1-\lfloor\tau T\rfloor)}\tilde{\sigma}_{j,N,j}}\right|\frac{1}{\sqrt{T}}\sum_{j=\lfloor\tau T\rfloor+\xi_T+1}^{t}|T^{-1/2}\psi_T^{-(j-1-\lfloor\tau T\rfloor)}u_{j-1}| \\
=\ & \frac{\delta_T(T^dN)^{1/2}}{\max_{\lfloor\tau T\rfloor+\xi_T+1\leqslant j\leqslant t}|(T^{d-1}N)^{1/2}\psi_T^{-(j-1-\lfloor\tau T\rfloor)}\tilde{\sigma}_{j,N,j}|}\frac{1}{\sqrt{T}}\sum_{j=\lfloor\tau T\rfloor+\xi_T+1}^{t}|T^{-1/2}\psi_T^{-(j-1-\lfloor\tau T\rfloor)}u_{j-1}| \\
=\ & \frac{\delta_T(T^dN)^{1/2}}{\max_{\lfloor\tau T\rfloor+\xi_T+1\leqslant j\leqslant t}|(T^{d-1}N)^{1/2}\psi_T^{-(j-1-\lfloor\tau T\rfloor)}\tilde{\sigma}_{j,N,j}|}\frac{1}{\sqrt{T}}(t-1-\lfloor\tau T\rfloor-\xi_T)|T^{-1/2}u_{\lfloor\tau T\rfloor}|(1+o_p(1)).
\end{aligned}
$$

Again using Lemma A.9, we have that $\max_{\lfloor\tau T\rfloor+\xi_T+1\leqslant j\leqslant t}|(T^{d-1}N)^{1/2}\psi_T^{-(j-1-\lfloor\tau T\rfloor)}\tilde{\sigma}_{j,N,j}| = O_p(1)$. Moreover, $|T^{-1/2}u_{\lfloor\tau T\rfloor}| = O_p(1)$ and is non-degenerate, and so for $t\geqslant\lfloor\tau^*T\rfloor$, with $\tau^*>\tau$, $|C_{t1}/(c_t\sqrt{t})| = O_p(\delta_T(T^dN)^{1/2})$. Notice that $\delta_T(T^dN)^{1/2} = O(T^{-d/2}N^{1/2})\to\infty$ because $d\leqslant 2/3$. Observing that $\max_{\lfloor\tau T\rfloor+\xi_T+1\leqslant j\leqslant t}|(T^{d-1}N)^{1/2}\psi_T^{-(j-1-\lfloor\tau T\rfloor)}\tilde{\sigma}_{j,N,j}|$ appears in the denominator, its stochastic upper bound order of $O_p(1)$ gives the lower bound of the divergence rate for C_{t1}.

Next, we observe that $A_T = O_p(1)$, regardless of the value of t (i.e. it has the same order in probability throughout the monitoring period). B_{ξ_T} represents the sum of ξ_T+1 terms immediately after the structural break to an explosive regime; its order also does not depend on t. Notice that since it cannot cancel exactly with C_t, which has a changing end point as the monitoring process goes on, the derived divergence rate lower bound we have derived for C_t also serves as a divergence rate lower bound for the monitoring statistic, regardless of the specific order of B_{ξ_T}. The monitoring statistic will always diverge relative to the boundary function, $c_t\sqrt{t}$, with at least a rate $O(T^{-d/2}N^{1/2})$ due to the C_t term, and so the stated result follows. □

Proof of Theorem A.1. Observe that

$$
\begin{aligned}
\frac{1}{\sqrt{T}}\sum_{j=T+1}^{\lfloor Tr\rfloor}\frac{e_j}{\sigma_j} &= \frac{1}{\sqrt{T}}\sum_{j=T+1}^{\lfloor Tr\rfloor}\frac{\sigma_j\eta_j-(\hat{\phi}_{j-1}-\phi)'z_j}{\sigma_j} \\
&= \frac{1}{\sqrt{T}}\sum_{j=T+1}^{\lfloor Tr\rfloor}\eta_j-\frac{1}{\sqrt{T}}\sum_{j=T+1}^{\lfloor Tr\rfloor}\frac{(\hat{\phi}_{j-1}-\phi)'z_j}{\sigma_j}.
\end{aligned}
$$

Using the same proof strategy as that of Theorem 1, we can again show that the first term above weakly converges to a standard Brownian motion, while the second term weakly converges to an integral of a Gaussian process as in (A.25). However, the proof is simpler because each z_j in the present setting now has zero mean, and so the second term in the equation above can be shown to be of $o_p(1)$, from which the stated result follows straightforwardly. □

Proof of Theorem A.2. Using the result in Theorem A.1, the stated result follows if we can show that the errors induced by the nonparametric estimation of the variance function are asymptotically negligible. This can be done along exactly the same lines as in the second part of the proof of Theorem 1, and we therefore omit the details to avoid repetition. □

Proof of Theorem A.3. Again, the proof will follow along the same lines as the proof of Theorem 2, using the observation that no aspect of that proof requires the covariates to have strictly non-zero means. We therefore omit the proof to avoid repetition. □

Additional References

Billingsley, P. (1999). Convergence of Probability Measures. Wiley Series in Probability and Statistics.

Chang, Y. and Park, J.Y. (2002). On the asymptotics of ADF tests for unit roots. Econometric Reviews, 21(4), 431-447.

Davidson, J. (2021). Stochastic Limit Theory: An Introduction for Econometricians, 2nd edition. Oxford University Press.

Hall, P. and Heyde, C.C. (1980). Martingale Limit Theory and its Application. Academic press.

A.3 Empirical Application

In this section we investigate the performance of our proposed monitoring procedures had they been applied ahead of Black Monday in 1987, and the dotcom bubble episode of the early 1990s. To do so we use the monthly dataset of Welch and Goyal (2008) which can be obtained from http://www.hec.unil.ch/agoyal/ as well as the 10 Year US Treasury Constant Maturity Rate which can be obtained from https://fred.stlouisfed.org/series/GS10.

Following PSY, the series tested for bubbles will be the price-dividend ratio (Index/D12) plotted in Figure A.1. Applying the $GSADF$ test of PSY to the sample of data used for our empirical analysis (October 1968 - December 1997) using the authors' recommended settings yields a test statistic of 2.873, which is in excess of the 1% critical value of 2.582, which gives strong evidence in agreement with the findings of PSY that one or more bubbles are present during this period. The candidate covariates are earnings (E12), the book-to-market ratio (b/m), the treasury-bill rate (tbl), corporate bond returns on AAA and BAA rated bonds (AAA and BAA), the 10 Year US Treasury Constant Maturity Rate (GS10) long term yield (lty), net equity expansion (ntis), the risk free rate (rfree), inflation (infl), long term rate of returns (ltr), long term corporate bond returns (corpr), stock variance (svar), the cross sectional premium (csp), the dividend payout ratio (de:=D12/E12), the earnings-price ratio (ep:=E12/Index), the default yield spread (dfy:=BAA-AAA), the term spread (tms:=lty-tbl) and the default return spread (dfr:=corpr-ltr).

We begin by examining how a monitoring exercise that began in January 1987, ahead of Black Monday in October 1987, would have played out, examining the performance of the CUSUM^V, CUSUM^{V*} and CUSUM^{WMV} monitoring procedures. For simplicity, and to help determine which covariates are individually useful, we apply the CUSUM^{WMV} procedure using only a single covariate at a time. Our training sample begins in October 1968 such that its length is equal to $T = 219$ as in the Monte Carlo simulations in Section 5. We use the same bandwidth selection rule as in Section 5 and, again, use the BIC procedures outlined in section 5.1 to select p and q, and whether or not to include the covariates, in the null regression model, (5), again setting the maximum permitted values of p and q to 4 and 2, respectively. We set the value of $b_\alpha = 0.0883$ such that the monitoring procedures would have an empirical FPR of 0.10 after 1 year if the price-dividend data were a pure unit root process driven by NIID innovations under the null.

Before applying the CUSUM^{WMV} procedure we must first ensure that any covariates used do not contain a unit root. We, therefore, pre-test the candidate covariates for a unit root using the training sample observations that would have been available at the commencement of the monitoring procedure. We apply the (heteroskedasticity-robust) wild bootstrap ADF unit root test of Cavaliere and Taylor (2009) at the 5% level allowing for an intercept using the authors' recommended settings where the number of lagged differences in the ADF regression is determined using the MAIC of Ng and Perron (2001) with the maximum number of lags k given by $k_{\max} = \lfloor 12(T/100)^{0.25} \rfloor$. Critical values are obtained using $B = 499$ bootstrap replications where the bootstrap data is generated using the recolouring scheme outlined in Cavaliere and Taylor (2009) using the same value of k selected by the MAIC for the test statistic of interest, with the value of k used to construct the bootstrap test statistics again determined by the MAIC. From this we found the variables ltr, corpr and dfr to be I(0) so there variables are utilised in levels, whereas we found rfree, infl, svar, tms, E12, b/m, tbl, GS10, AAA, BAA, lty, ntis, csp, de, ep and dfy to be I(1) so these variables are utilised in first differences.

At the commencement of the monitoring procedure applying the BIC to (5) indicates that the covariates that are individually relevant for monitoring the price-dividend series are Δ(b/m), Δ(tbl), Δ(GS10),Δ(AAA), Δ(BAA), Δ(lty), ltr, corpr, Δ(csp) and Δ(ep) and so we only report results for the use of these covariates in the CUSUM^{WMV} procedure. For the CUSUM^{V*} procedure the BIC selects a lag length of $p = 0$ and so that this procedure is identical to the CUSUM^{V} procedure, we therefore report results only for the latter procedure. Figures A.2-A.3 report plots of the individual test statistics underlying the monitoring procedures, as well as the boundary function $c_t\sqrt{t}$, with a rejection of the no-bubble null indicated by any test statistic exceeding this boundary function. The vertical dashed lines are used to indicate the first date each monitoring procedure rejects the null of no bubble. The plots of these test statistics shows that the Black Monday bubble episode was rather short lived, with only a small window of opportunity for detection before the collapse of the price-divided ratio. In spite of this we see that the CUSUM^{WMV} procedure would have detected this bubble in July 1987 when utilising any of Δ(GS10), Δ(AAA), Δ(lty), ltr, corpr or Δ(ep) as a covariate, which is earlier than the first rejection in August 1987 displayed by CUSUM^{V}. For the other covariates the CUSUM^{WMV} procedure first rejects at the same time as CUSUM^{V}, excepting Δ(csp) where the CUSUM^{WMV} procedure marginally fails to reject in August 1987. We also extended the analysis to allow for multiple covariates, letting the BIC select from any combination of the covariates that were found to be individually relevant. In this case the BIC suggested including both $\Delta(b/m)$ and

$\Delta(ep)$, with the resulting procedure rejecting slightly later than when using $\Delta(ep)$ alone, highlighting the fact that including additional covariates may not always be beneficial.

We also examined how a monitoring procedure that began in January 1994, ahead of the dotcom bubble, would have played out. The monitoring procedures were performed exactly as for the Black Monday exercise except that the training sample of data were of length $T = 72$, running from January 1988 to December 1993 so as to avoid the abrupt collapse in the price-dividend ratio witnessed at the end of 1987 following Black Monday; cf. item 3 in Section 5.3. This necessitated setting $b_\alpha = 0.2672$ to retain an FPR of 0.10 after 1 year, again assuming the price-dividend data were a purely unit root process driven by NIID innovations under the null. Once again, the BIC selected $p = 0$ for the CUSUM^{V*} procedure so we report results only for CUSUM^{V}. For the CUSUM^{WMV} procedure we utilise the covariates that proved to be useful during the Black Monday bubble episode, namely Δ(GS10) Δ(AAA), Δ(lty), ltr, corpr or Δ(ep). Figure A.4 again reports plots of the individual test statistics underlying the monitoring procedures, as well as the boundary function $c_t\sqrt{t}$. The CUSUM^{V} procedure which utilises no covariate augmentation first rejects the null of no bubble in January 1996, whereas the CUSUM^{WMV} procedure rejects earlier when using any of the six candidate covariates, with a first rejection in October 1994 when using corpr, March 1995 when using Δ(ep), July 1995 when using Δ(AAA), Δ(GS10) or ltr and September 1995 when using Δ(lty). We also extended the analysis to allow for multiple covariates, letting the BIC select from any combination of the covariates that were found to be individually relevant. In this case the BIC suggested using only $\Delta(ep)$ which was previously shown to lead to the second earliest rejection among all individual candidate covariates.

As an additional robustness check, we also performed unit root tests on all of the covariates employed in the CUSUM^{WMV} procedure across the entire sample range of data used for our empirical analysis (October 1968 - December 1997) and obtained the same conclusions as when these unit root tests were performed on the initial training sample, suggesting no change in persistence occurred for any of the employed covariates during the monitoring periods. We also applied the $GSADF$ test of PSY to each of these covariates over the full sample period and found no evidence of bubbles, rendering it unlikely that any co-explosive behaviour was present between the covariates and the series of interest.

Additional References

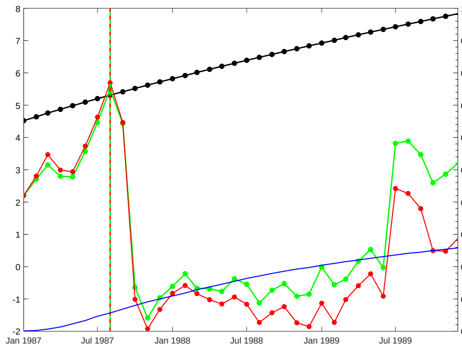
Cavaliere, G. and Taylor, A.M.R. (2009). Bootstrap M unit root tests. Econometric Reviews 28, 393–421.

Ng, S. and Perron, P. (2011). Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power. Econometrica 69, 1519-1554.

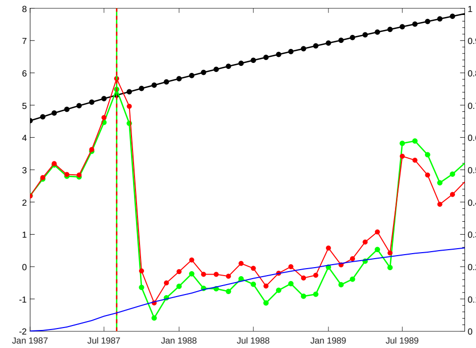
Figure A.1: Price Dividend Ratio - 1968-2000

Figure A.2: Test Statistics vs Critical Value - Black Monday

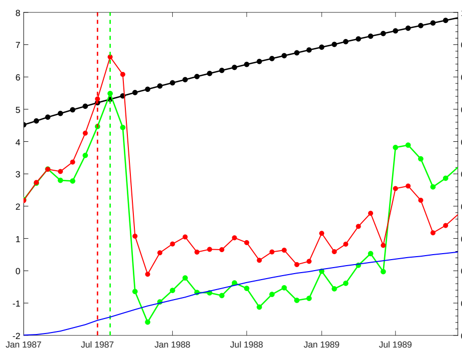
(a) Δ(b/m)



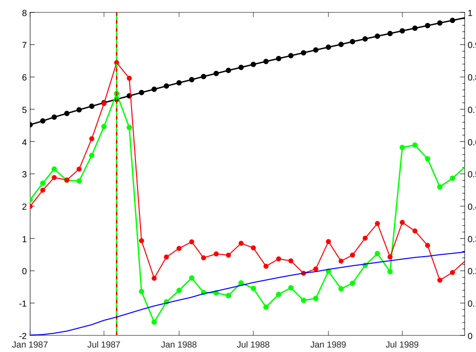
(b) Δ(tbl)



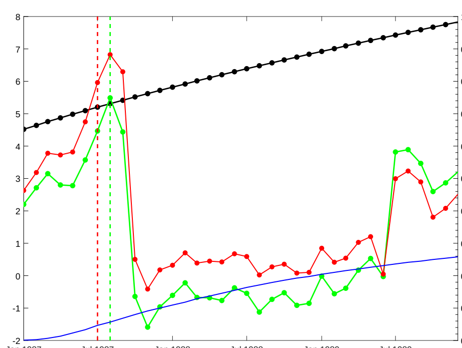
(c) Δ(AAA)



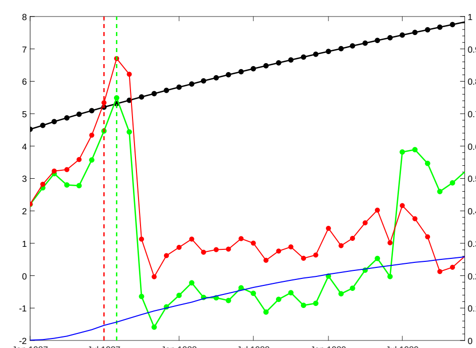
(d) Δ(BAA)



(e) Δ(lty)



(f) ΔGS10

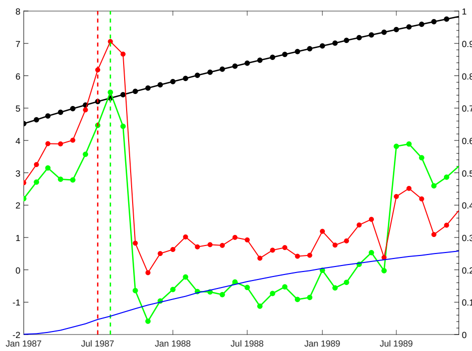


LEFT AXIS: Critical Value:——, $CUSUM^{V}$, $CUSUM^{V*}$:——,$CUSUM^{WMV}$:——

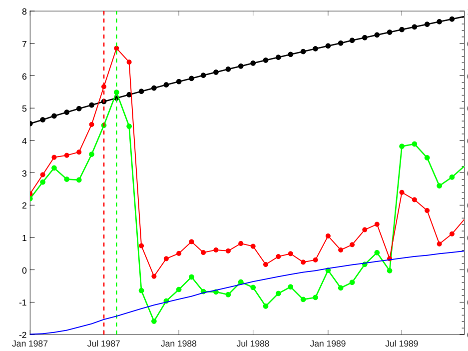
RIGHT AXIS: False Positive Rate:——

Figure A.3: Test Statistics vs Critical Value - Black Monday

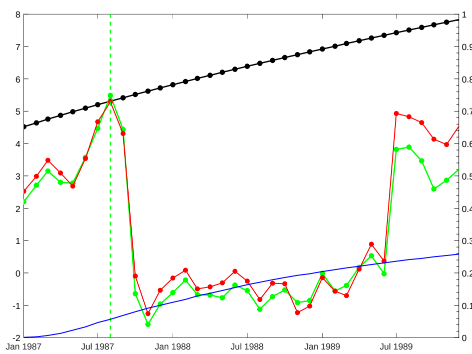
(a) ltr



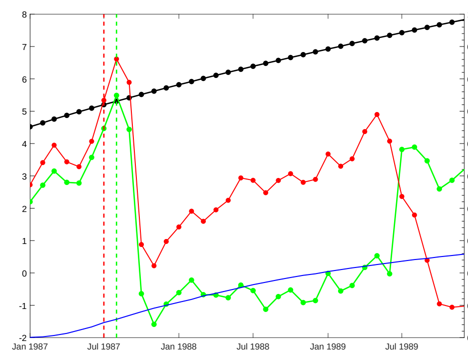
(b) corpr



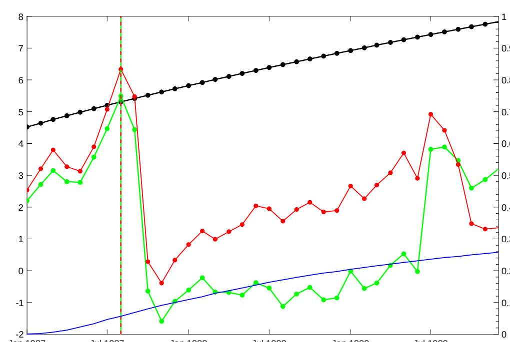
(c) Δ(csp)



(d) Δ(ep)



(e) Δ(b/m) & Δ(ep)

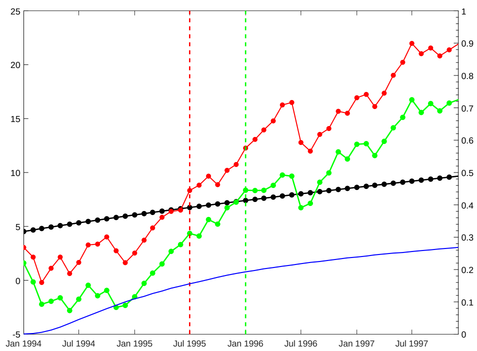


LEFT AXIS: Critical Value:———, CUSUM^{V}, CUSUM^{V*}:———,CUSUM^{WMV}:———

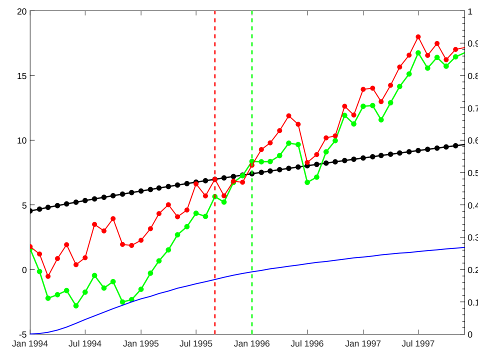
RIGHT AXIS: False Positive Rate:———

Figure A.4: Test Statistics vs Critical Value - dotcom

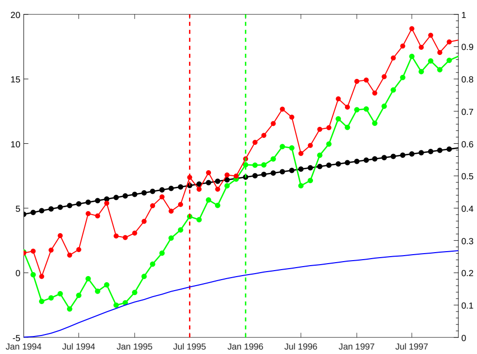
(a) Δ(AAA)



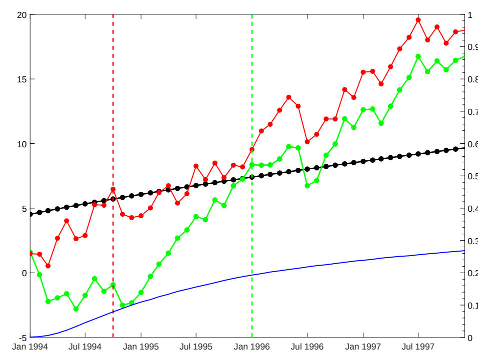
(b) Δ(lty)



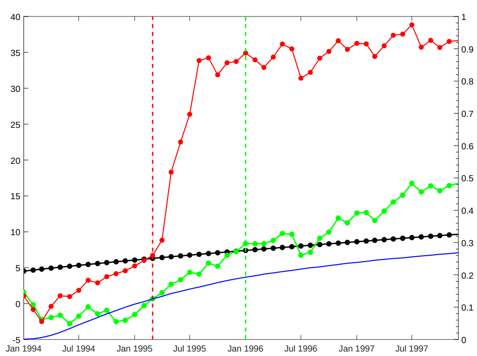
(c) ltr



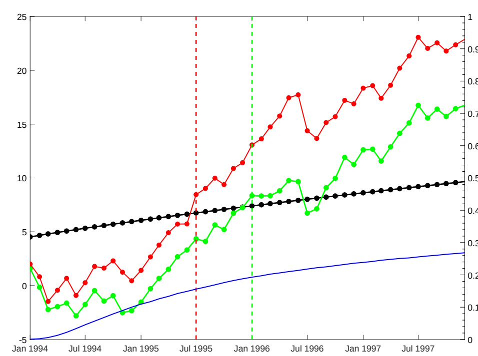
(d) corpr



(e) Δ(ep)



(f) Δ(GS10)



LEFT AXIS: Critical Value:———, CUSUM^{V}, CUSUM^{V*}:———,CUSUM^{WMV}:———

RIGHT AXIS: False Positive Rate:———

A.4 Additional Monte Carlo Simulation Results

This section reports Monte Carlo simulations results additional to those reported in the main paper. Similar to the main paper all results are reported for the DGP in (1)-(2), and under the alternative we set $\lfloor \tau T \rfloor = 220$ with $\delta = 0.005$.

A.4.1 Additional Simulations - Further Parameter Constellations

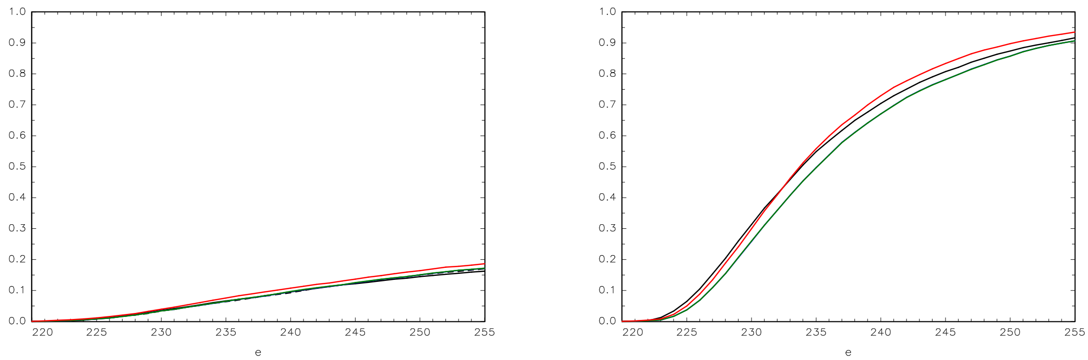
Figure A.5 reports the FPR of the procedures under the null and TPR under the alternative for the CSS type DGP for v_t and x_t given by (23)-(24) with $\rho = \sigma_{12} = \alpha_1 = 0$ and $\beta = 0.5$. The reported results are qualitatively similar to those reported in Figure 2 for the case when $\beta = 0.8$ reported in the main paper.

Figure A.6 reports the FPR and TPR of the procedures for the CSS type DGP for v_t and x_t given by (23)-(24) with $\sigma_{12} = 0.4$, $\alpha_1 = 0.2$ and $\beta = \rho = 0.0$ such that the covariate x_t is irrelevant but the innovations v_t are serially correlated. The FPR and TPR profile of both the CUSUM^{WMV} and CUSUM^{V*} monitoring procedures are practically identical to each other, identified by the green and red lines being almost indistinguishable from one another. This is due to the fact that the BIC deems the candidate covariate to be irrelevant, so that the CUSUM^{WMV} procedure reduces to the CUSUM^{V*} procedure, in the vast majority of replications. Both the CUSUM and CUSUM^{V} procedures exhibit substantial FPR distortions in this scenario due to the unmodelled serial correlation present in v_t.

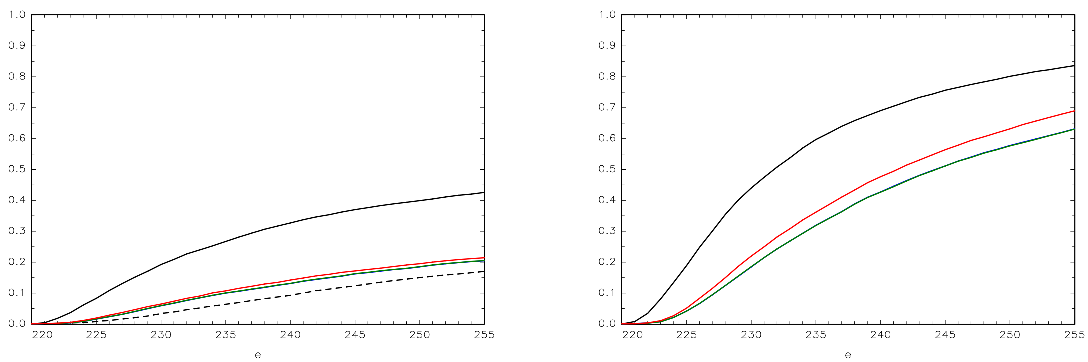
Figures A.7-A.12 present the FPR of the procedures under the null and TPR under the alternative for the CSS type DGP for v_t and x_t given by (23)-(24) when $\alpha_1 = 0.2$ and $\sigma_{12} = 0.4$ for the combinations of β and ρ considered by CSS not reported in Figures 3-4 in the main paper. Once again the results are all qualitatively similar to those reported in the main paper.

Figure A.5: $\beta = 0.5$, $\rho = \sigma_{12} = \alpha_1 = 0$ - Left Panel=FPR, Right Panel =TPR. ($\varrho^2 = 0.800$)

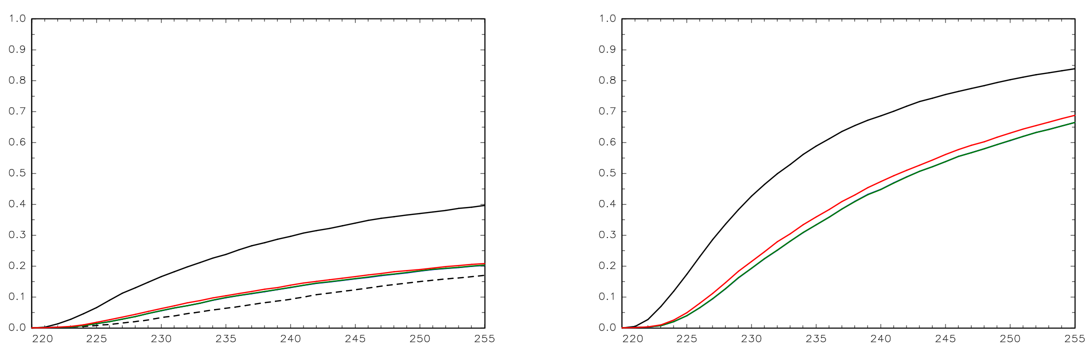
(a) Homoskedastic



(b) $\sigma_{1,t}$, $\sigma_{2,t}$ Shift.



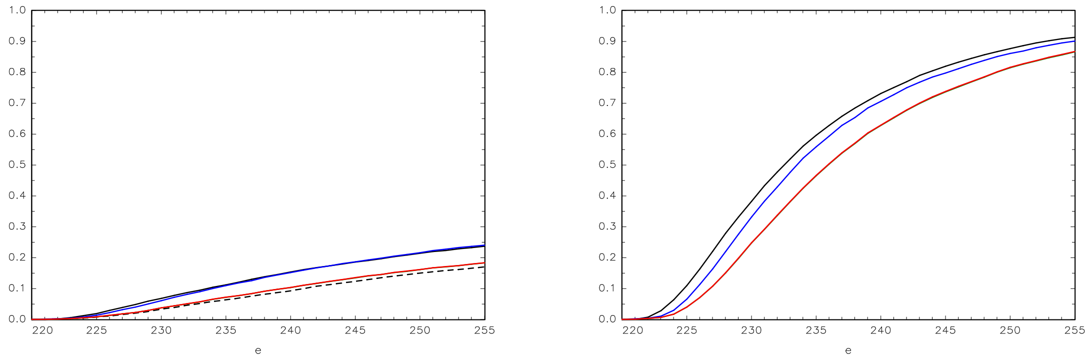
(c) $\sigma_{1,t}$ Shift.



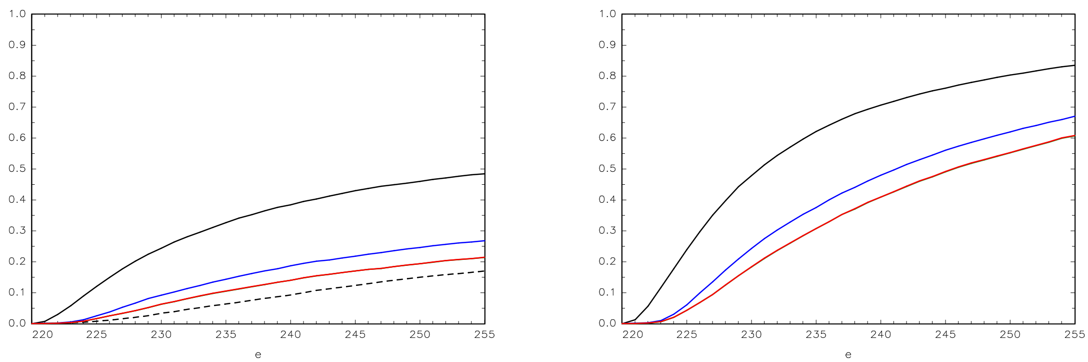
$\text{FPR}_{\text{i.i.d.}}$:– –, CUSUM:——, CUSUM^{V}:——, CUSUM^{V*}:——, CUSUM^{WMV}:——

Figure A.6: $\beta = 0.0$, $\rho = 0.0$, $\sigma_{12} = 0.4$, $\alpha_1 = 0.2$ - Left Panel=FPR, Right Panel =TPR. ($\varrho^2 = 1.000$)

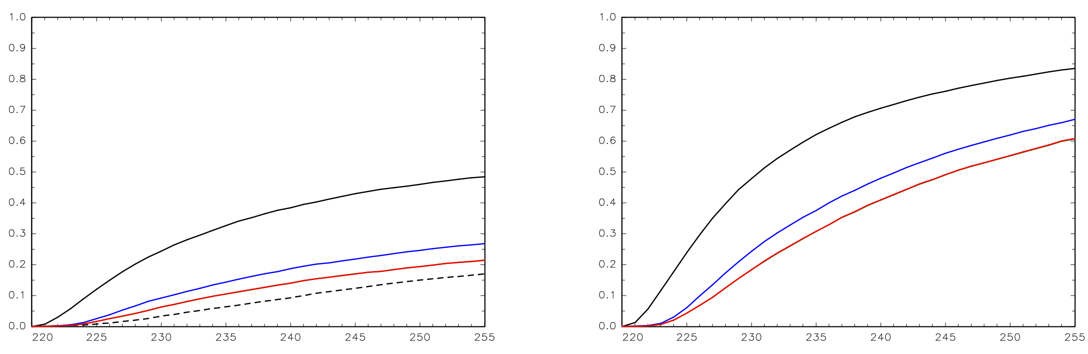
(a) Homoskedastic



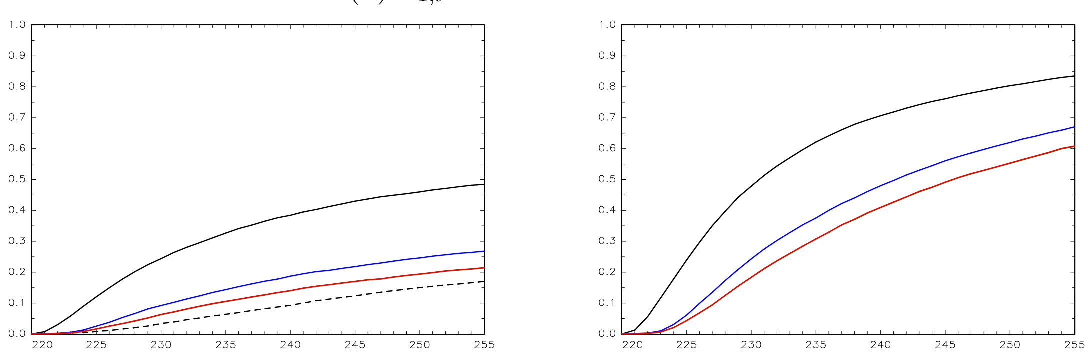
(b) $\sigma_{1,t}$, $\sigma_{2,t}$ Shift. Fixed Correlation



(c) $\sigma_{1,t}$ Shift. Fixed Correlation



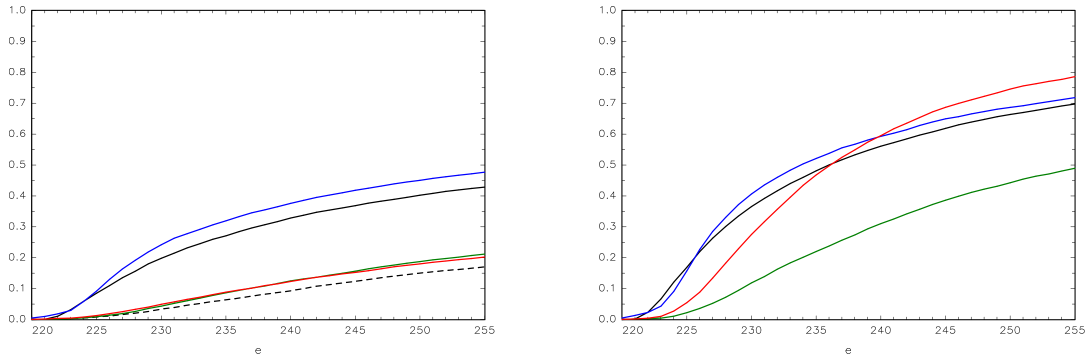
(d) $\sigma_{1,t}$ Shift. Correlation Varies



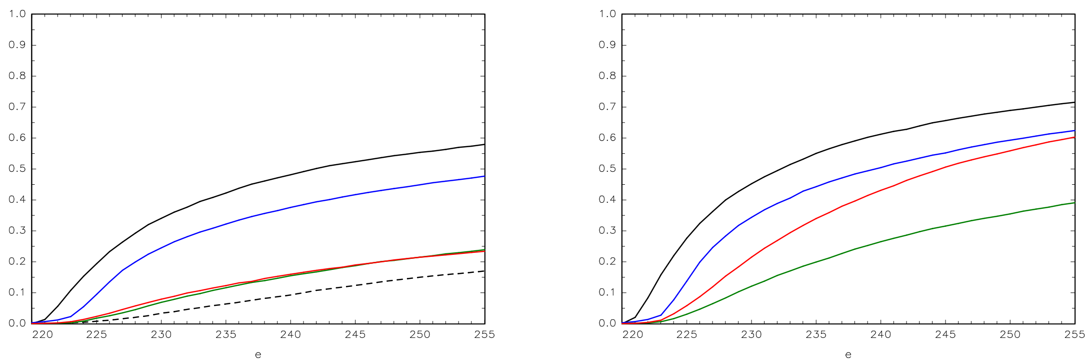
$\text{FPR}_{\text{i.i.d.}}$:– –, CUSUM:——, CUSUM^V:——, CUSUM^{V*}:——, CUSUM^{WMV}:——

Figure A.7: $\beta = 0.5$, $\rho = 0.8$, $\sigma_{12} = 0.4$, $\alpha_1 = 0.2$ - Left Panel=FPR, Right Panel =TPR. ($\varrho^2 = 0.432$)

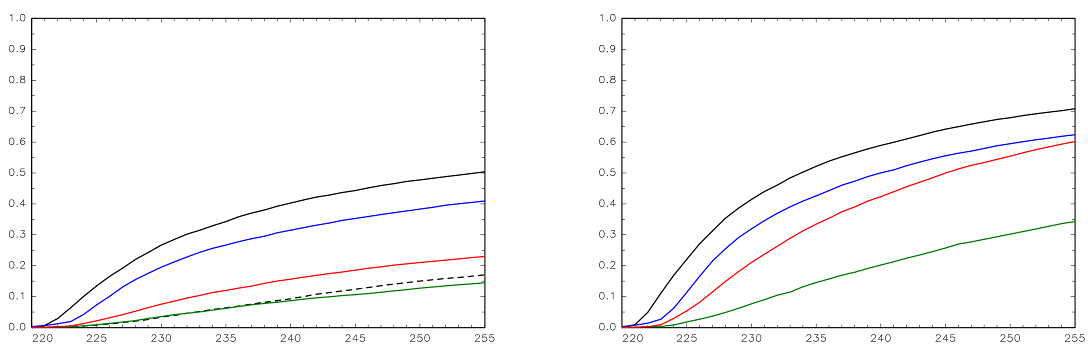
(a) Homoskedastic



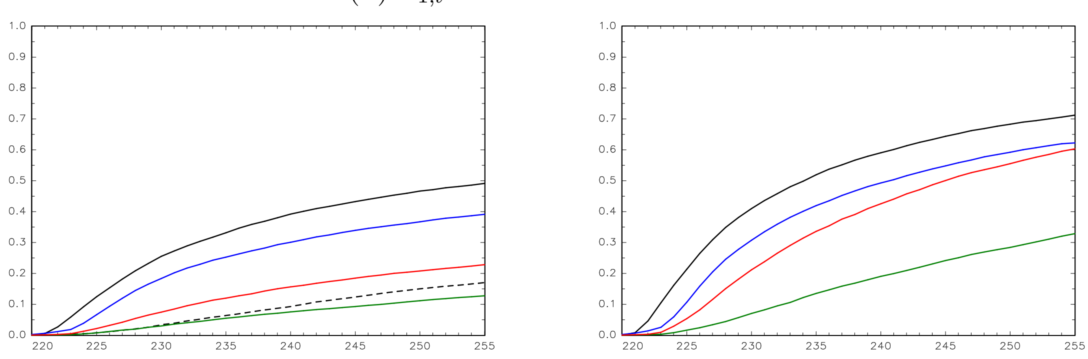
(b) $\sigma_{1,t}$, $\sigma_{2,t}$ Shift. Fixed Correlation



(c) $\sigma_{1,t}$ Shift. Fixed Correlation



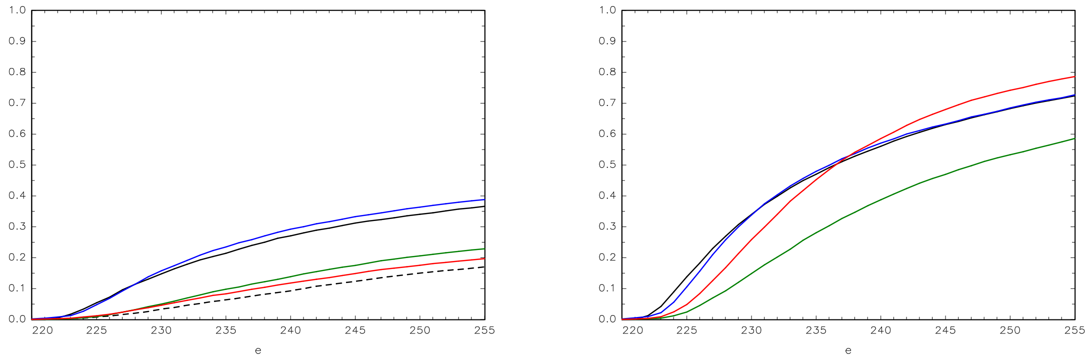
(d) $\sigma_{1,t}$ Shift. Correlation Varies



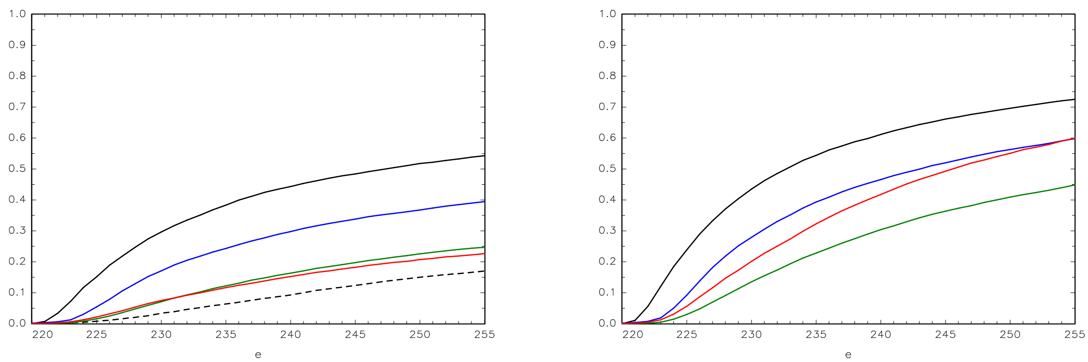
$\text{FPR}_{\text{i.i.d.}}$:– –, CUSUM:——, CUSUM^{V}:——, CUSUM^{V*}:——, CUSUM^{WMV}:——

Figure A.8: $\beta = -0.5$, $\rho = 0.8$, $\sigma_{12} = 0.4$, $\alpha_1 = 0.2$ - Left Panel=FPR, Right Panel =TPR. ($\varrho^2 = 0.000$)

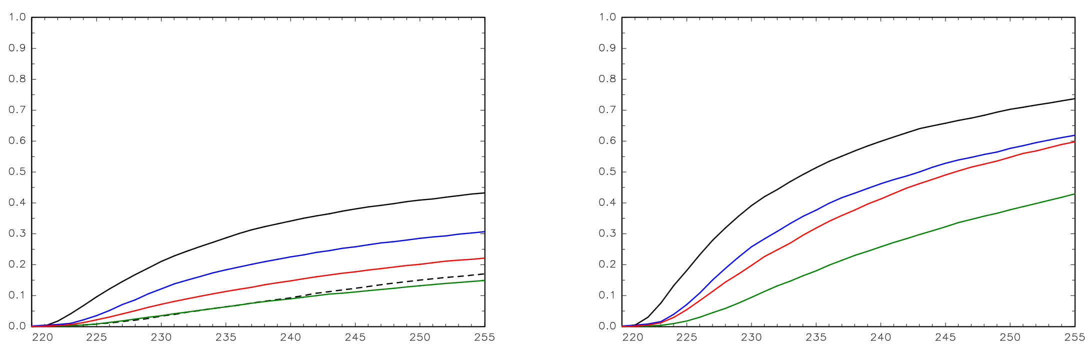
(a) Homoskedastic



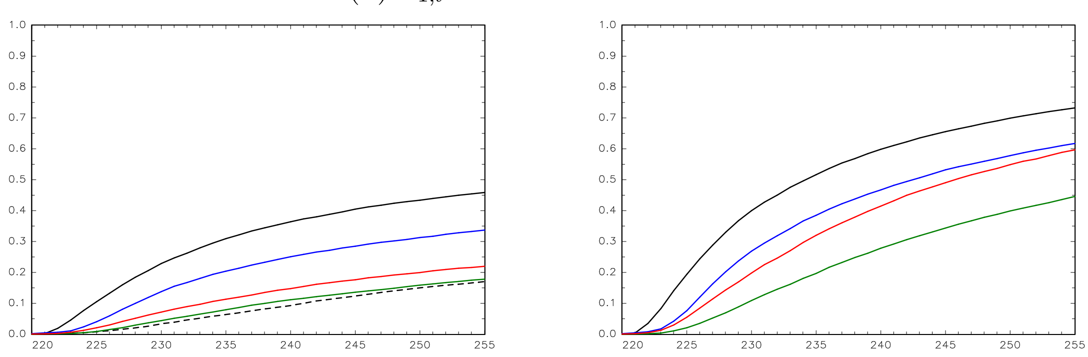
(b) $\sigma_{1,t}$, $\sigma_{2,t}$ Shift. Fixed Correlation



(c) $\sigma_{1,t}$ Shift. Fixed Correlation



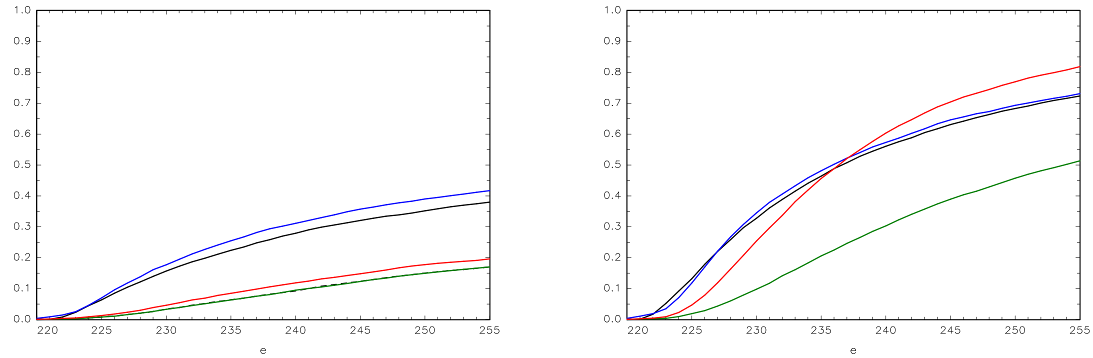
(d) $\sigma_{1,t}$ Shift. Correlation Varies



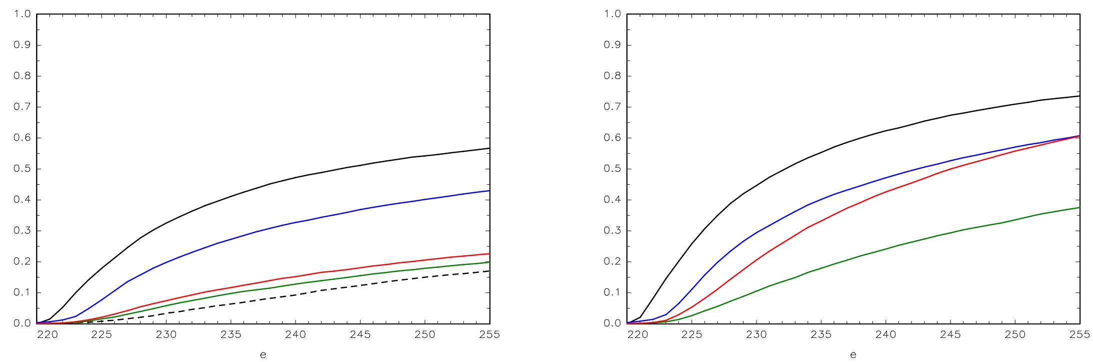
$\text{FPR}_{\text{i.i.d.}}$:$--$, CUSUM:———, CUSUM^V:———, CUSUM^{V*}:———, CUSUM^{WMV}:———

Figure A.9: $\beta = 0.8$, $\rho = 0.5$, $\sigma_{12} = 0.4$, $\alpha_1 = 0.2$ - Left Panel=FPR, Right Panel =TPR. ($\varrho^2 = 0.556$)

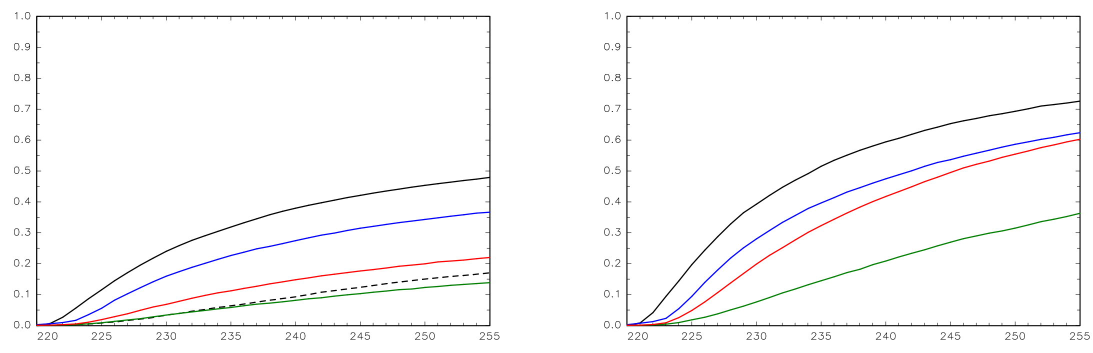
(a) Homoskedastic



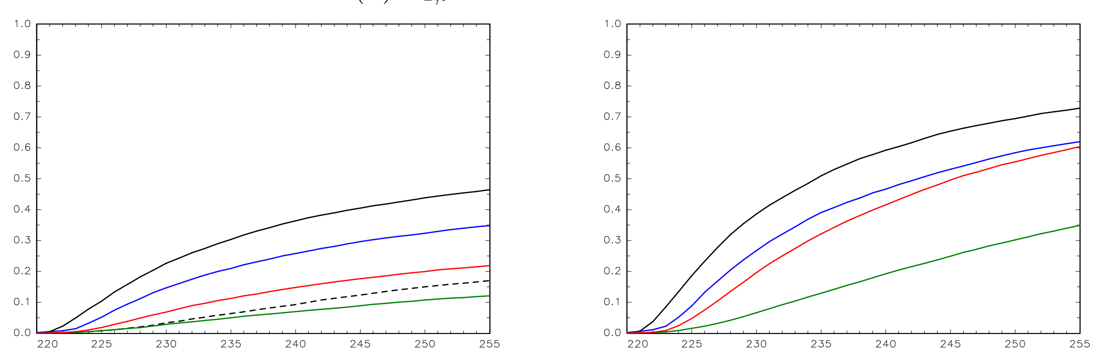
(b) $\sigma_{1,t}$, $\sigma_{2,t}$ Shift. Fixed Correlation



(c) $\sigma_{1,t}$ Shift. Fixed Correlation



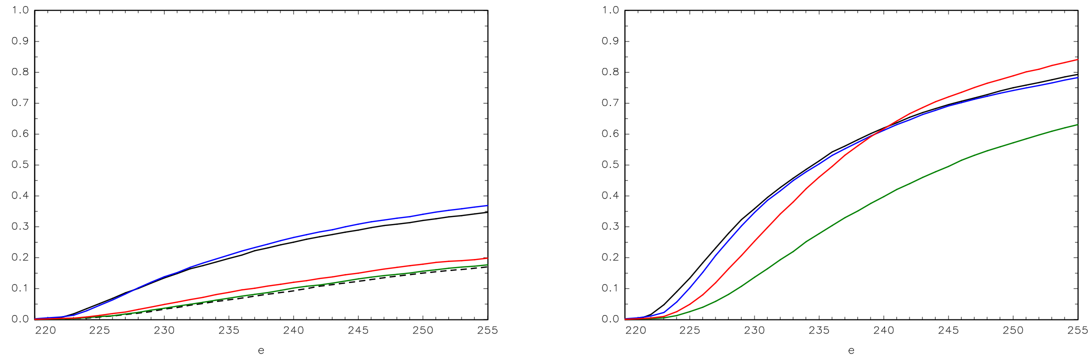
(d) $\sigma_{1,t}$ Shift. Correlation Varies



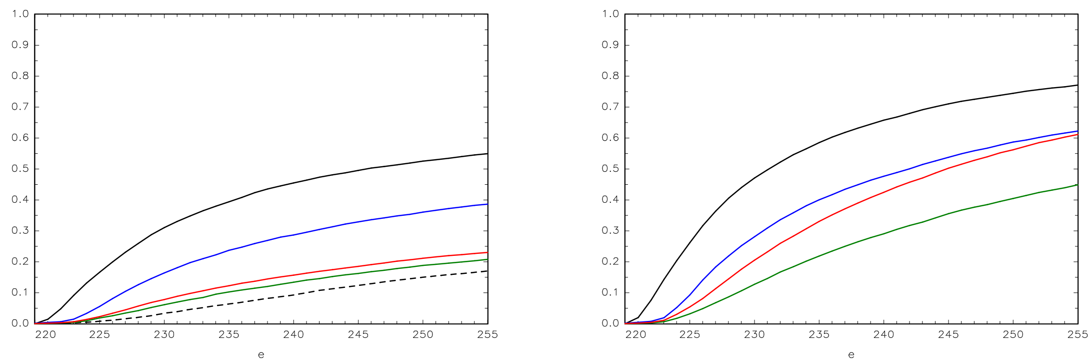
$\text{FPR}_{\text{i.i.d.}}$:− −, CUSUM:——, CUSUM^{V}:——, CUSUM^{V*}:——, CUSUM^{WMV}:——

Figure A.10: $\beta = 0.5$, $\rho = 0.5$, $\sigma_{12} = 0.4$, $\alpha_1 = 0.2$ - Left Panel=FPR, Right Panel =TPR. ($\varrho^2 = 0.700$)

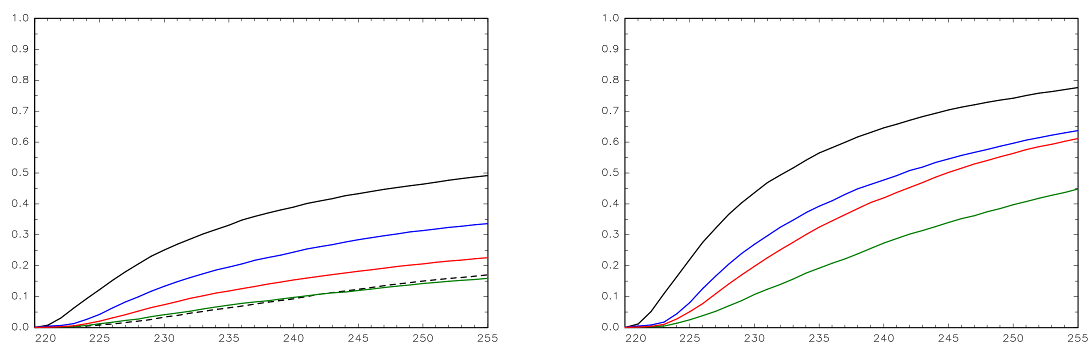
(a) Homoskedastic



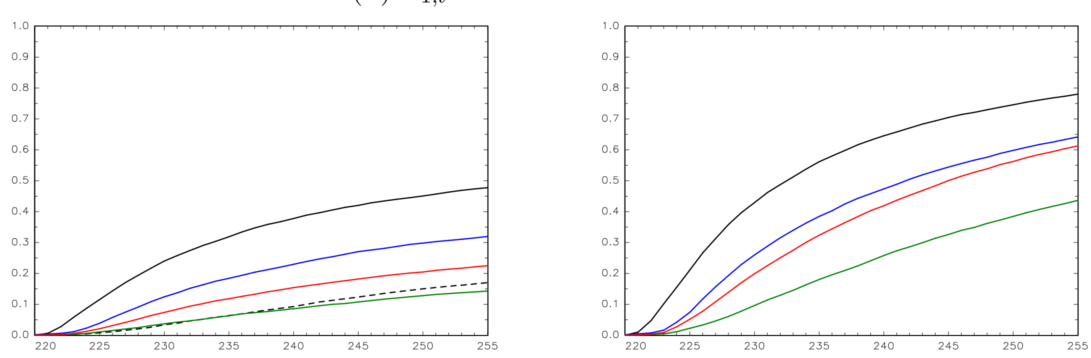
(b) $\sigma_{1,t}$, $\sigma_{2,t}$ Shift. Fixed Correlation



(c) $\sigma_{1,t}$ Shift. Fixed Correlation



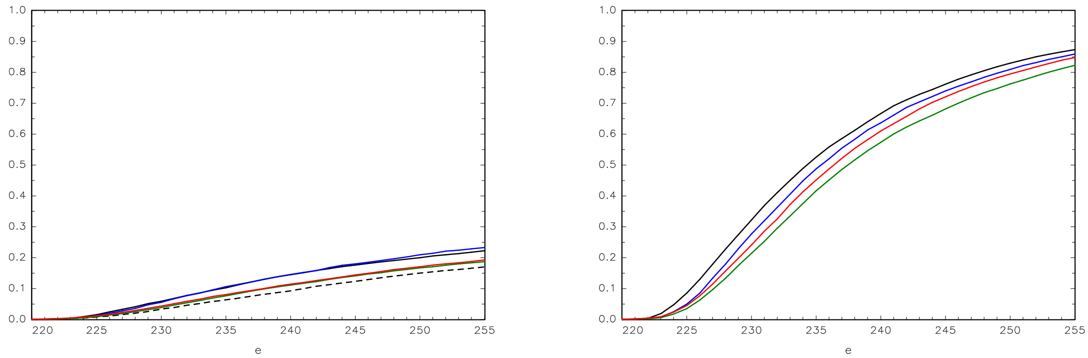
(d) $\sigma_{1,t}$ Shift. Correlation Varies



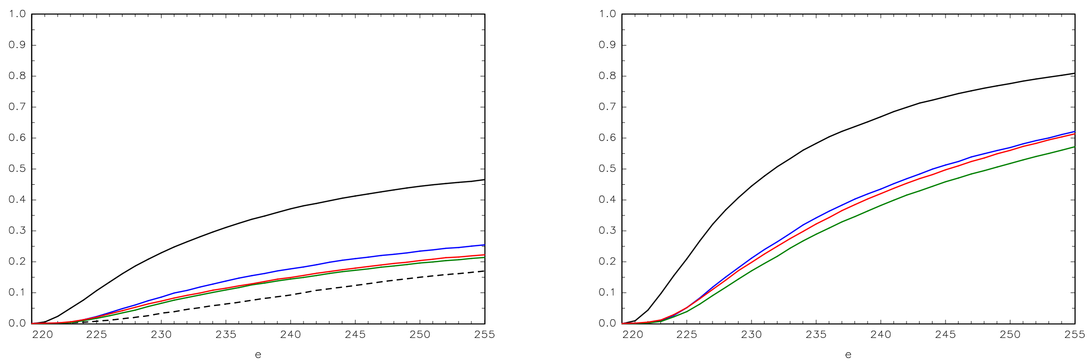
$\text{FPR}_{\text{i.i.d.}}$:– –, CUSUM:——, CUSUM^{V}:——, CUSUM^{V*}:——, CUSUM^{WMV}:——

Figure A.11: $\beta = -0.5$, $\rho = 0.5$, $\sigma_{12} = 0.4$, $\alpha_1 = 0.2$ - Left Panel=FPR, Right Panel =TPR. ($\varrho^2 = 0.300$)

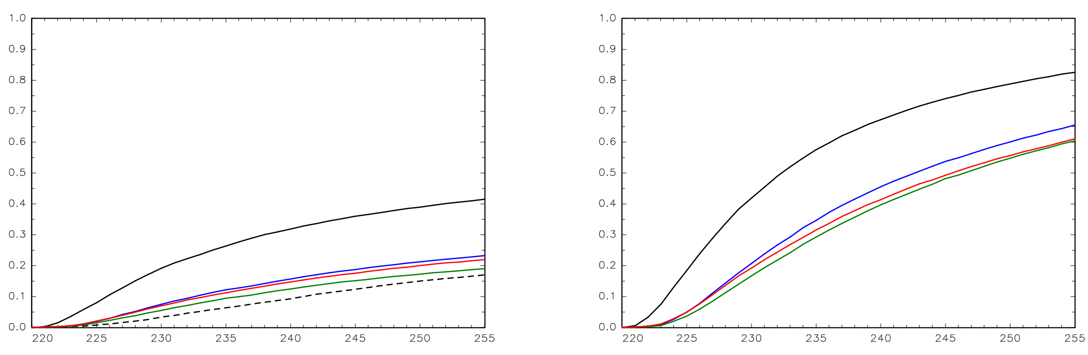
(a) Homoskedastic



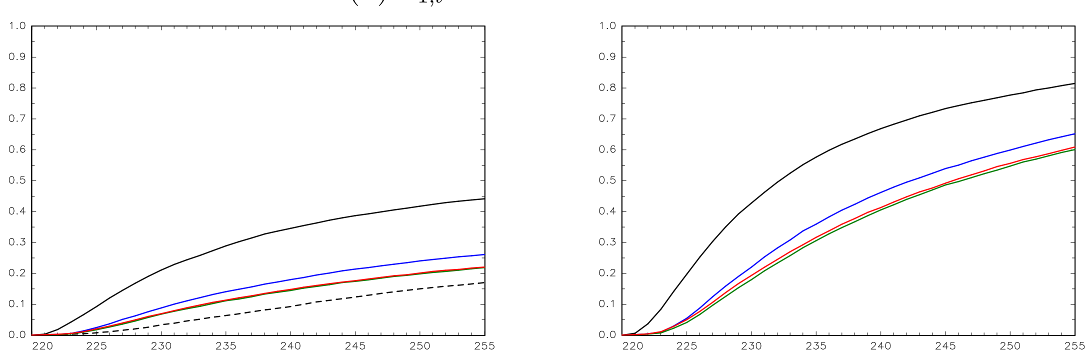
(b) $\sigma_{1,t}$, $\sigma_{2,t}$ Shift. Fixed Correlation



(c) $\sigma_{1,t}$ Shift. Fixed Correlation



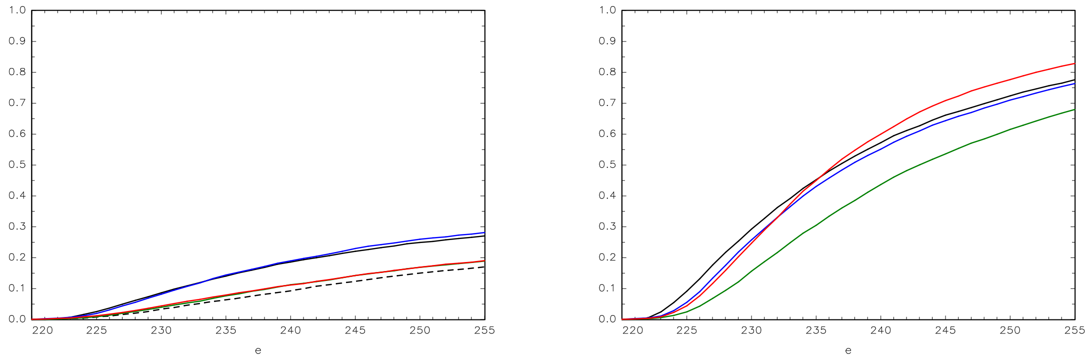
(d) $\sigma_{1,t}$ Shift. Correlation Varies



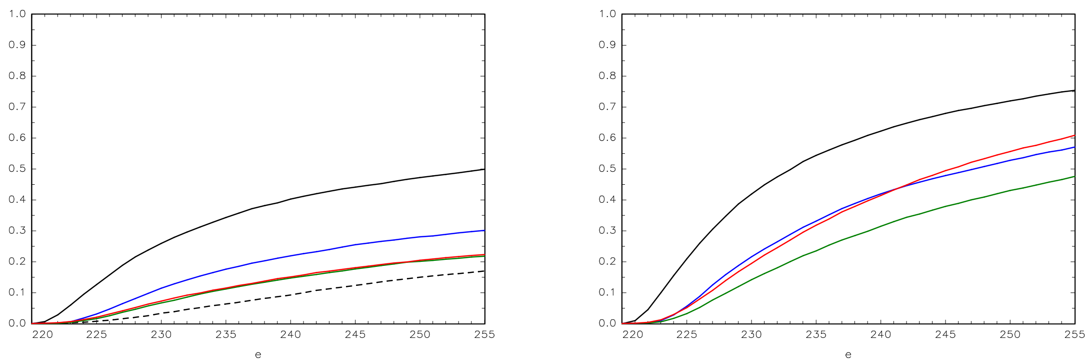
$\text{FPR}_{\text{i.i.d.}}$:− −, CUSUM:——, CUSUM^{V}:——, CUSUM^{V*}:——, CUSUM^{WMV}:——

Figure A.12: $\beta = -0.8$, $\rho = 0.5$, $\sigma_{12} = 0.4$, $\alpha_1 = 0.2$ - Left Panel=FPR, Right Panel =TPR. ($\varrho^2 = 0.057$)

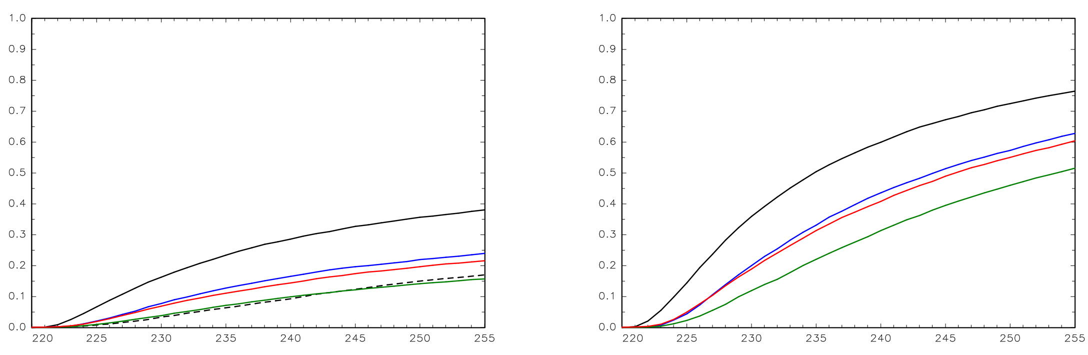
(a) Homoskedastic



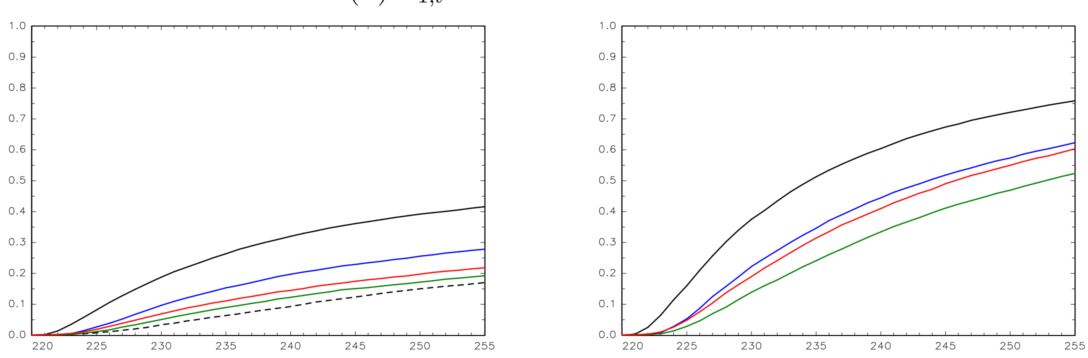
(b) $\sigma_{1,t}$, $\sigma_{2,t}$ Shift. Fixed Correlation



(c) $\sigma_{1,t}$ Shift. Fixed Correlation



(d) $\sigma_{1,t}$ Shift. Correlation Varies



$\text{FPR}_{\text{i.i.d.}}$:– –, CUSUM:——, CUSUM^V:——, CUSUM^{V*}:——, CUSUM^{WMV}:——

A.4.2 Additional Simulations - Smooth Volatility Shift in Innovations to Covariate

Figures A.13-A.15 report the TPR and FPR of the monitoring procedures when a smooth volatility shift is present in the innovations to the covariate only. Data were generated according to (1)-(2) and (23)-(25), with $\sigma_{1,t} = 1$, $\sigma_{2,t} := 1 + (\sqrt{4} - 1)\,[1 + \exp(-\theta(t - 219))]^{-1}$ $\sigma_{12,t} = \sigma_{12}\sigma_{2,t}$ and we again set $u_0 = 100$, $w_0 = \mu = 0$. Under the null we set $\delta = 0$, whereas under the alternative we set $\delta = 0.005$, $\tau_1 T = 220$ and $\tau_2 T = \lambda T$.

Figure A.13 (a) shows that when x_t does not enter the DGP for y_t in any way that, unsurprisingly, none of the monitoring procedures are impacted in any meaningful way by the volatility shift present in the innovations to the covariate. The FPR and TPR for the CUSUM^{WMV} procedure is near identical to the homoskedastic case, and the FPR and TPR of all other procedures are exactly identical to the homoskedastic case reported in Figure 1 (a) of the main paper.

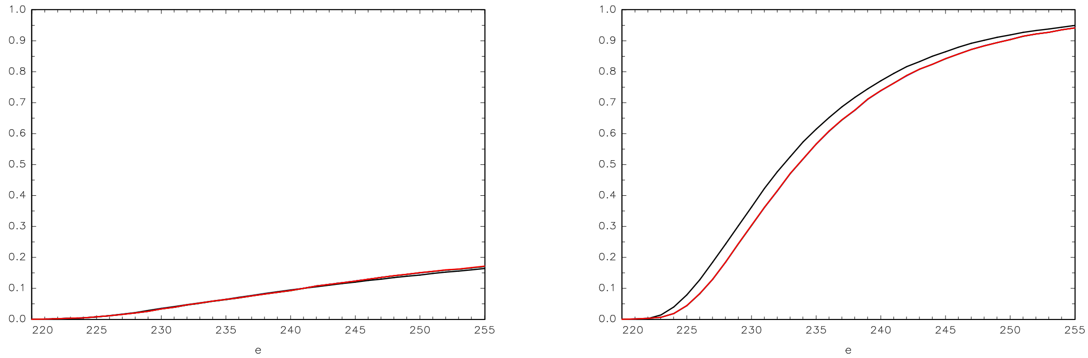
Figure A.13 (b) and (c) report results for the case where the covariate is relevant but no serial correlation is present in v_t. In this scenario the heteroskedasticity in x_t feeds through into the values of Δy_t that are used to construct the statistics underlying the CUSUM and CUSUM^V procedures. The former suffers FPR distortions as a consequence, whereas the CUSUM^V procedure is still able to control FPR to a decent extent due to the use of the kernel based variance estimator. Due to the lack of serial correlation the BIC reduces the CUSUM^{V*} procedure to the CUSUM^V procedure in a great majority of replications so that the blue and green lines almost exactly coincide. The CUSUM^{WMV} procedure is, expectedly, also FPR controlled in these scenarios.

Figure A.13 (d) reports results for the case where the covariate is irrelevant, but serial correlation is present in v_t. Here both the standard CUSUM and CUSUM^V procedures exhibit FPR distortions, where the CUSUM^{WMV} and CUSUM^{V*} procedures display much better FPR control. The latter two procedures display almost identical FPR/TPR profiles in this scenario as when the covariate is irrelevant the BIC reduces the CUSUM^{WMV} procedure to the CUSUM^{V*} procedure in a vast majority of replications.

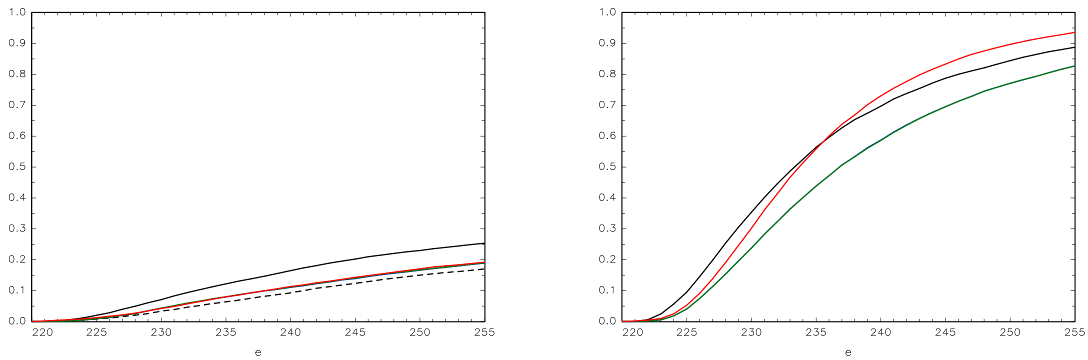
The remaining figures A.14 and A.15 report the FPR and TPR of the procedures for the same parameter constellations as considered by CSS. Across these scenarios both the standard CUSUM and CUSUM^V procedures display very poor FPR control as neither are able to model the serial correlation present in v_t, and the former uses a standard variance estimator that is unable to account for heteroskedasticity. The FPR control of the CUSUM^{V*} procedure is also poor, with this procedure exhibiting upward FPR distortions rather than the downward FPR distortions it exhibits when the variance of only $\epsilon_{1,t}$ shifts. The best FPR control overall is clearly displayed by the CUSUM^{WMV} procedure. The CUSUM^{WMV} procedure also displays far superior TPR performance to the CUSUM^{V*} procedure across all scenarios.

Figure A.13: Volatility Shift in $\varepsilon_{2,t}$ only - Left Panel=FPR, Right Panel =TPR.

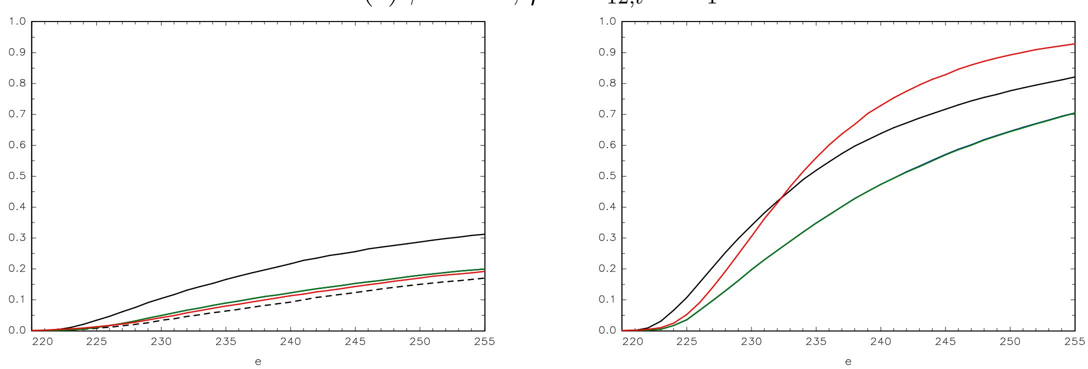
(a) $\beta = \rho = \sigma_{12} = \alpha_1 = 0$



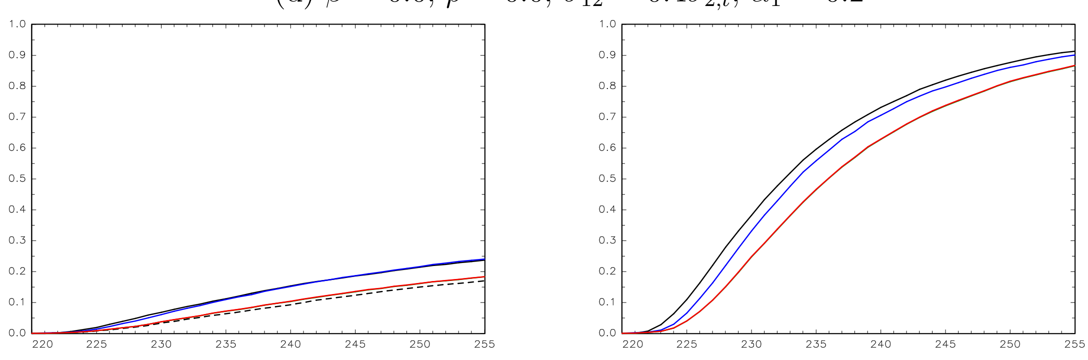
(b) $\beta = 0.5,\ \rho = \sigma_{12} = \alpha_1 = 0$



(c) $\beta = 0.8,\ \rho = \sigma_{12,t} = \alpha_1 = 0$



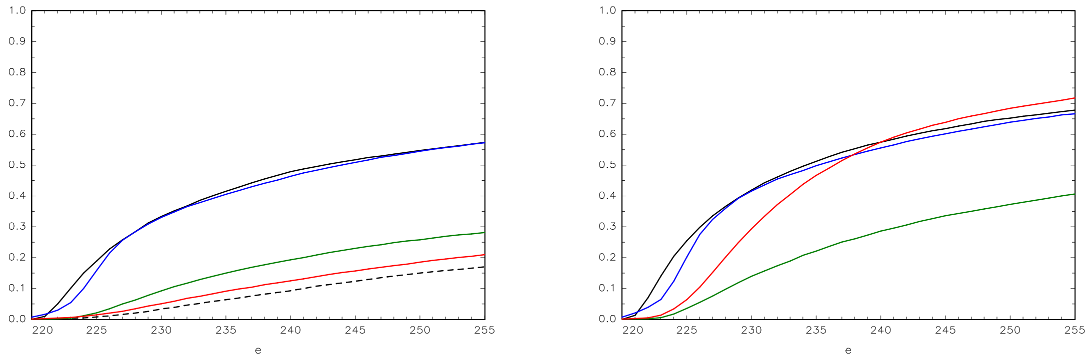
(d) $\beta = 0.0,\ \rho = 0.0,\ \sigma_{12} = 0.4\sigma_{2,t},\ \alpha_1 = 0.2$



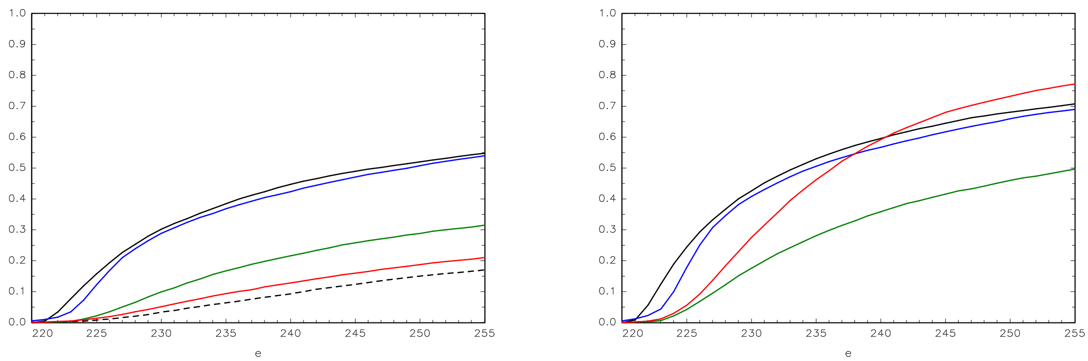
$\text{FPR}_{\text{i.i.d.}}$:– –, CUSUM:——, CUSUM^{V}:——, CUSUM^{V*}:——, CUSUM^{WMV}:——

Figure A.14: Volatility Shift in $\varepsilon_{2,t}$ only - Left Panel=FPR, Right Panel =TPR.

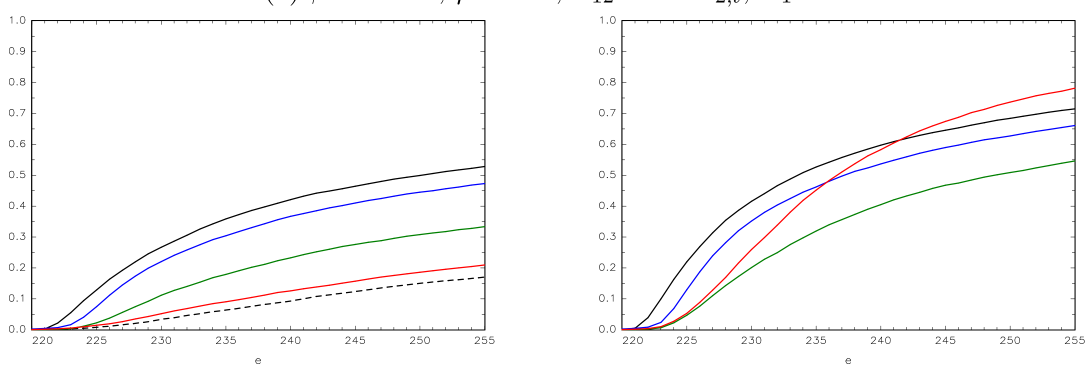
(a) $\beta = 0.8$, $\rho = 0.8$, $\sigma_{12} = 0.4\sigma_{2,t}$, $\alpha_1 = 0.2$



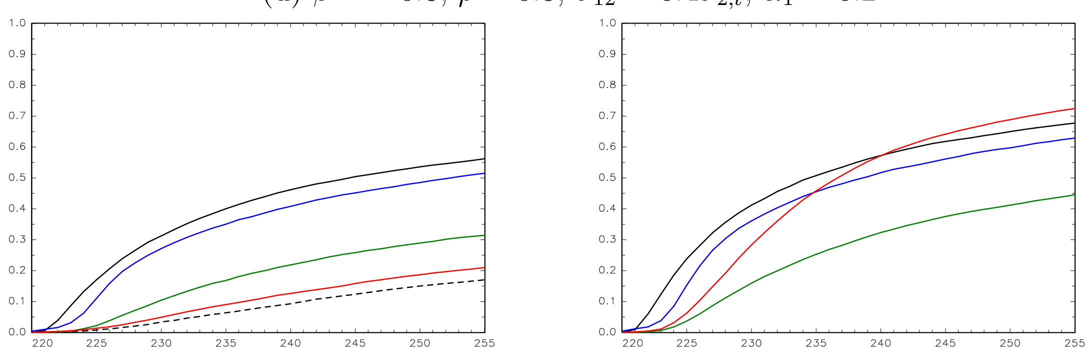
(b) $\beta = 0.5$, $\rho = 0.8$, $\sigma_{12} = 0.4\sigma_{2,t}$, $\alpha_1 = 0.2$



(c) $\beta = -0.5$, $\rho = 0.8$, $\sigma_{12} = 0.4\sigma_{2,t}$, $\alpha_1 = 0.2$



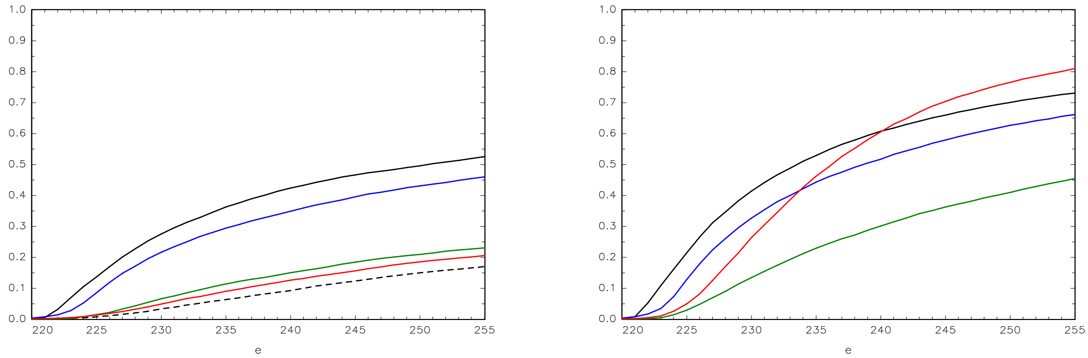
(d) $\beta = -0.8$, $\rho = 0.8$, $\sigma_{12} = 0.4\sigma_{2,t}$, $\alpha_1 = 0.2$



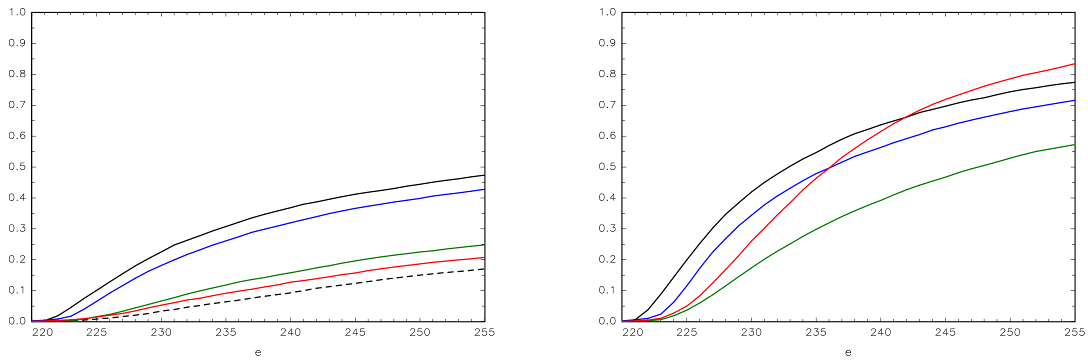
$\text{FPR}_{\text{i.i.d.}}$:– –, CUSUM:——, CUSUM^{V}:——, CUSUM^{V*}:——, CUSUM^{WMV}:——

Figure A.15: Volatility Shift in $\varepsilon_{2,t}$ only - Left Panel=FPR, Right Panel =TPR.

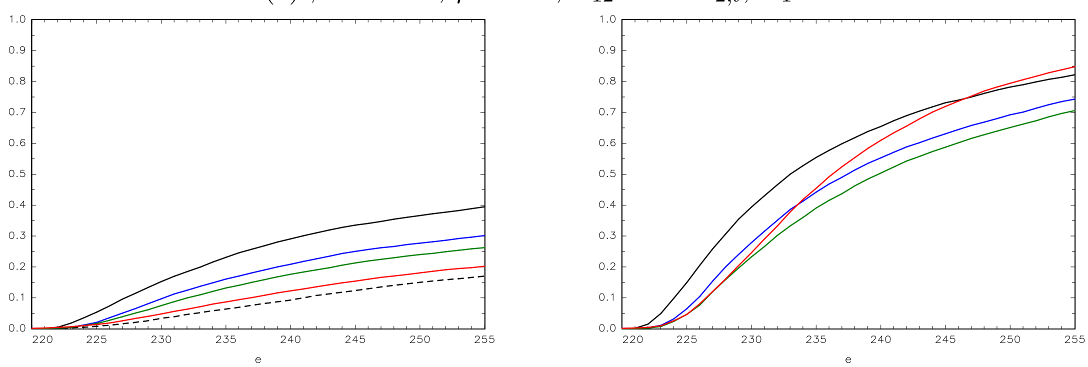
(a) $\beta = 0.8$, $\rho = 0.5$, $\sigma_{12} = 0.4\sigma_{2,t}$, $\alpha_1 = 0.2$



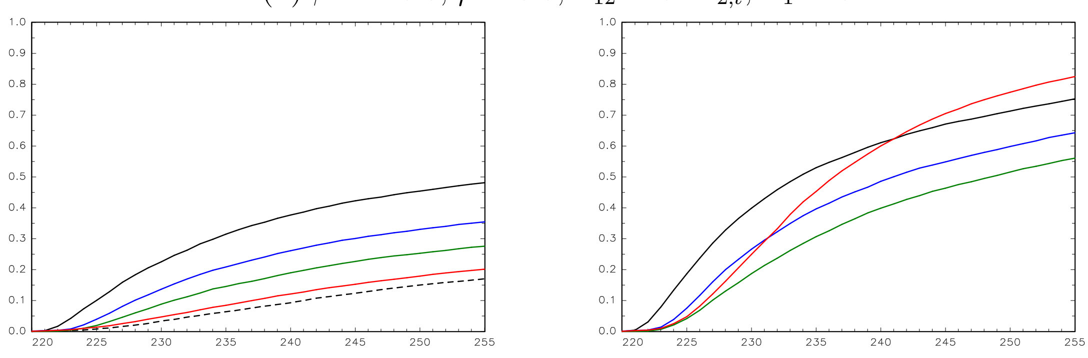
(b) $\beta = 0.5$, $\rho = 0.5$, $\sigma_{12} = 0.4\sigma_{2,t}$, $\alpha_1 = 0.2$



(c) $\beta = -0.5$, $\rho = 0.5$, $\sigma_{12} = 0.4\sigma_{2,t}$, $\alpha_1 = 0.2$



(d) $\beta = -0.8$, $\rho = 0.5$, $\sigma_{12} = 0.4\sigma_{2,t}$, $\alpha_1 = 0.2$



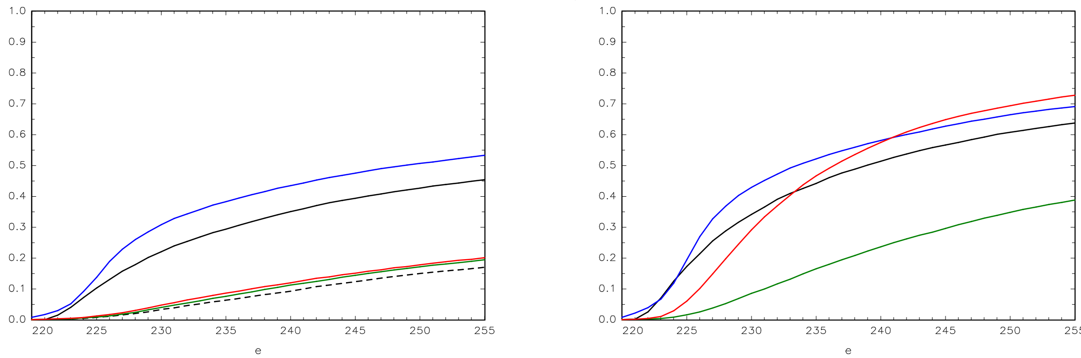
$\text{FPR}_{\text{i.i.d.}}$:– –, CUSUM:——, CUSUM^{V}:——, CUSUM^{V*}:——, CUSUM^{WMV}:——

A.4.3 Additional Simulations - Covariate Observed with Measurement Error

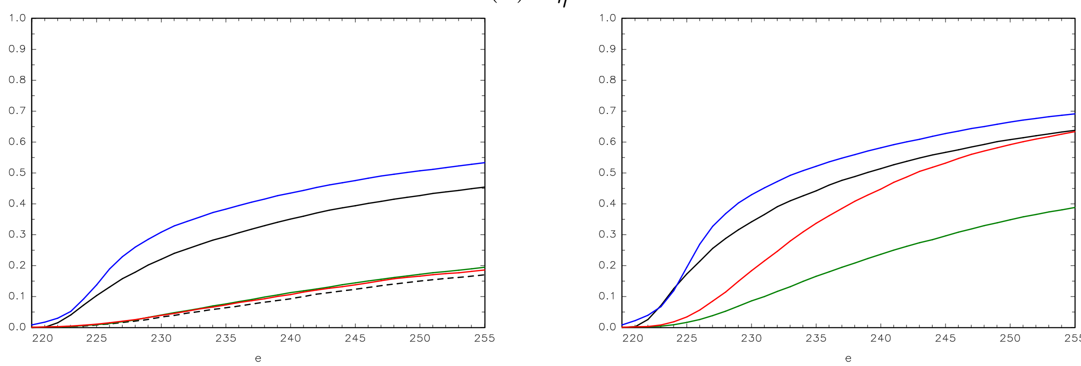
Figures A.16-A.23 report the FPR and TPR of the monitoring procedures in the same scenarios as panel (a) in Figures 3-4 and A.7-A.12 but where the covariate x_t is observed subject to a measurement error. Data were generated according to (1)-(2) and (23)-(25), with $\sigma^2_{1,t} = \sigma^2_{2,t} = 1$, $\sigma_{12,t} = 0.4$, $\alpha_1 = 0.2$ and we again set $u_0 = 100$, $w_0 = \mu = 0$. Under the null we set $\delta = 0$, whereas under the alternative we set $\delta = 0.005$, $\tau_1 T = 220$ and $\tau_2 T = \lambda T$. Results are reported for the case where the covariate $z_t = x_t + \eta_t$ is used in the CUSUM^{WMV} procedure where $\eta_t \sim N(0, \sigma^2_\eta)$ (Setting $\sigma^2_\eta = 0$ corresponds to the case where the covariate is observed without measurement error). In all cases, the FPR of our proposed CUSUM^{WMV} monitoring procedure is impacted very little by the measurement error, whereas under the alternative the TPR of the procedure is somewhat reduced relative to the case where the covariate is observed without measurement error, with this power reduction increasing in the value of σ^2_η. The FPR and TPR of the CUSUM, CUSUM^V and CUSUM^{V*} procedures does not change with the value of σ^2_η as the true DGP remains the same across the various values of σ^2_η and these procedures make no use of the observed covariate z_t. In all cases the TPR of the CUSUM^{WMV} procedure is superior to that of the CUSUM^{V*} procedure which is the only other FPR controlled test in these scenarios.

Figure A.16: Covariate observed with error:$\beta = 0.8$, $\rho = 0.8$, $\sigma_{12} = 0.4$, $\alpha_1 = 0.2$ - Left=FPR, Right=TPR

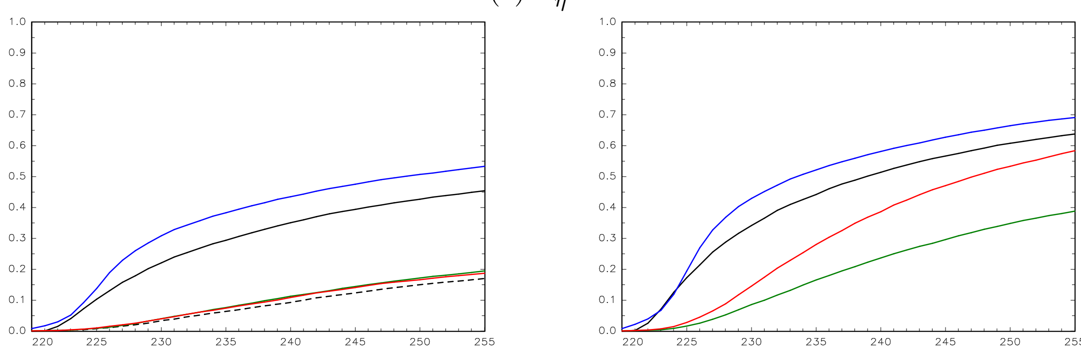
(a) $\sigma_\eta^2 = 0$



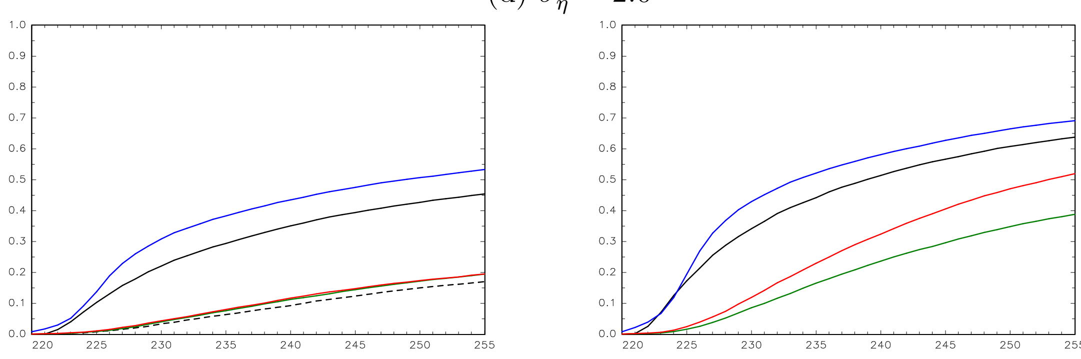
(b) $\sigma_\eta^2 = 0.5$



(c) $\sigma_\eta^2 = 1.0$



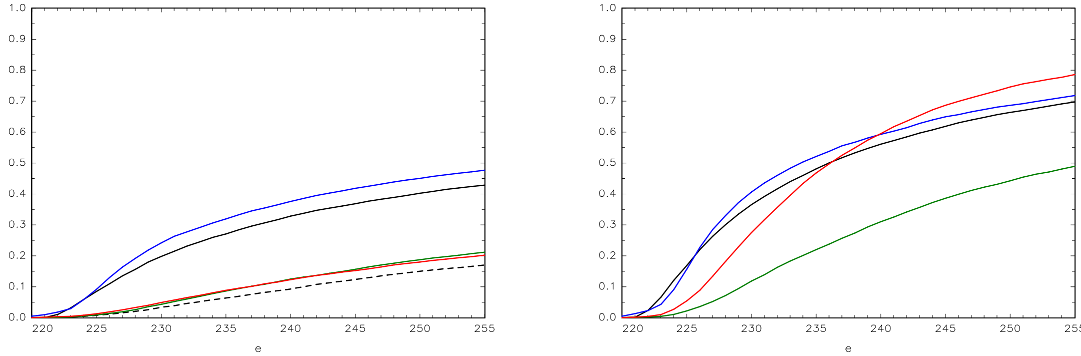
(d) $\sigma_\eta^2 = 2.0$



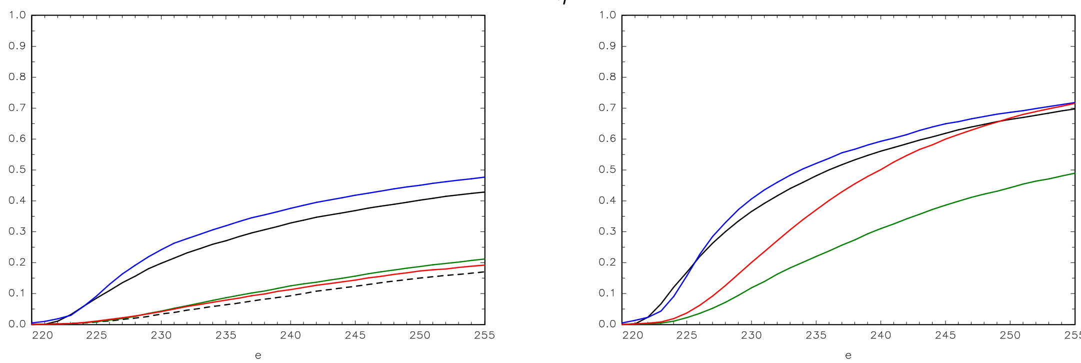
$\text{FPR}_{\text{i.i.d.}}$:– –, CUSUM:——, CUSUM^{V}:——, CUSUM^{V*}:——, CUSUM^{WMV}:——

Figure A.17: Covariate observed with error: $\beta = 0.5$, $\rho = 0.8$, $\sigma_{12} = 0.4$, $\alpha_1 = 0.2$ - Left=FPR, Right=TPR

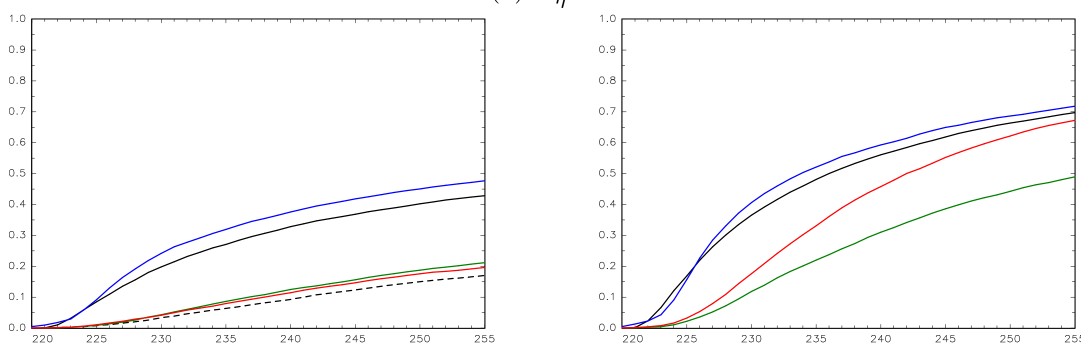
(a) $\sigma_\eta^2 = 0$



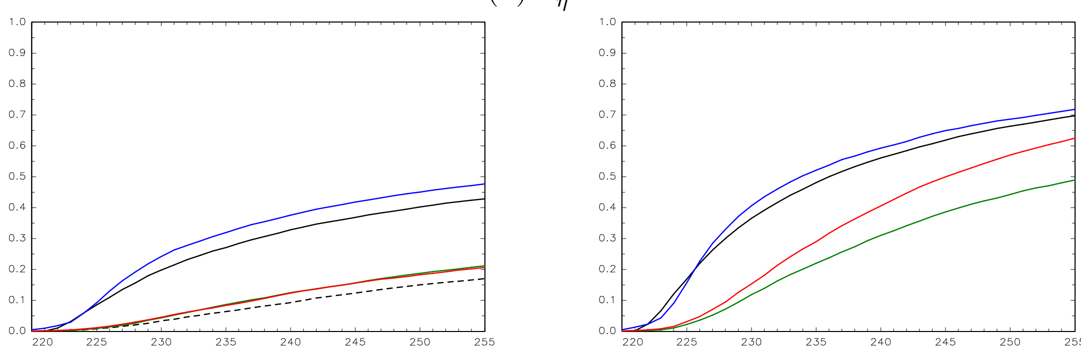
(b) $\sigma_\eta^2 = 0.5$



(c) $\sigma_\eta^2 = 1.0$



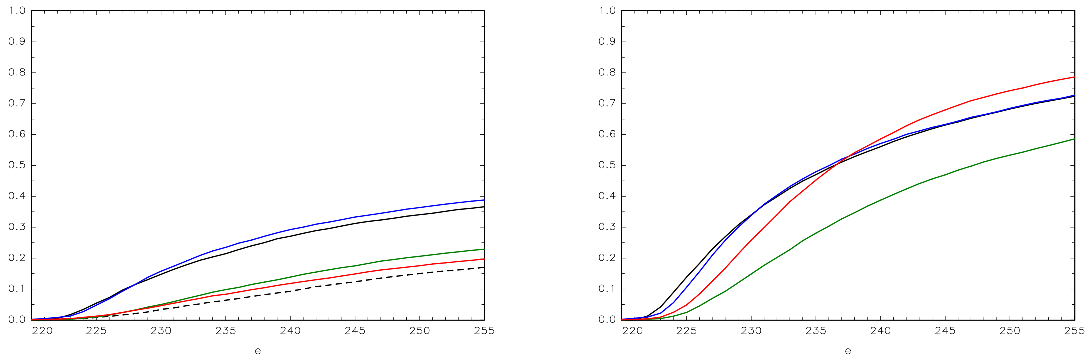
(d) $\sigma_\eta^2 = 2.0$



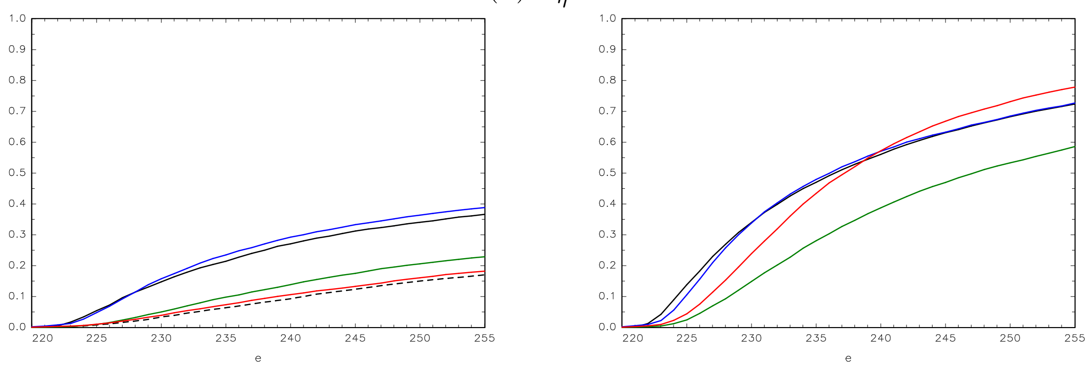
$\text{FPR}_{\text{i.i.d.}}$:– –, CUSUM:——, CUSUM^{V}:——, CUSUM^{V*}:——, CUSUM^{WMV}:——

Figure A.18: Covariate observed with error: $\beta = -0.5$, $\rho = 0.8$, $\sigma_{12} = 0.4$, $\alpha_1 = 0.2$ - Left=FPR, Right=TPR

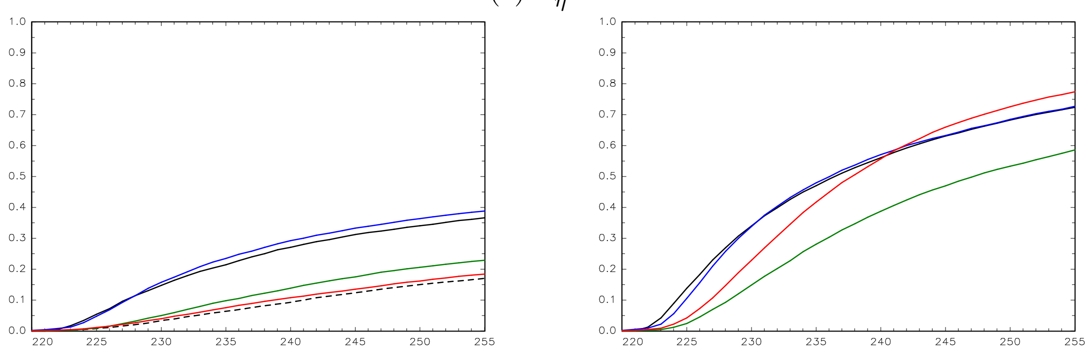
(a) $\sigma_\eta^2 = 0$



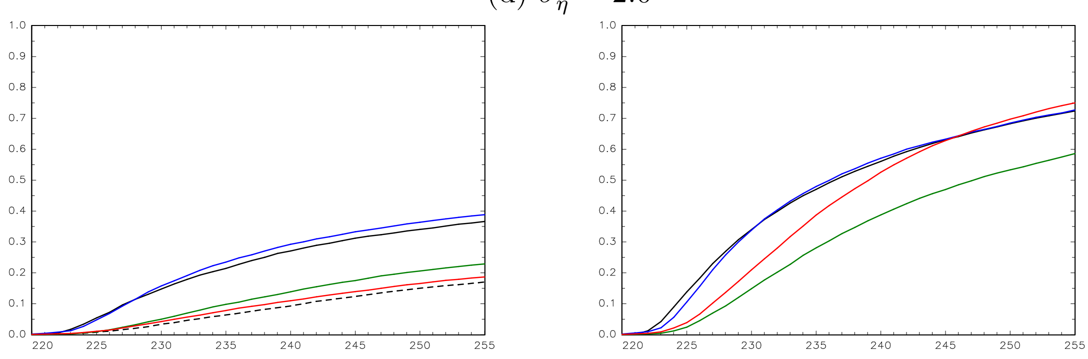
(b) $\sigma_\eta^2 = 0.5$



(c) $\sigma_\eta^2 = 1.0$



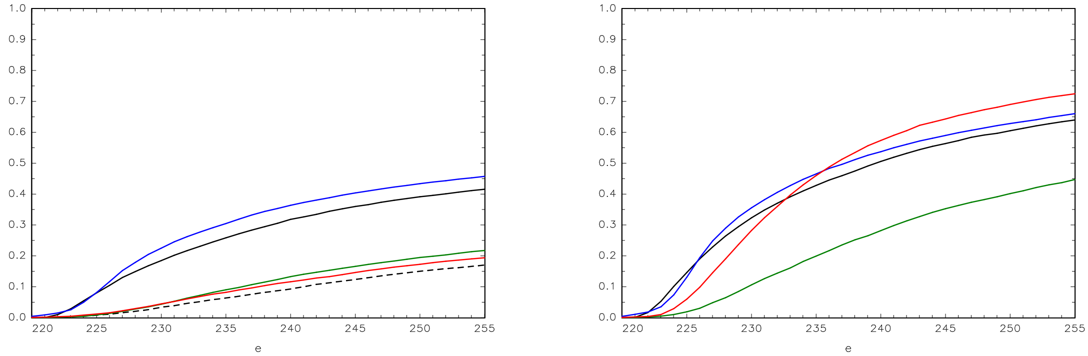
(d) $\sigma_\eta^2 = 2.0$



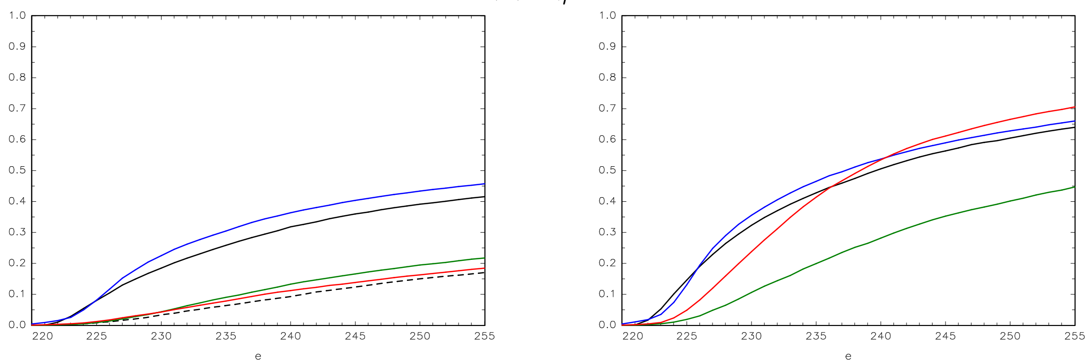
$FPR_{\text{i.i.d.}}$: – –, CUSUM: ——, $CUSUM^V$: ——, $CUSUM^{V*}$: ——, $CUSUM^{WMV}$: ——

Figure A.19: Covariate observed with error:$\beta = -0.8$, $\rho = 0.8$, $\sigma_{12} = 0.4$, $\alpha_1 = 0.2$ - Left=FPR, Right=TPR

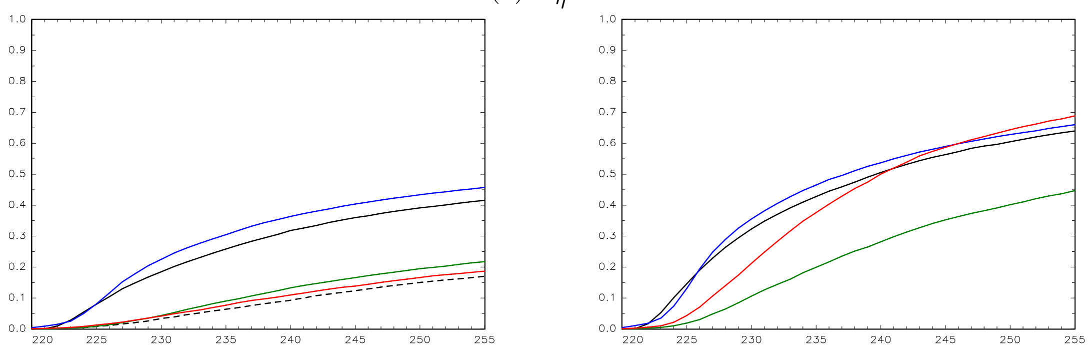
(a) $\sigma^2_\eta = 0$



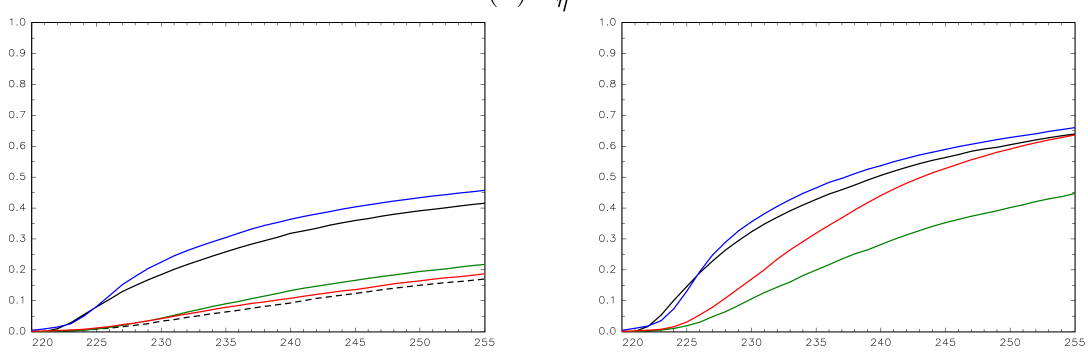
(b) $\sigma^2_\eta = 0.5$



(c) $\sigma^2_\eta = 1.0$



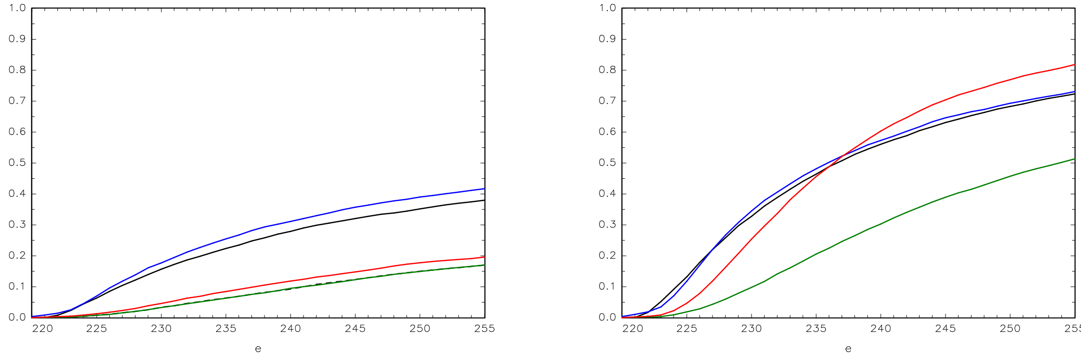
(d) $\sigma^2_\eta = 2.0$



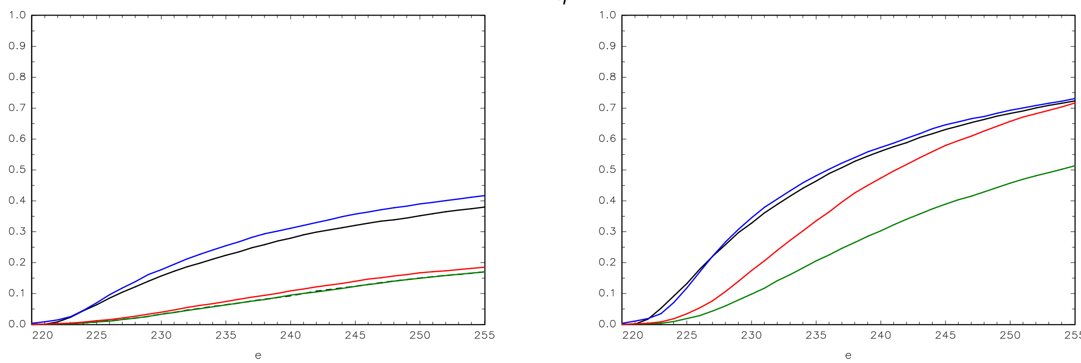
$\text{FPR}_{\text{i.i.d.}}$:– –, CUSUM:——, CUSUM^{V}:——, CUSUM^{V*}:——, CUSUM^{WMV}:——

Figure A.20: Covariate observed with error:$\beta = 0.8$, $\rho = 0.5$, $\sigma_{12} = 0.4$, $\alpha_1 = 0.2$ - Left=FPR, Right=TPR

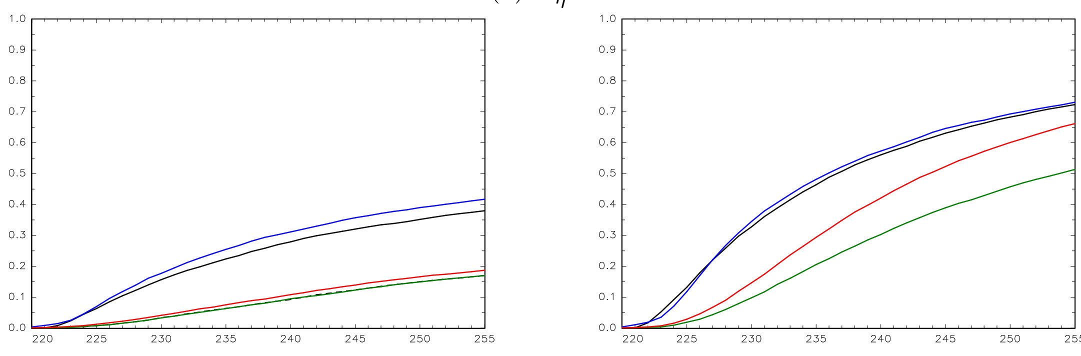
(a) $\sigma_\eta^2 = 0$



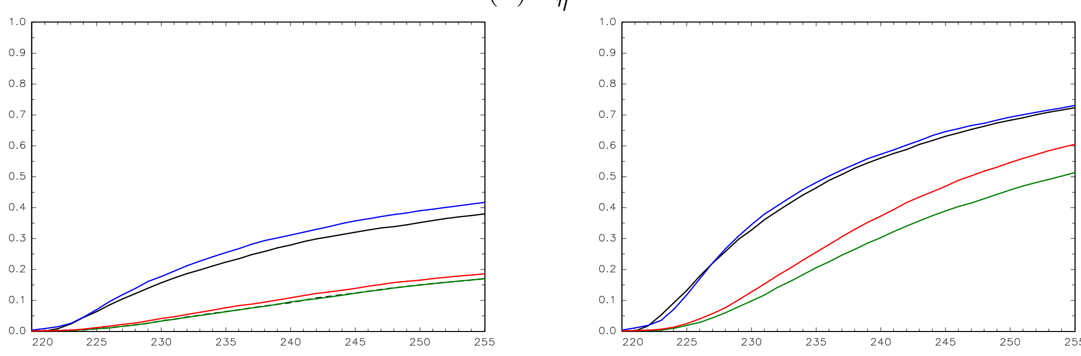
(b) $\sigma_\eta^2 = 0.5$



(c) $\sigma_\eta^2 = 1.0$



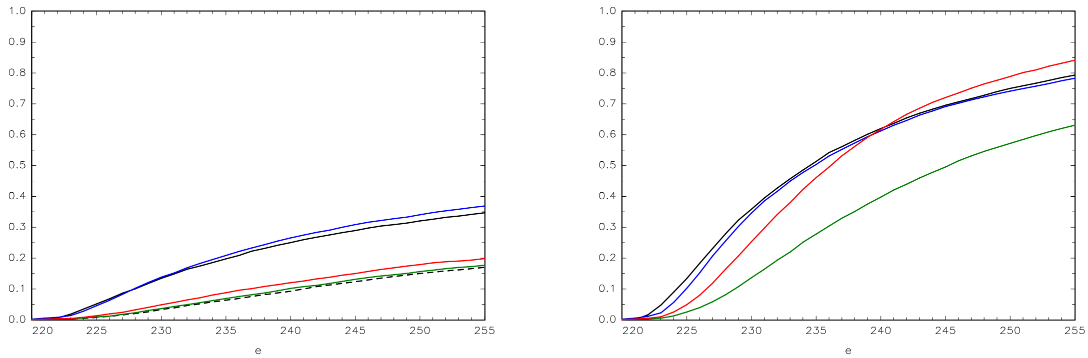
(d) $\sigma_\eta^2 = 2.0$



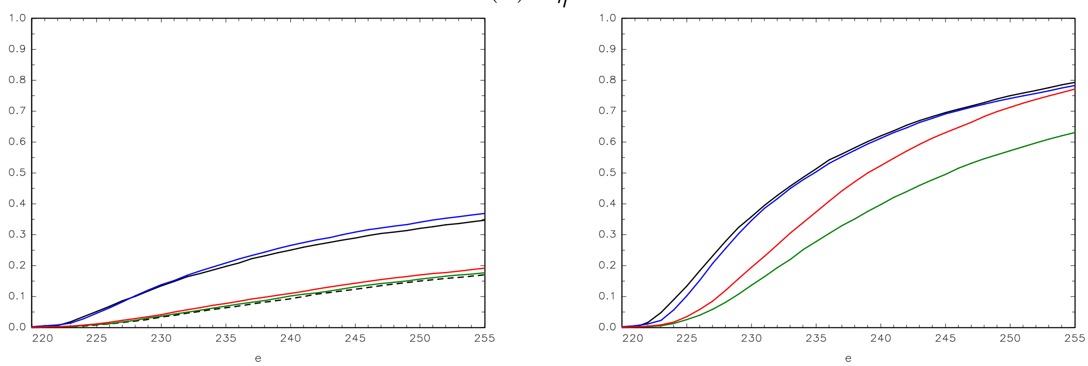
$FPR_{\text{i.i.d.}}$:− −, CUSUM:——, $CUSUM^{V}$:——, $CUSUM^{V*}$:——, $CUSUM^{WMV}$:——

Figure A.21: Covariate observed with error: $\beta = 0.5$, $\rho = 0.5$, $\sigma_{12} = 0.4$, $\alpha_1 = 0.2$ - Left=FPR, Right=TPR

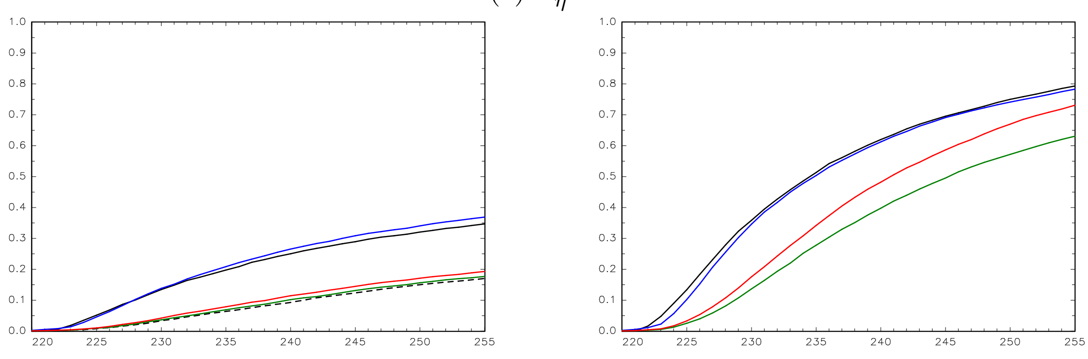
(a) $\sigma_\eta^2 = 0$



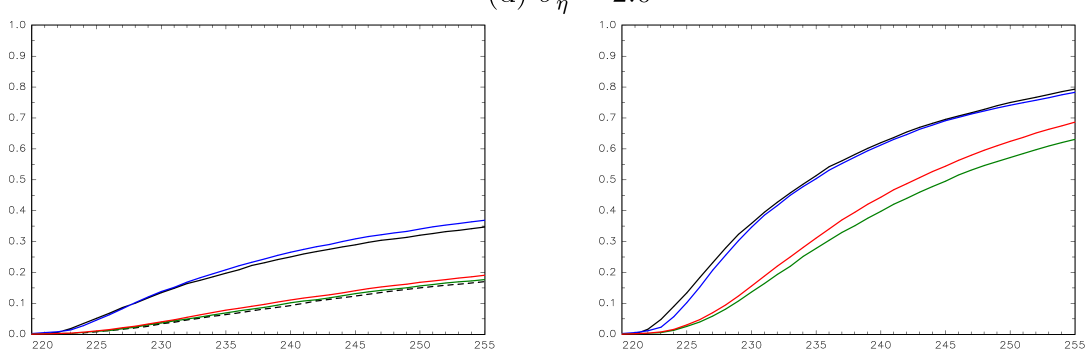
(b) $\sigma_\eta^2 = 0.5$



(c) $\sigma_\eta^2 = 1.0$



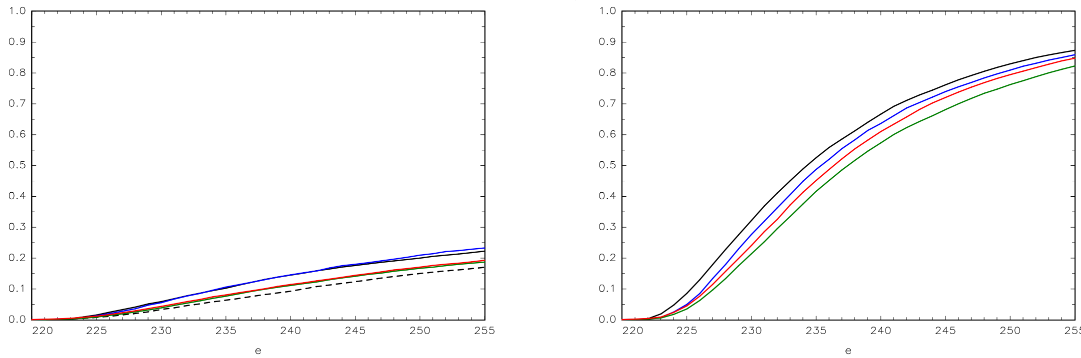
(d) $\sigma_\eta^2 = 2.0$



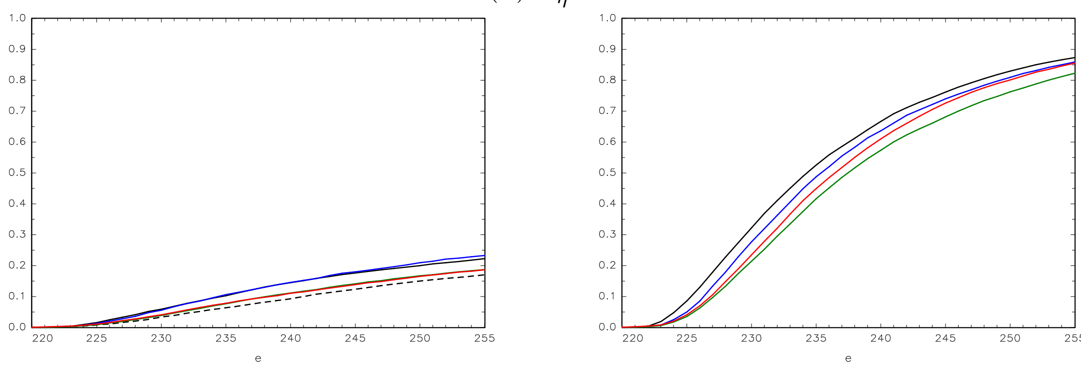
$\text{FPR}_{\text{i.i.d.}}$:– –, CUSUM:——, CUSUM^{V}:——, CUSUM^{V*}:——, CUSUM^{WMV}:——

Figure A.22: Covariate observed with error:$\beta = -0.5$, $\rho = 0.5$,$\sigma_{12} = 0.4$,$\alpha_1 = 0.2$ - Left=FPR, Right=TPR

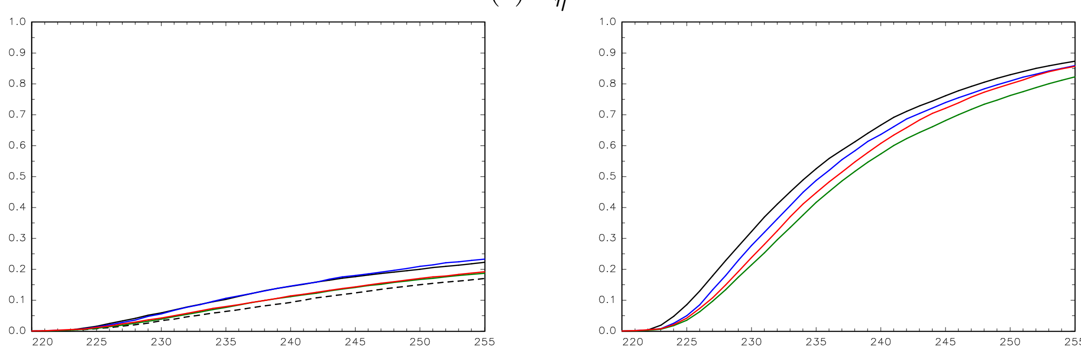
(a) $\sigma_\eta^2 = 0$



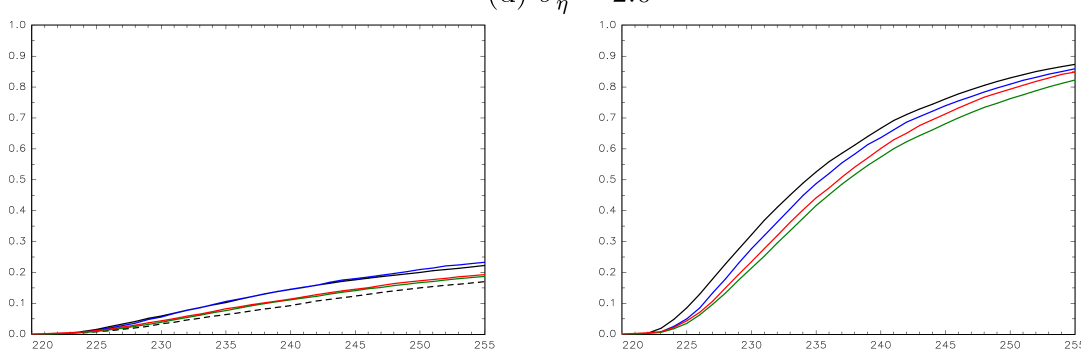
(b) $\sigma_\eta^2 = 0.5$



(c) $\sigma_\eta^2 = 1.0$



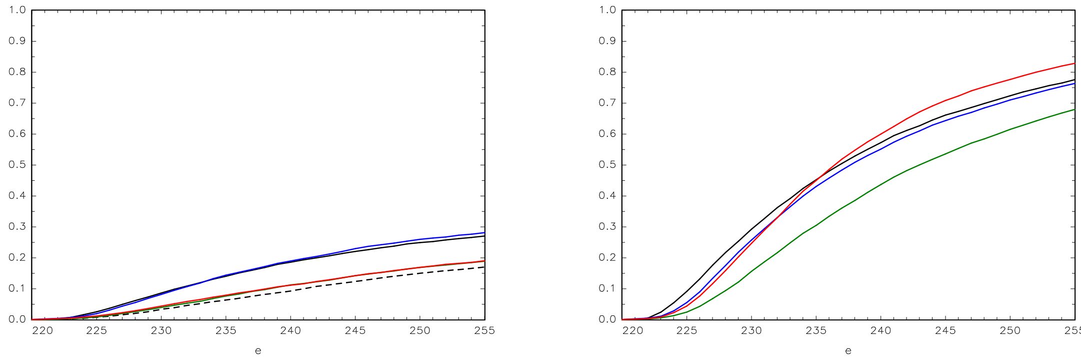
(d) $\sigma_\eta^2 = 2.0$



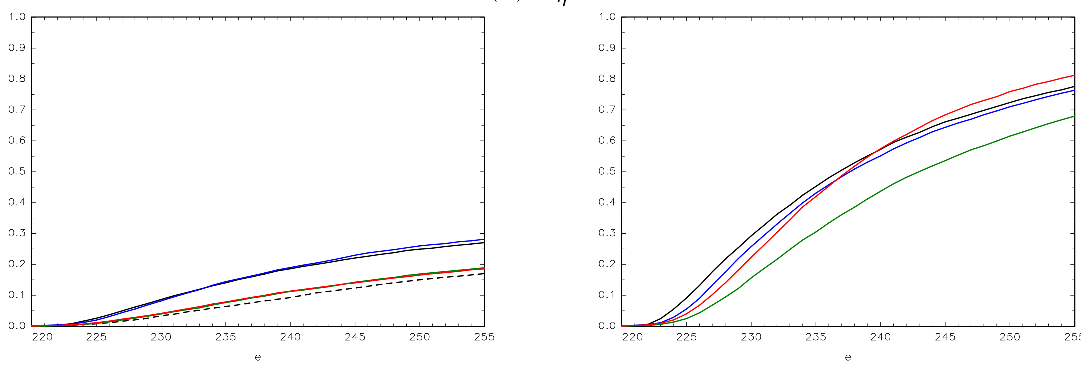
$\text{FPR}_{\text{i.i.d.}}$:– –, CUSUM:——, CUSUM^{V}:——, CUSUM^{V*}:——, CUSUM^{WMV}:——

Figure A.23: Covariate observed with error:$\beta = -0.8$,$\rho = 0.5$,$\sigma_{12} = 0.4$,$\alpha_1 = 0.2$ - Left=FPR, Right=TPR

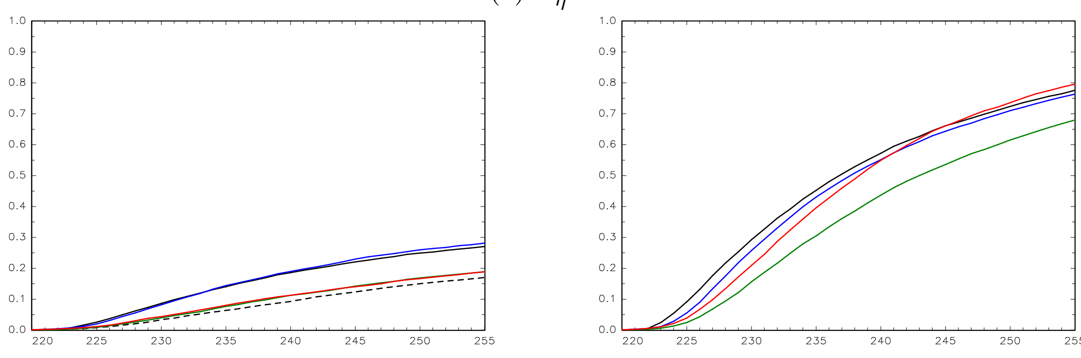
(a) $\sigma_\eta^2 = 0$



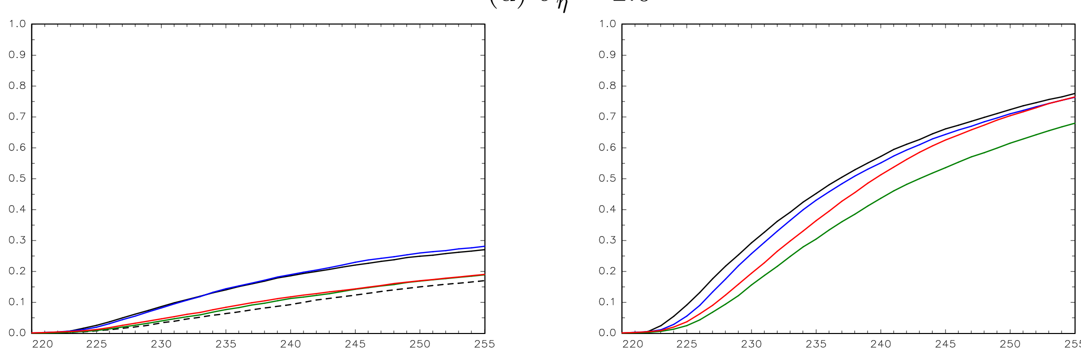
(b) $\sigma_\eta^2 = 0.5$



(c) $\sigma_\eta^2 = 1.0$



(d) $\sigma_\eta^2 = 2.0$



$\text{FPR}_{\text{i.i.d.}}$:– –, CUSUM:——, CUSUM^{V}:——, CUSUM^{V*}:——, CUSUM^{WMV}:——

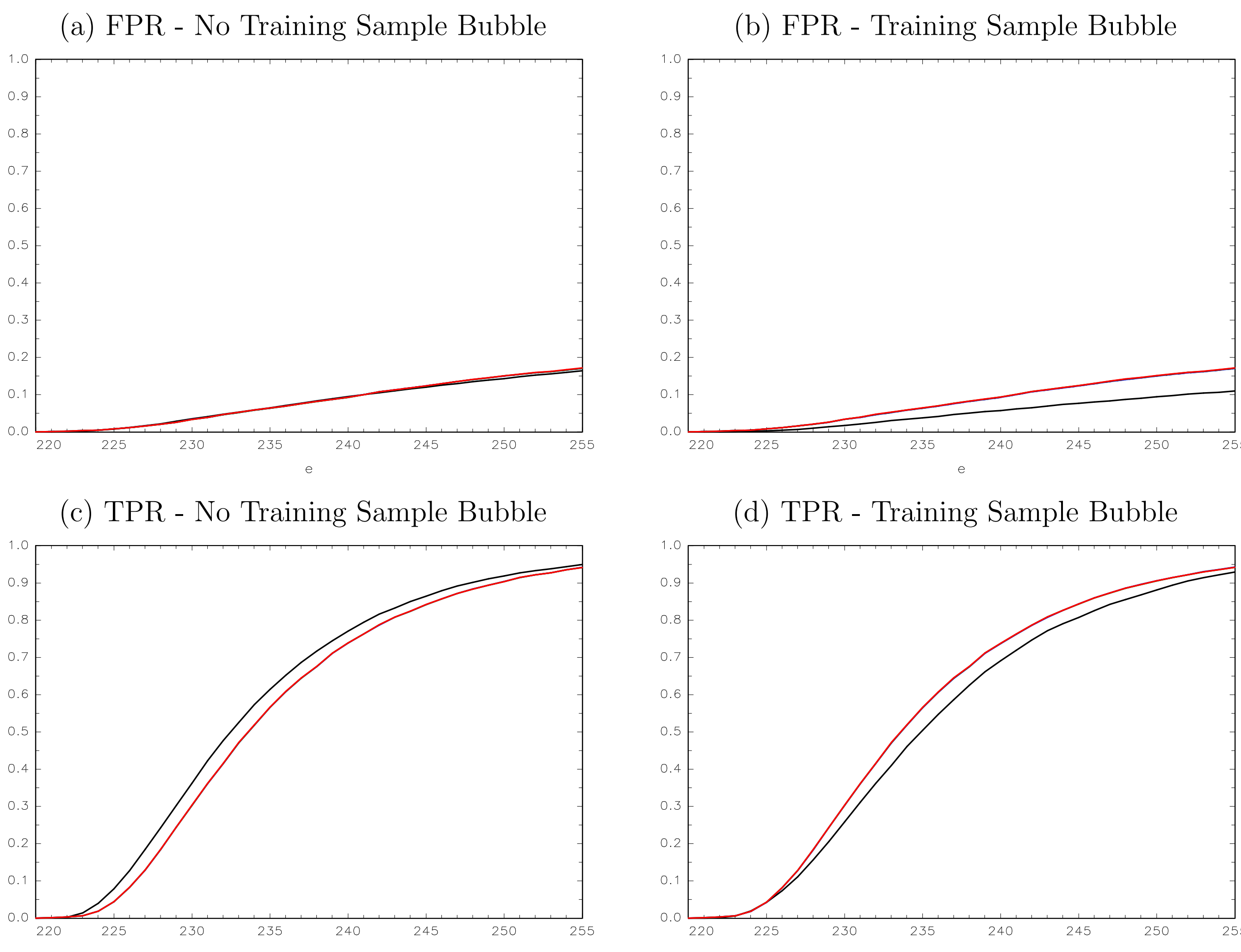
A.4.4 Additional Simulations - Training Sample Bubble

Figures A.24-A.32 report the FPR and TPR of the monitoring procedures when a single collapsed bubble is present in the training sample, with this training sample bubble running from $t = 96$ to $t = 110$ with an explosive offset of 0.005. Data were generated according to (1) with u_t generated as

$$u_t = \begin{cases} u_{t-1} + v_t & t = 1, \ldots, 95 \\ 1.005u_{t-1} + v_t & t = 96, \ldots, 110 \\ u_{111} = u_{95} + v_{111} & \\ u_{t-1} + v_t & t = 112, \ldots, T \\ u_{t-1} + v_t & t = T + 1, \ldots, \lfloor \tau_1 T \rfloor \\ (1 + \delta)u_{t-1} + v_t & t = \lfloor \tau_1 T \rfloor + 1, \ldots, \lfloor \tau_2 T \rfloor \\ u_{t-1} + v_t & t = \lfloor \tau_2 T \rfloor + 1, \ldots, \lfloor \lambda T \rfloor \end{cases} \tag{A.28}$$

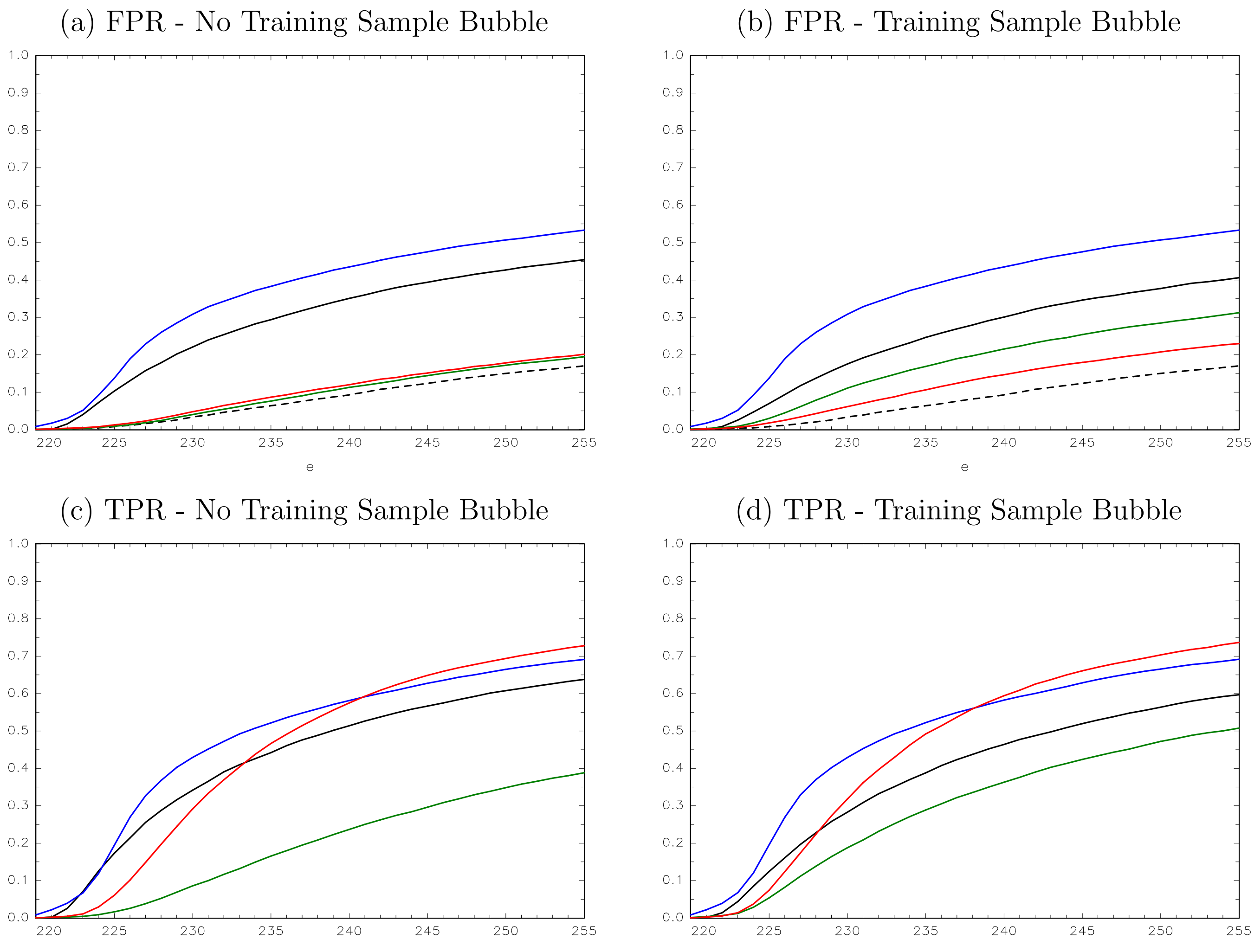
We generated v_t and x_t according to (23)-(25), with $\sigma^2_{1,t} = \sigma^2_{2,t} = 1$, $\sigma_{12,t} = \sigma_{12}$, and we again set $u_0 = 100$, $x_0 = \mu = 0$. Under the null we set $\delta = 0$, whereas under the alternative we set $\delta = 0.005$, $\tau_1 T = 220$ and $\tau_2 T = \lambda T$. When $\beta = 0$ and $\rho = 0$, so that the covariate is irrelevant, the past training sample bubble has minimal impact on the FPR and TPR of our proposed CUSUM^{WMV} procedure. This is because in almost all replications the BIC chooses to include no dynamics at all in the pre-whitening regression for y_t so that the CUSUM^{WMV} and CUSUM^{V*} procedures reduce to the CUSUM^{V} procedure which, when using a maximum bandwidth of 20 for the volatility estimator, will never use any of the observations associated with this past bubble. In all cases where the covariate is relevant, the past bubble causes a slight inflation of the FPR of the CUSUM^{WMV} and CUSUM^{V*} procedures relative to the case where no past bubble is present. This also leads to a slight increase in the TPR of the procedures when a past bubble is present relative to the case where no past bubble is present.

Figure A.24: FPR and TPR - Training Sample Bubble. $\beta = 0.0 = \rho = \sigma_{12} = \alpha_1 = 0.0$



$\text{FPR}_{\text{i.i.d.}}$:– –, CUSUM:——, CUSUM^{V}:——, CUSUM^{V*}:——, CUSUM^{WMV}:——

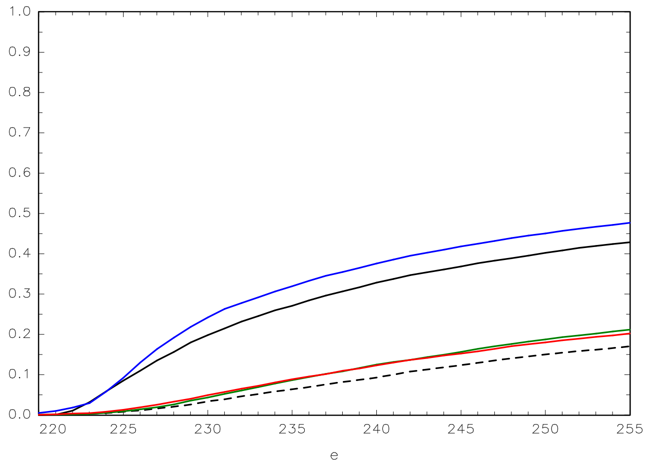
Figure A.25: FPR and TPR - Training Sample Bubble. $\beta = 0.8$, $\rho = 0.8$, $\sigma_{12} = 0.4$, $\alpha_1 = 0.2$



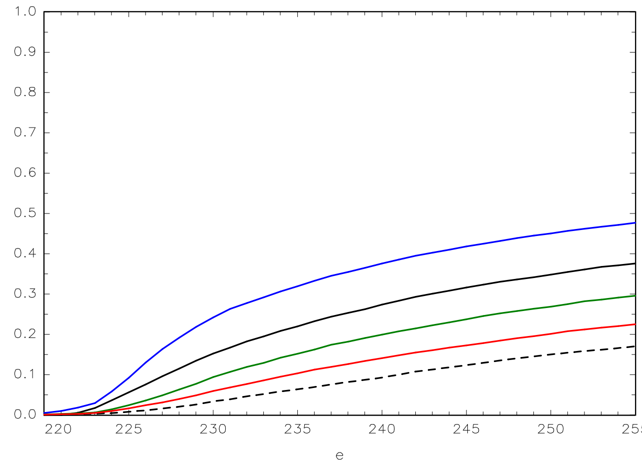
$\text{FPR}_{\text{i.i.d.}}$:– –, CUSUM:——, CUSUM^{V}:——, CUSUM^{V*}:——, CUSUM^{WMV}:——

Figure A.26: FPR and TPR - Training Sample Bubble. $\beta = 0.5$, $\rho = 0.8$, $\sigma_{12} = 0.4$, $\alpha_1 = 0.2$

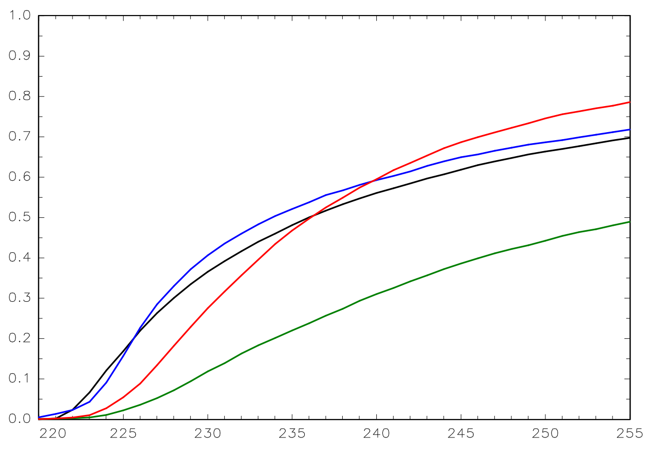
(a) FPR - No Training Sample Bubble



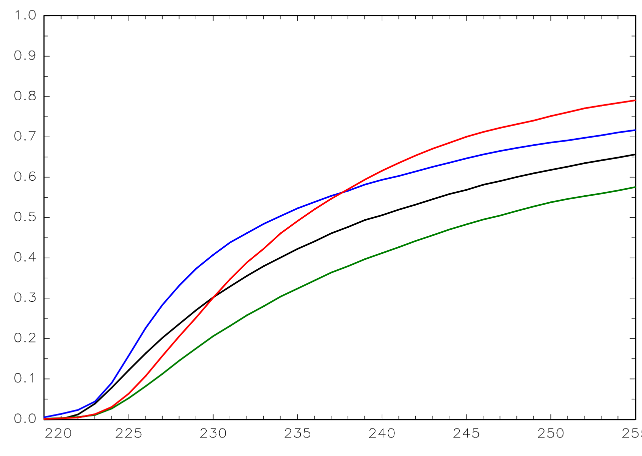
(b) FPR - Training Sample Bubble



(c) TPR - No Training Sample Bubble



(d) TPR - Training Sample Bubble



$\text{FPR}_{\text{i.i.d.}}$:– –, CUSUM:——, CUSUM^{V}:——, CUSUM^{V*}:——, CUSUM^{WMV}:——

Figure A.27: FPR and TPR - Training Sample Bubble. $\beta = -0.5$, $\rho = 0.8$, $\sigma_{12} = 0.4$, $\alpha_1 = 0.2$

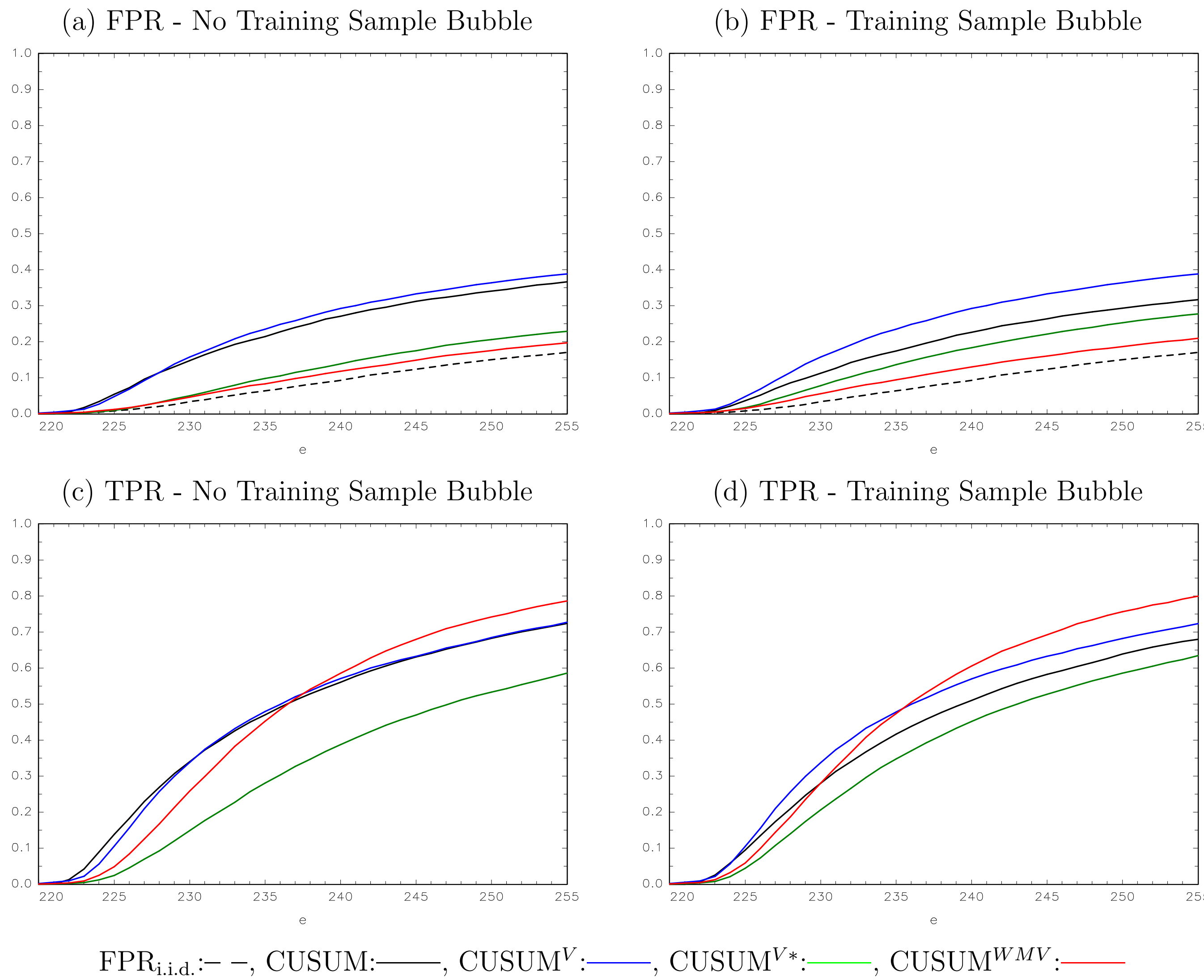
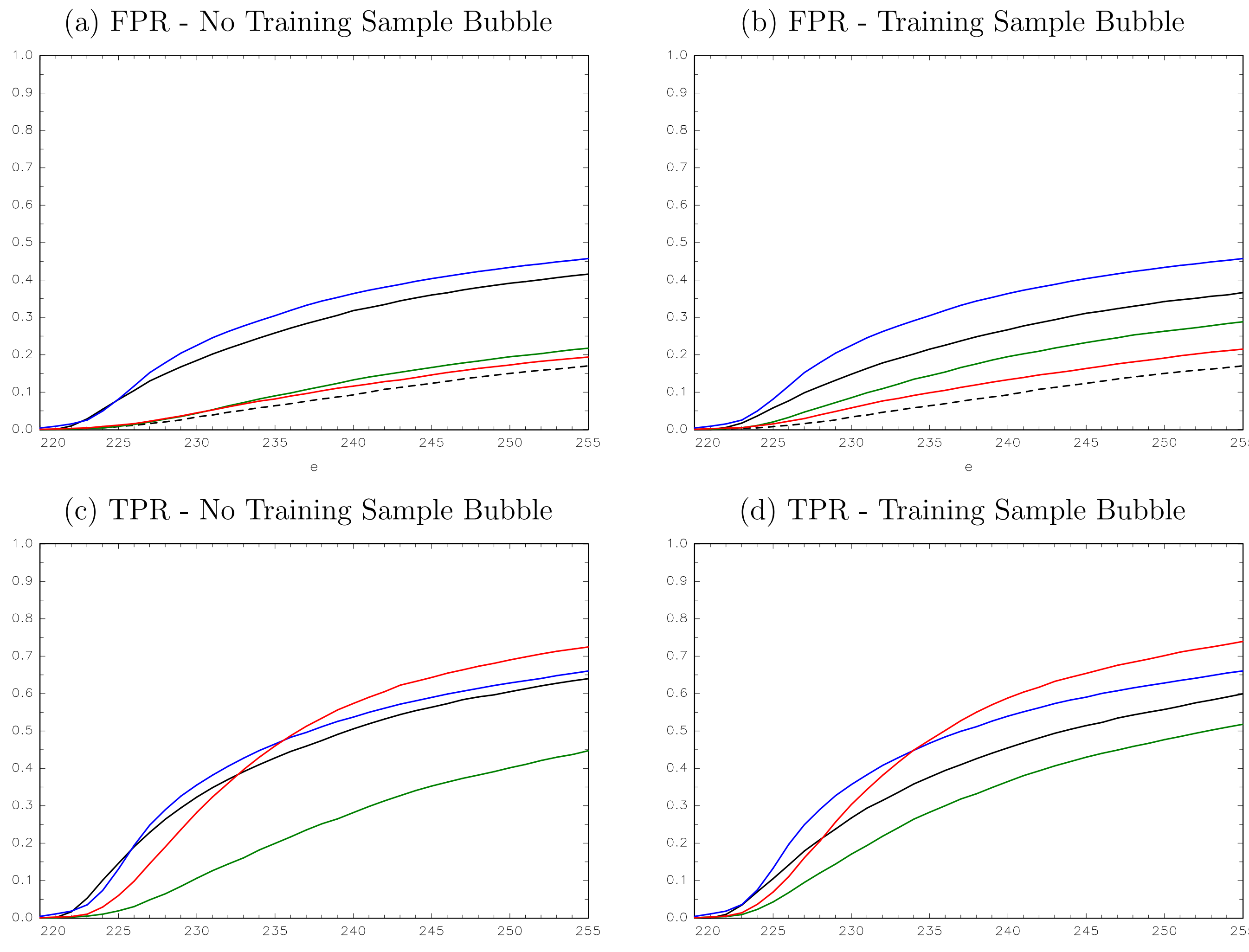
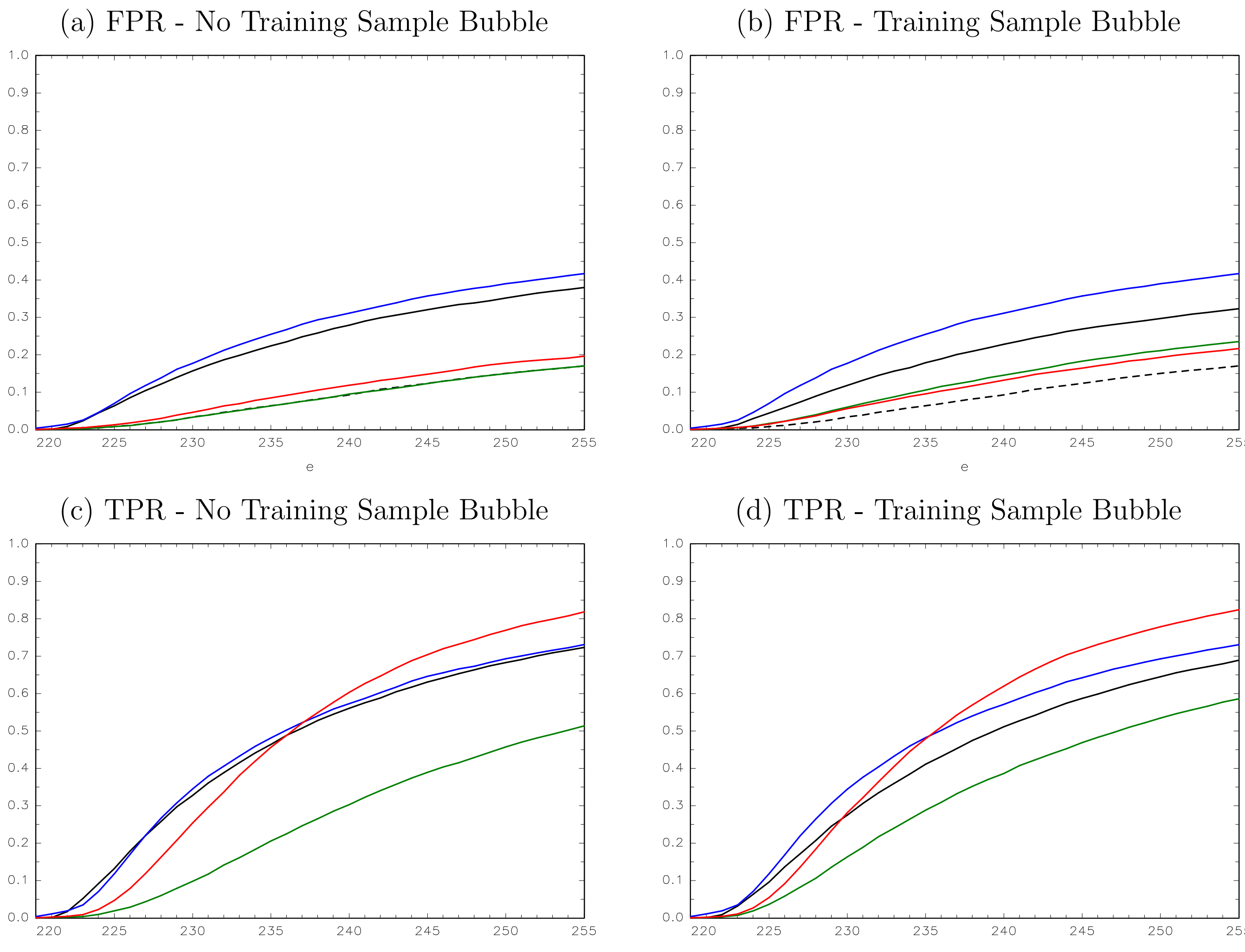


Figure A.28: FPR and TPR - Training Sample Bubble. $\beta = -0.8$, $\rho = 0.8$, $\sigma_{12} = 0.4$, $\alpha_1 = 0.2$



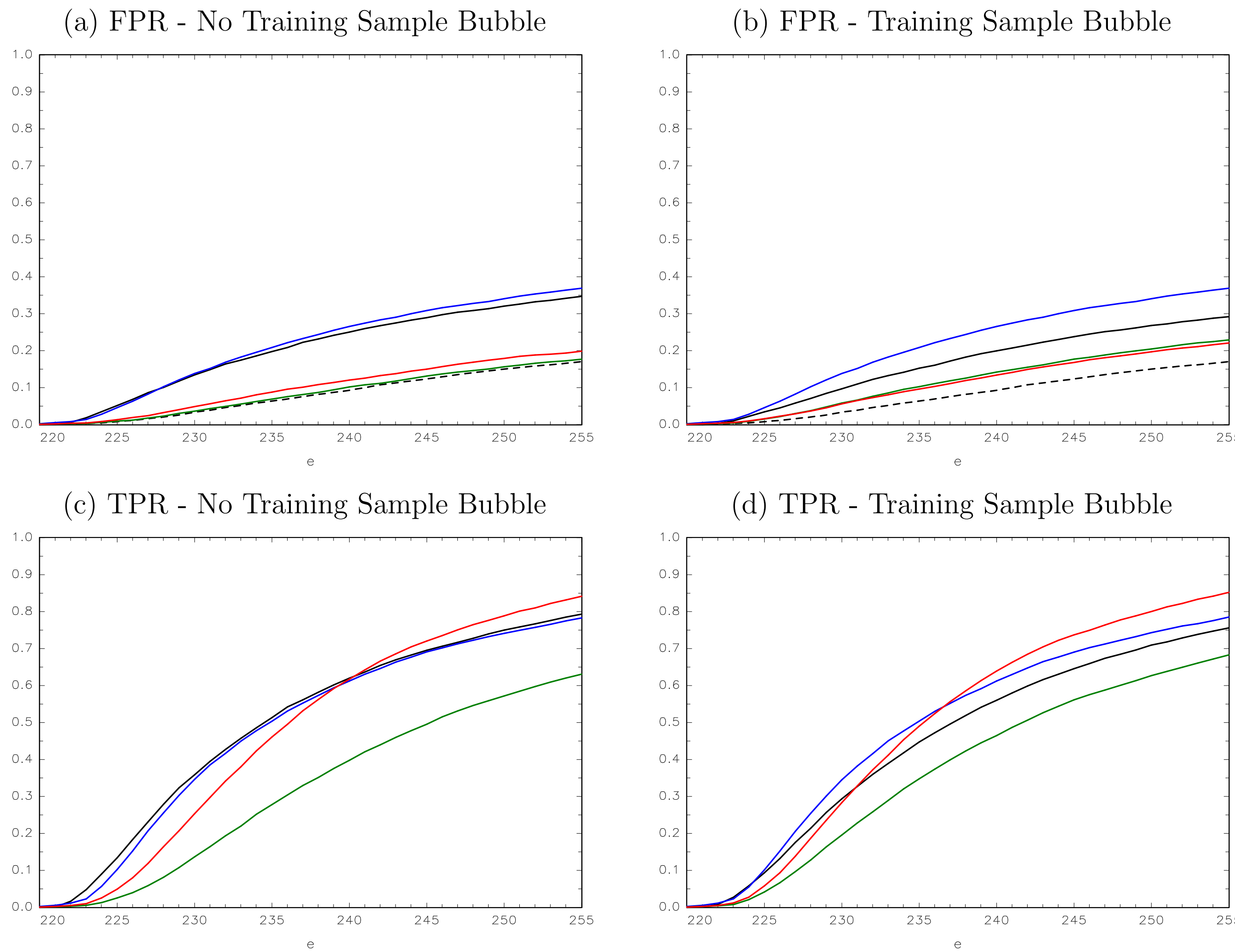
$FPR_{i.i.d.}$:− −, CUSUM:——, $CUSUM^{V}$:——, $CUSUM^{V*}$:——, $CUSUM^{WMV}$:——

Figure A.29: FPR and TPR - Training Sample Bubble. $\beta = 0.8$, $\rho = 0.5$, $\sigma_{12} = 0.4$, $\alpha_1 = 0.2$



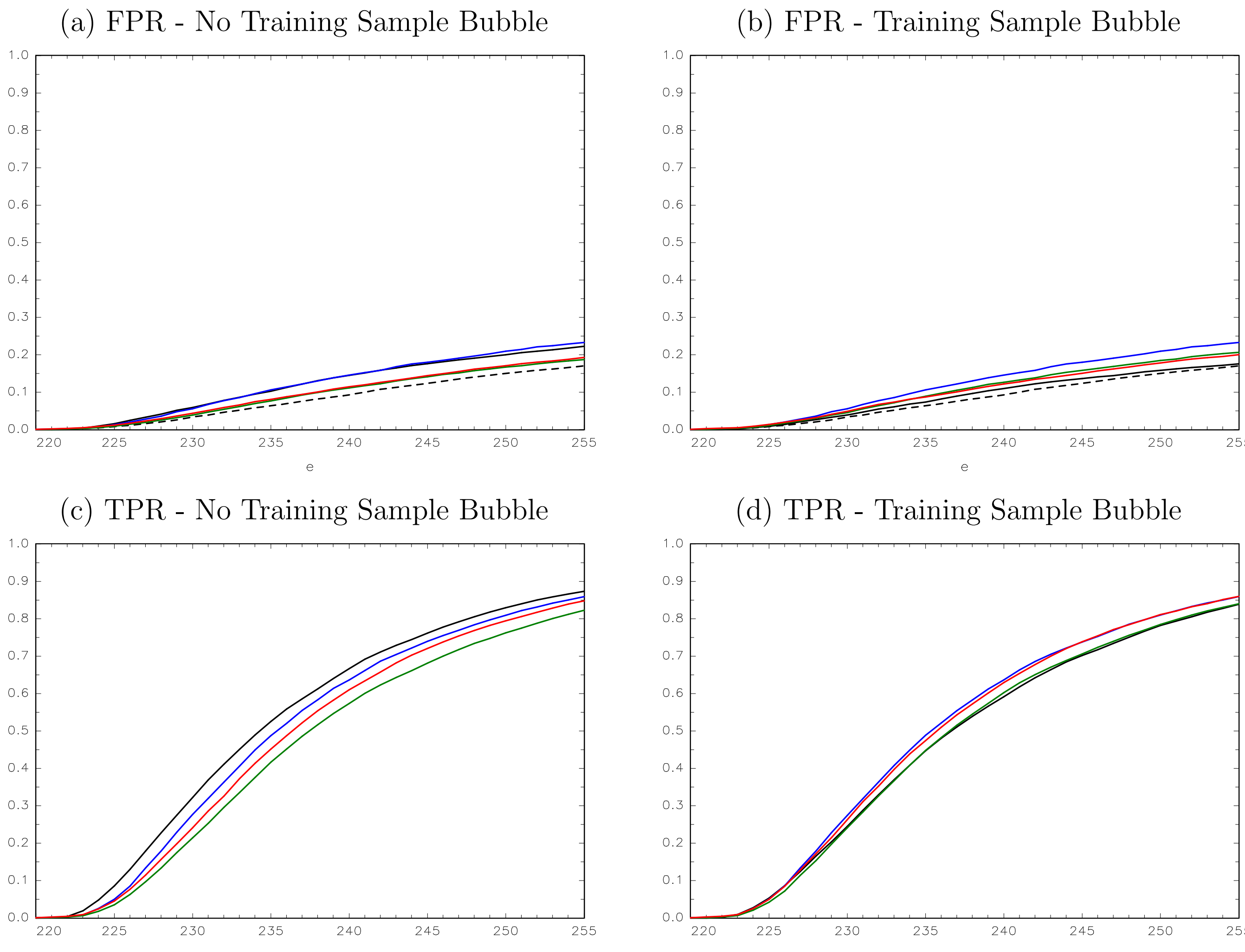
$\text{FPR}_{\text{i.i.d.}}$:− −, CUSUM:——, CUSUM^{V}:——, CUSUM^{V*}:——, CUSUM^{WMV}:——

Figure A.30: FPR and TPR - Training Sample Bubble. $\beta = 0.5$, $\rho = 0.5$, $\sigma_{12} = 0.4$, $\alpha_1 = 0.2$



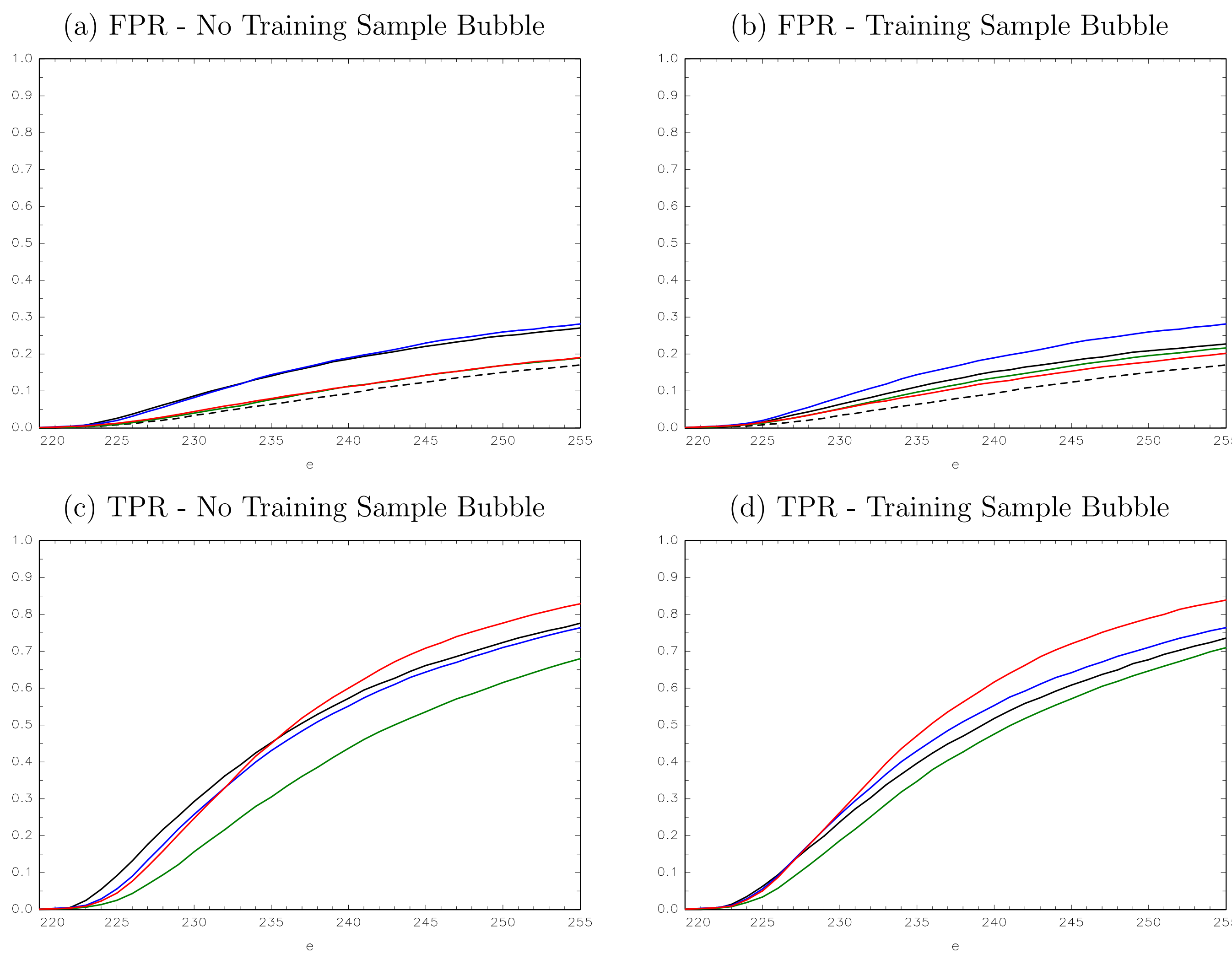
$\text{FPR}_{\text{i.i.d.}}$:– –, CUSUM:——, CUSUM^{V}:——, CUSUM^{V*}:——, CUSUM^{WMV}:——

Figure A.31: FPR and TPR - Training Sample. $\beta = -0.5$, $\rho = 0.5$, $\sigma_{12} = 0.4$, $\alpha_1 = 0.2$



$FPR_{i.i.d.}$:– –, CUSUM:——, $CUSUM^V$:——, $CUSUM^{V*}$:——, $CUSUM^{WMV}$:——

Figure A.32: FPR and TPR - Training Sample. $\beta = -0.8$, $\rho = 0.5$, $\sigma_{12} = 0.4$, $\alpha_1 = 0.2$

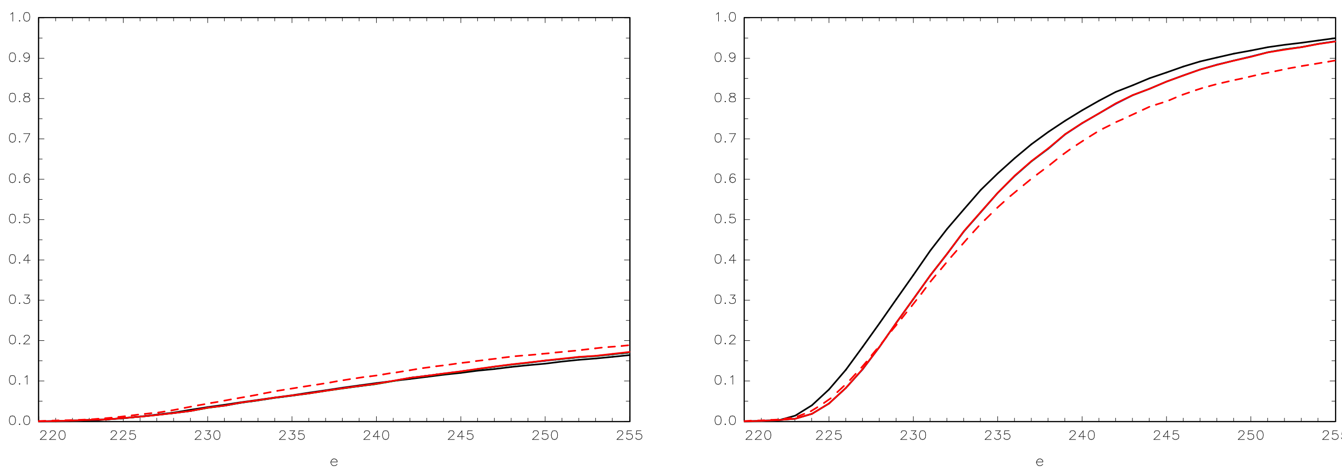


$\text{FPR}_{\text{i.i.d.}}$:– –, CUSUM:——, CUSUM^{V}:——, CUSUM^{V*}:——, CUSUM^{WMV}:——

A.4.5 Additional Simulations - $I(1)$ Covariate

Figure A.33 reports the FPR and TPR of the procedures when an irrelevant $I(1)$ covariate is considered for inclusion in the CUSUM^{WMV} procedure. Data were therefore generated according to (1)-(2) and (23)-(24) with $\beta = \rho = \sigma_{12} = \alpha_1 = 0$, $\rho = 1$ and $\sigma^2_{1,t} = \sigma^2_{2,t} = 1$ $\forall t$. In a vast majority of replications the BIC model selection procedure correctly determines the covariate to be irrelevant and so there is almost no effect on the FPR and TPR of the CUSUM^{WMV} procedure. Consequently, to get a better idea of the impact of including this irrelevant $I(1)$ covariate we include a line on the figures for the case where the covariate is forcibly included in the CUSUM^{WMV} procedure. We see that including this covariate leads to a slight increase in FPR under the null and a moderate decrease in TPR under the alternative, compared to the correctly specified univariate procedures which have an FPR profile identical to that seen in Figure 1(a) in the main paper.

Figure A.33: $\beta = 0.0$, $\rho = 1.0$, $\sigma_{12} = 0.0$, $\alpha_1 = 0.0$ - Left Panel=FPR, Right Panel =TPR. ($I(1)$ Covariate)



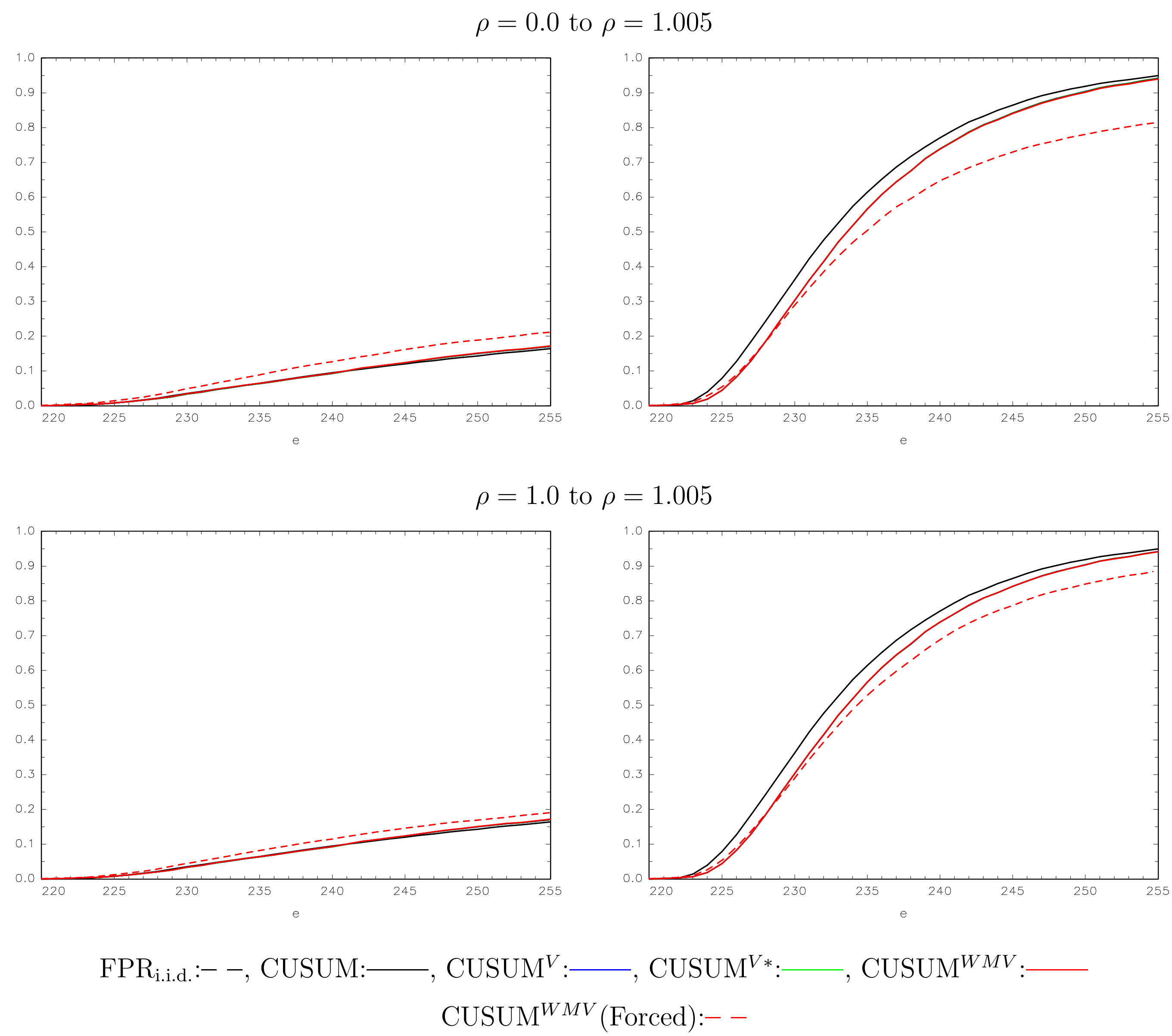
FPR$_{\text{i.i.d.}}$:– –, CUSUM:———, CUSUM^{V}:———, CUSUM^{V*}:———, CUSUM^{WMV}:———

CUSUM^{WMV}(Forced):– –

A.4.6 Additional Simulations - Bubble in Irrelevant Covariate

Figure A.34 reports the FPR and TPR of the procedures when an irrelevant covariate that exhibits explosive behaviour in the monitoring period is considered for inclusion in the CUSUM^{WMV} procedure. Data were therefore generated according to (1)-(2) and (23)-(24) with $\beta = \sigma_{12} = \alpha_1 = 0$, $\sigma^2_{1,t} = \sigma^2_{2,t} = 1 \ \forall t$, $\rho_t = \rho$, $\rho \in \{0.0, 1.0\}$ for $t = 1, ..., 220$ and $\rho_t = 1.005$ for $t = 221, ..., \lambda T$. The covariate therefore behaves as either an $I(0)$ process or an $I(1)$ process up until the start of monitoring, before exhibiting explosive behaviour until the end of the monitoring period. Again, in a majority of replications the BIC model selection procedure correctly determines the covariate to be irrelevant and so there is almost no effect on the FPR and TPR of the CUSUM^{WMV} procedure. To get a better idea of the impact of including this explosive covariate we also include a line on the figures for the case where the covariate is forcibly included in the CUSUM^{WMV} procedure. We see that always including this covariate leads to a slight increase in the FPR of CUSUM^{WMV} under the null and a modest decrease in the TPR under the alternative, compared to the correctly specified univariate procedures, which have an FPR profile identical to that seen in Figure 1(a) in the main paper, with these effects seen to be more pronounced for $\rho = 0.0$ than for $\rho = 1.0$.

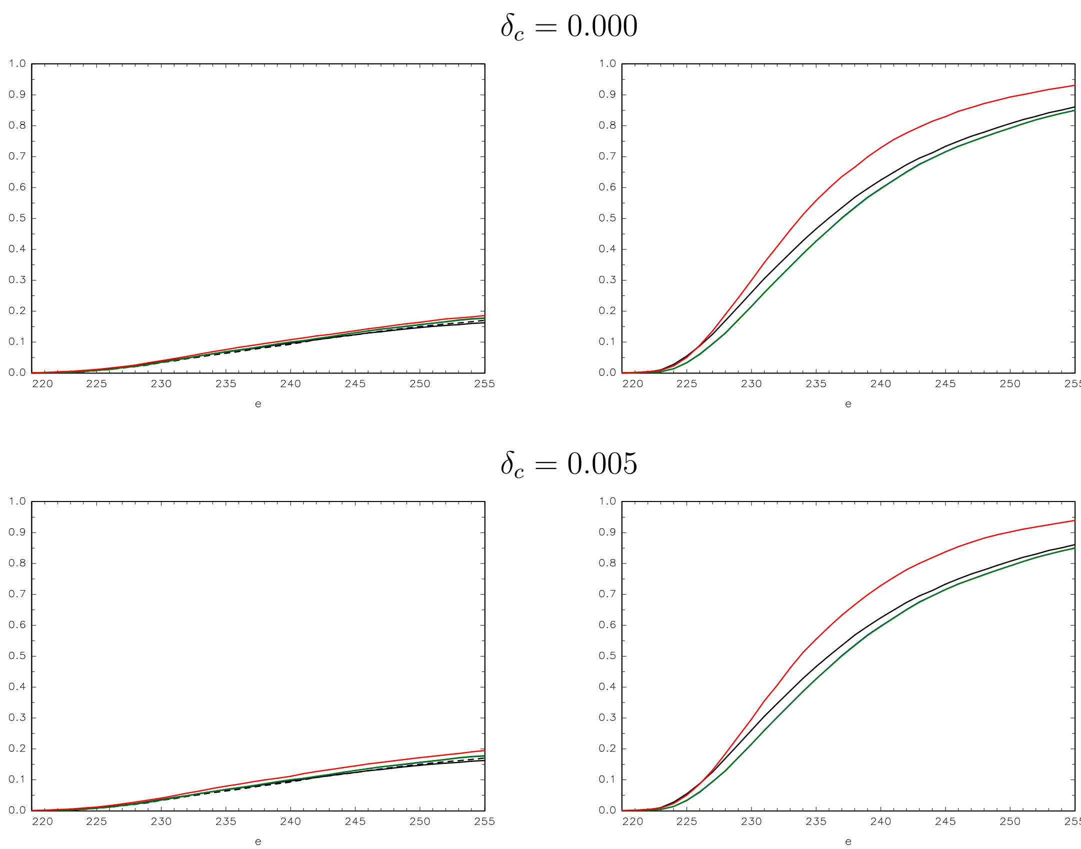
Figure A.34: $\beta = 0.0$, $\sigma_{12} = 0.0$, $\alpha_1 = 0.0$ - Left Panel=FPR, Right Panel =TPR. Covariate Bubble in Monitoring Period



A.4.7 Additional Simulations - Bubble in Relevant Covariate

Figure A.35 reports the FPR and TPR of the procedures when a bubble may be present in a relevant covariate considered for use in the CUSUM^{WMV} procedure. Data were therefore generated according to (1)-(2) and (23)-(24) with $\beta = 0.8$, $\rho = \sigma_{12} = \alpha_1 = 0.0$ and $\sigma^2_{1,t} = \sigma^2_{2,t} = 1$ $\forall t$. We generate $z_t = z_{t-1} + x_t$, $t = 1, ..., \lfloor \tau T \rfloor$ and $z_t = (1 + \delta_c)z_{t-1} + x_t$, $t = \lfloor \tau T \rfloor + 1, ..., \lfloor \lambda T \rfloor$, again setting $\lfloor \tau T \rfloor = 220$, and use Δz_t as a covariate in the CUSUM^{WMV} procedure. Results are reported for $\delta_c \in \{0.000, 0.005\}$. We see that for $\delta_c = 0$ the FPR and TPR of the procedures is identical to those reported in Figure 2 panel(a), as would be expected because in this case $\Delta z_t = x_t$. For $\delta_c = 0.005$, such that the covariate is contaminated by explosivity during the monitoring period, we see that the FPR of the CUSUM^{WMV} procedure is very slightly inflated, relative to the case where $\delta_c = 0$. Under the alternative, the TPR of CUSUM^{WMV} when $\delta_c = 0.005$ is essentially the same as when $\delta_c = 0$.

Figure A.35: $\beta = 0.8$, $\rho = \sigma_{12} = \alpha_1 = 0.0$ - Left Panel=FPR, Right Panel =TPR. Impact of Explosivity in Covariate



$\text{FPR}_{\text{i.i.d.}}$:− −, CUSUM:——, CUSUM^{V}:——, CUSUM^{V*}:——, CUSUM^{WMV}:——

A.4.8 Additional Simulations - Mean Shift In Covariate

In this section we examine the FPR and TPR of the monitoring procedures when a covariate that includes a mean shift in the monitoring period is considered for inclusion in the CUSUM^{WMV} procedure. For reference we also report the cumulative rejection rate of a univariate monitoring procedure, which will be denoted CUSUM^{X}, that is designed to test for structural change in the covariate over the same monitoring period as CUSUM^{WMV}. This procedure is entirely analogous to the one-sided CUSUM^{WMV} procedure except that a two-tailed decision rule is used (to allow for positive or negative mean shifts), and the regressions used to construct the statistics underlying the CUSUM^{X} procedure are with-constant autoregressions fitted to x_t, with the autoregressive lag order determined by the BIC. The CUSUM^{X} procedure is calibrated so that it has a two-sided FPR of 20% at the same time that the one-sided bubble CUSUM procedures are calibrated to have an FPR of 10%. In practice the significance level used with CUSUM^{X} can of course be varied by the practitioner.

Given that in empirical applications candidate macroeconomic and financial covariates will generally be entered in first differences rather than levels (see Remark 2.8), we consider first what is arguably the empirically most relevant case where a mean shift occurs in a series that is used as a covariate in first differences in the CUSUM^{WMV} procedure. Data were therefore generated according to (1)-(2) and (23)-(24) with $\beta \in \{0.0, 0.5\}$, $\rho = \sigma_{12} = \alpha_1 = 0$, $\sigma^2_{1,t} = \sigma^2_{2,t} = 1 \ \forall t$. We then generate $z_t = z_{t-1} + x_t$ and add a 0.5 or 1.0 standard deviation magnitude mean shift to z_t at time $t = 230$. The first differenced series Δz_t is then used as a covariate in the CUSUM^{WMV} procedure.

We first consider the case $\beta = 0.0$, such that the covariate is irrelevant. Once again, in the vast majority of simulation replications the BIC model selection device correctly determines the covariate to be irrelevant and so there is almost no effect on the FPR and TPR of the CUSUM^{WMV} procedure. Consequently, in order to get a better idea of the impact of including this covariate we again include a line on the figures for the case where the covariate is forcibly included in the CUSUM^{WMV} procedure. Examining the results in Figures A.36-A.37 we see that, regardless of whether the mean shift is upwards or downwards, including this covariate leads to a marginal increase in FPR under the null and a marginal decrease in TPR under the alternative, in each case compared to the correctly specified univariate procedures which have an FPR profile identical to that seen in Figure 1(a) in the main paper. These effects are seen to be very small.

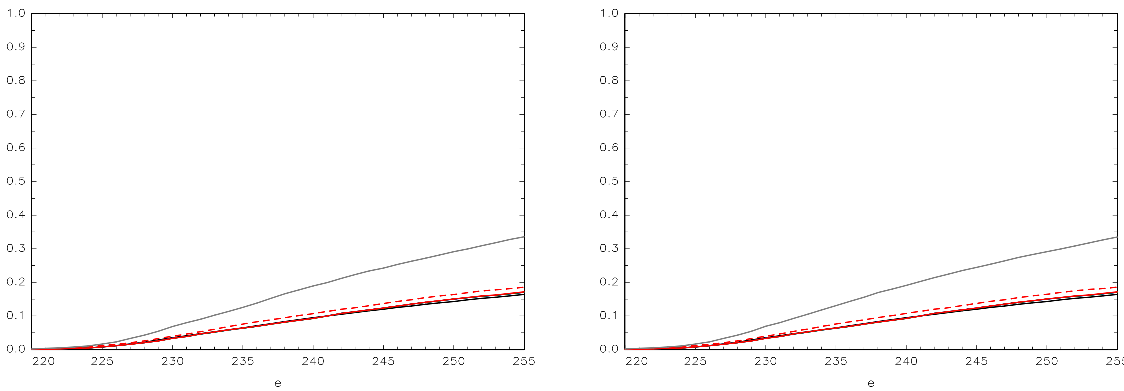
Moving to the case where $\beta = 0.5$, results for which are reported in Figures A.38-A.39, we see that the mean shift in z_t does lead to some slight FPR distortions in the CUSUM^{WMV} procedure, with upward mean shifts causing a slight decrease in FPR and downward mean shifts leading to a slight increase. The impact on TPR is similar, as is to be expected, with upward mean shifts causing a slight decrease in TPR and downward mean shifts leading to a slight increase. The effects are, however, relatively benign even for the case of a relatively large mean shift of one standard deviation.

We next turn to the case where the mean shift occurs in the candidate covariate, x_t, which enters the regression in levels. Data were therefore generated according to (1)-(2) and (23)-(24) with $\beta \in \{0.0, 0.5\}$, $\rho = \sigma_{12} = \alpha_1 = 0$, $\sigma^2_{1,t} = \sigma^2_{2,t} = 1$ $\forall t$ where we add a 0.5 or 1.0 standard deviation mean shift to x_t at time $t = 230$. We begin by examining results where $\beta = 0.0$ reported in Figures A.40-A.41, once again including a line on the figures for the case where the covariate is forcibly included in the CUSUM^{WMV} procedure, given that the covariate is irrelevant. We see that, once again, the mean shift makes almost no difference to the rejection rate of the CUSUM^{WMV} procedure.

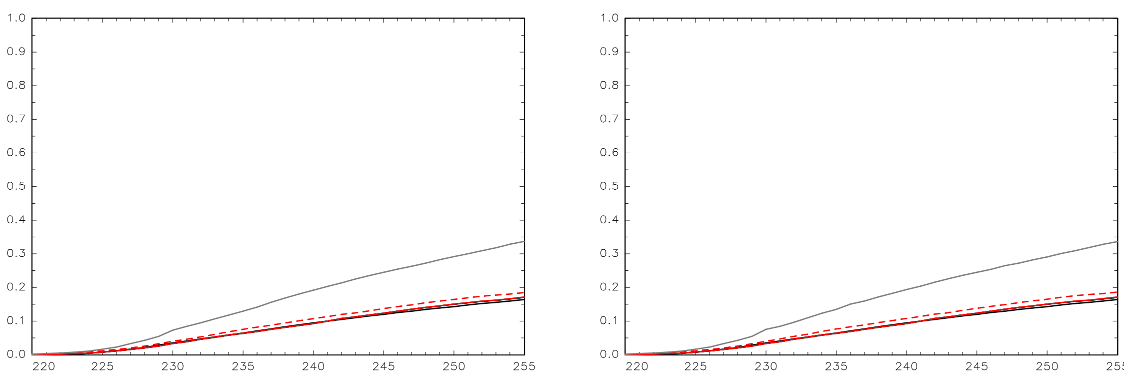
Results for the case where $\beta = 0.5$, reported in Figures A.42-A.43, are not so benign, with the mean shift in the utilised covariate causing large upward (downward) bias in the FPR of the CUSUM^{WMV} procedure for downward (upward) mean shifts, with a large downward bias in the TPR of the CUSUM^{WMV} procedure subsequently observed for upward mean shifts. Crucially, however, the empirical rejection frequency of the CUSUM^X procedure is considerably higher than that of the CUSUM^{WMV} procedure under the null of no bubble, even when the latter is subject to a large upward bias in FPR due to the presence of a negative mean shift. In the majority of cases, therefore, the simultaneous CUSUM^X procedure would alert the practitioner to the mean shift in the covariate and the practitioner would consequently revert to a covariate unaugmented procedure, for example CUSUM$^V*$.

Figure A.36: $\beta = 0.0$, $\rho = 0.0$, $\sigma_{12} = 0.0$, $\alpha_1 = 0.0$. FPR of Procedures. Mean Shift at $t = 230$ in an Irrelevant Covariate entered in First Differences. Left Panel = Upward Shift, Right Panel = Downward Shift.

Break Magnitude = 0.5σ



Break Magnitude = 1σ



$\text{FPR}_{\text{i.i.d.}}$:− −, CUSUM:——, CUSUM^{V}:——, CUSUM^{V*}:——, CUSUM^{WMV}:——

CUSUM^{WMV}(Forced):− −, CUSUM^{X}:——

Figure A.37: $\beta = 0.0$, $\rho = 0.0$, $\sigma_{12} = 0.0$, $\alpha_1 = 0.0$. TPR of Procedures. Mean Shift at $t = 230$ in an Irrelevant Covariate entered in First Differences. Left Panel = Upward Shift, Right Panel = Downward Shift.

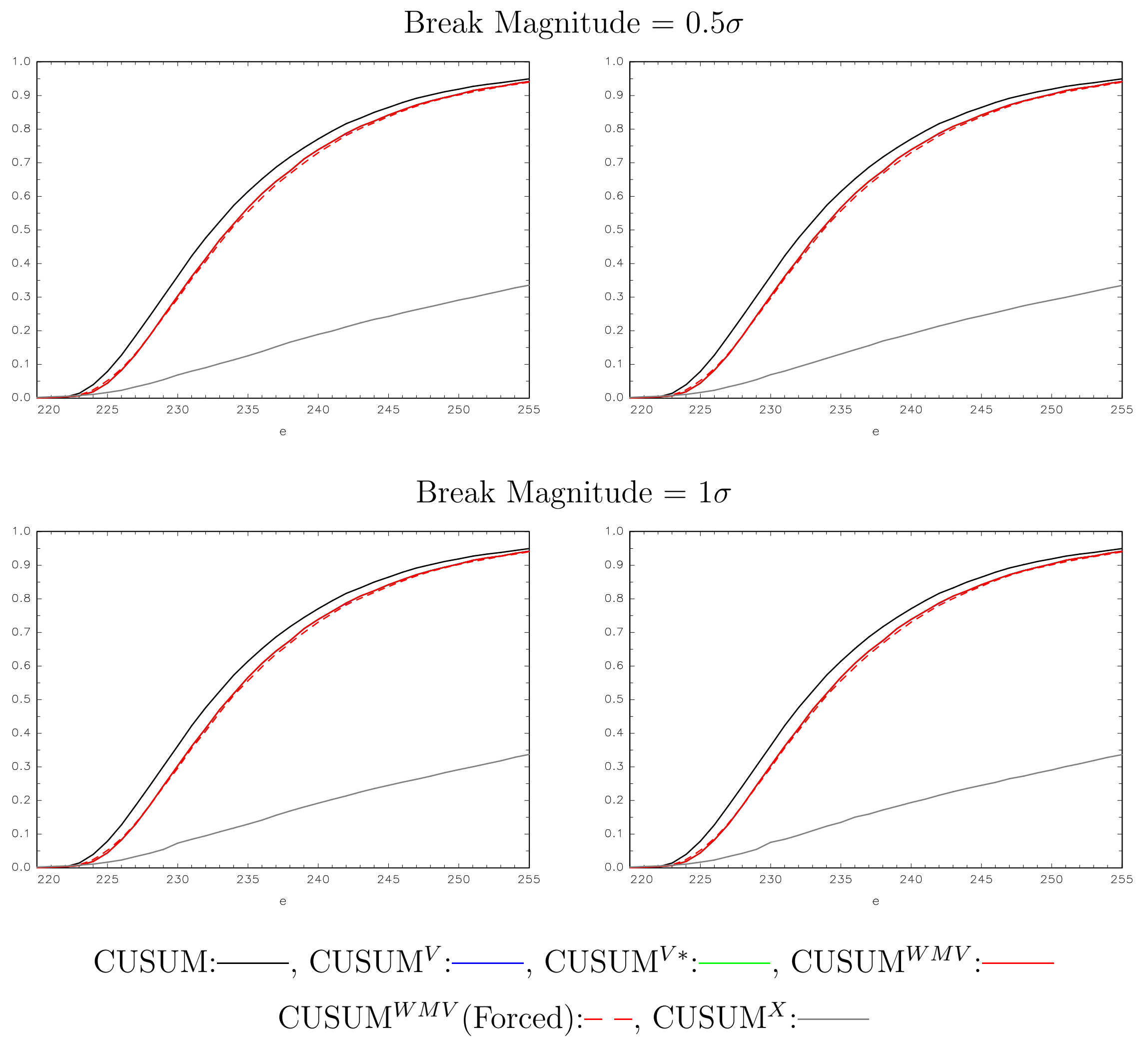


Figure A.38: $\beta = 0.5$, $\rho = 0.0$, $\sigma_{12} = 0.0$, $\alpha_1 = 0.0$. FPR of Procedures. Mean Shift at $t = 230$ in a Relevant Covariate entered in First Differences. Left Panel = Upward Shift, Right Panel = Downward Shift.

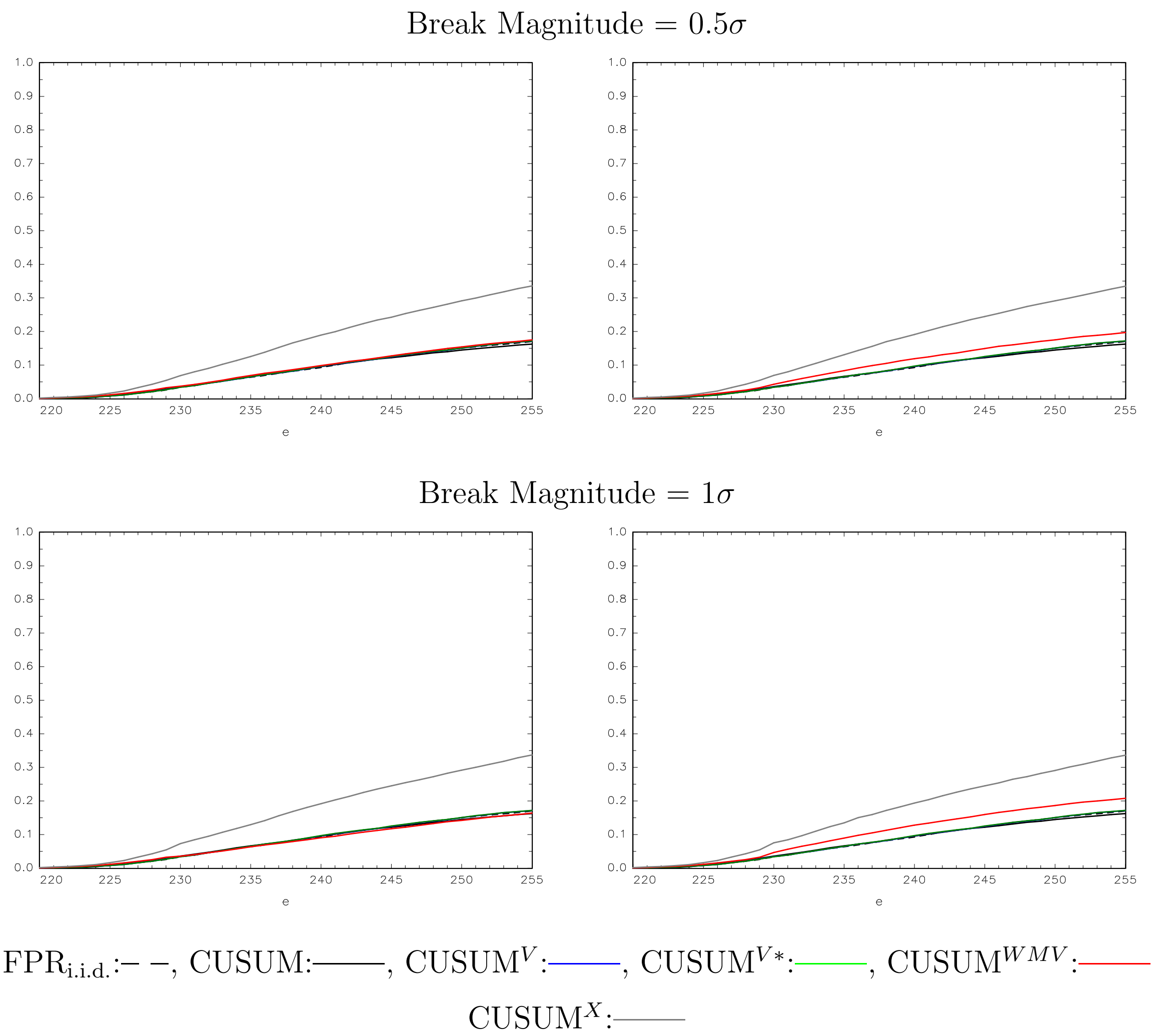
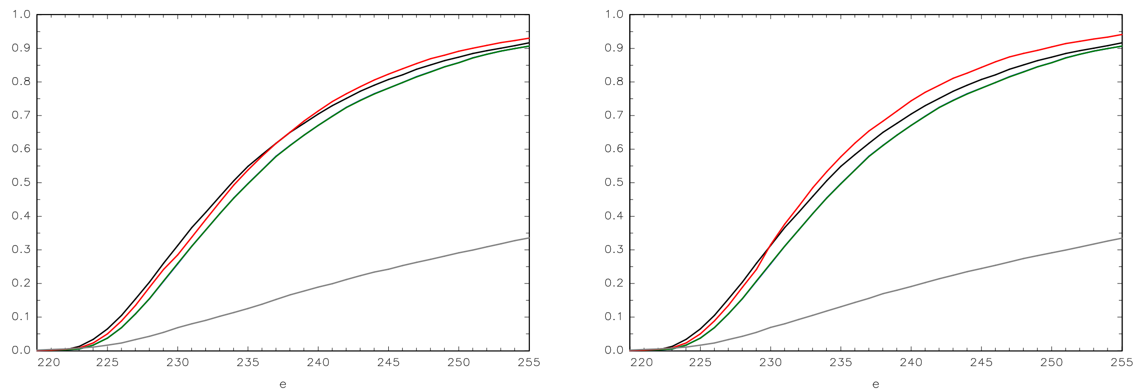
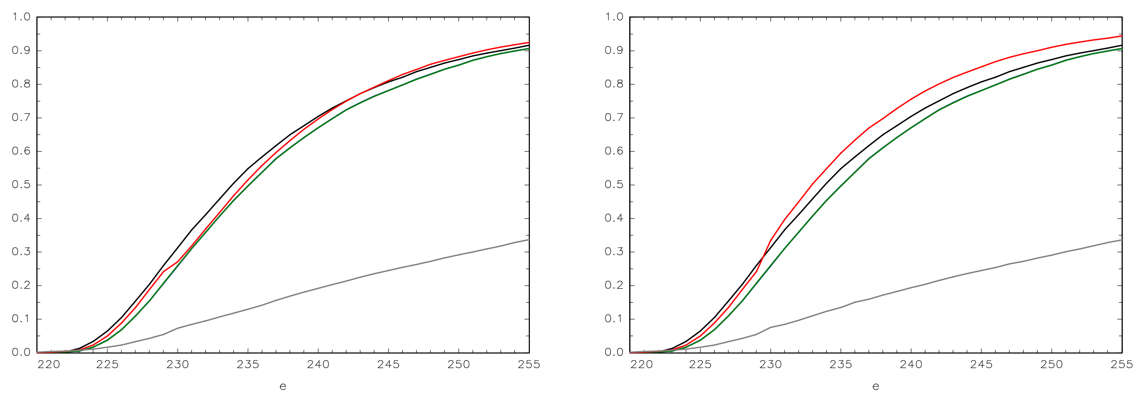


Figure A.39: $\beta = 0.5$, $\rho = 0.0$, $\sigma_{12} = 0.0$, $\alpha_1 = 0.0$. TPR of Procedures. Mean Shift at $t = 230$ in a Relevant Covariate entered in First Differences. Left Panel = Upward Shift, Right Panel = Downward Shift.

Break Magnitude = 0.5σ



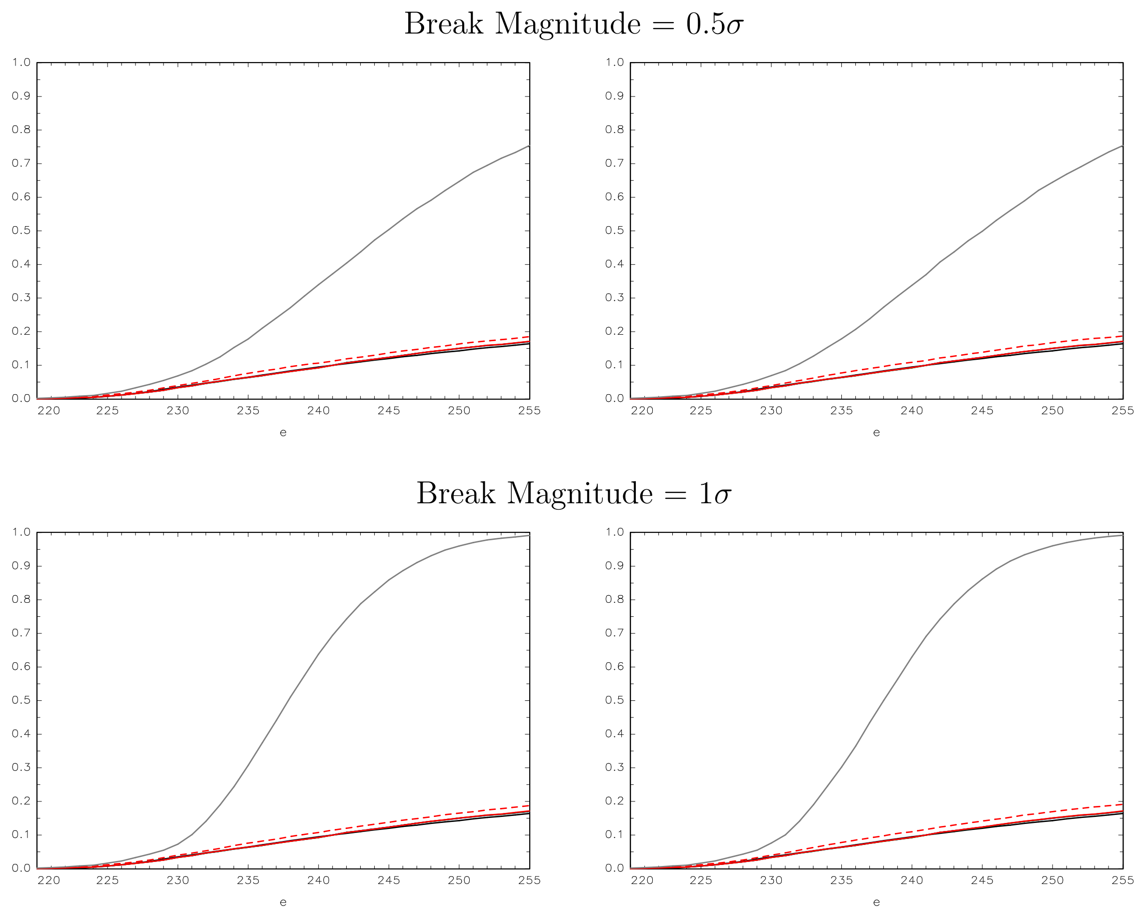
Break Magnitude = 1σ



CUSUM:——, CUSUM^{V}:——, CUSUM^{V*}:——, CUSUM^{WMV}:——

CUSUM^{X}:——

Figure A.40: $\beta = 0.0$, $\rho = 0.0$, $\sigma_{12} = 0.0$, $\alpha_1 = 0.0$. FPR of Procedures. Mean Shift at $t = 230$ in an Irrelevant Covariate entered in Levels. Left Panel = Upward Shift, Right Panel = Downward Shift.

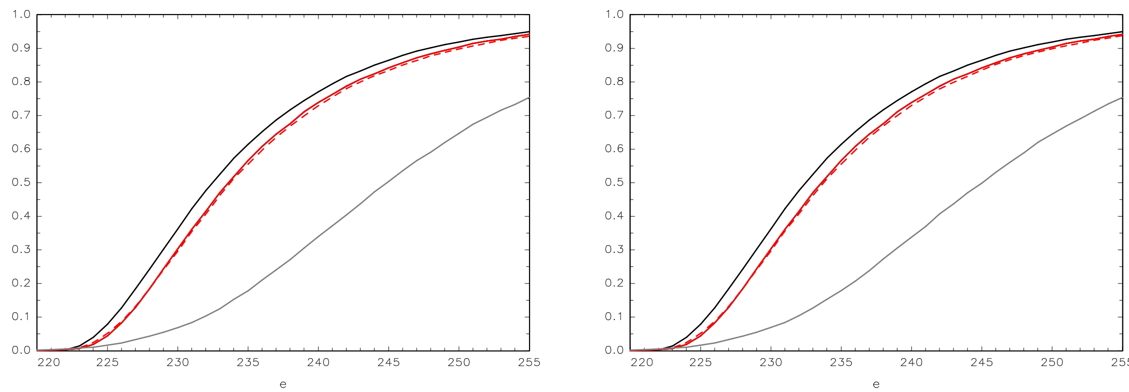


$\text{FPR}_{\text{i.i.d.}}$:– –, CUSUM:——, CUSUM^{V}:——, CUSUM^{V*}:——, CUSUM^{WMV}:——

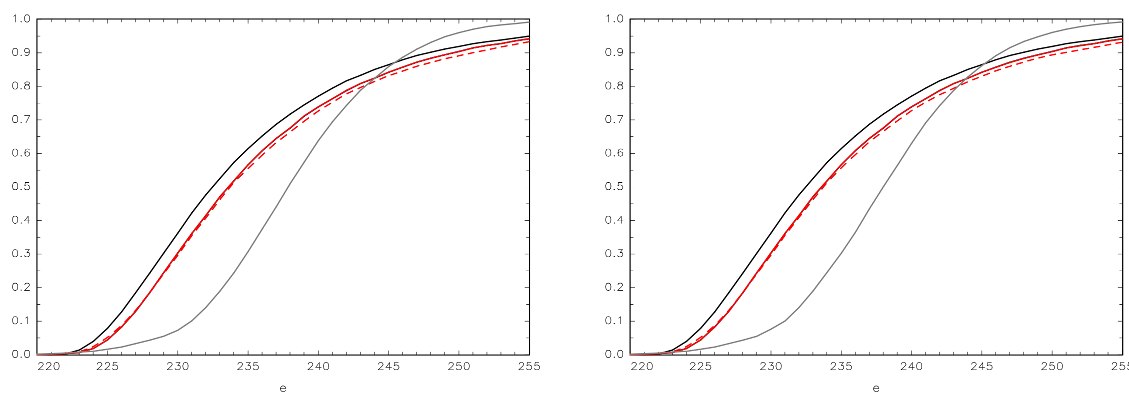
CUSUM^{WMV}(Forced):– –, CUSUM^{X}:——

Figure A.41: β = 0.0, ρ = 0.0, σ_{12} = 0.0, α_1 = 0.0. TPR of Procedures. Mean Shift at t = 230 in an Irrelevant Covariate entered in Levels. Left Panel = Upward Shift, Right Panel = Downward Shift.

Break Magnitude = 0.5σ



Break Magnitude = 1σ



CUSUM:——, CUSUM^V:——, CUSUM^{V*}:——, CUSUM^{WMV}:——

CUSUM^{WMV}(Forced):– –, CUSUM^X:——

Figure A.42: $\beta = 0.5$, $\rho = 0.0$, $\sigma_{12} = 0.0$, $\alpha_1 = 0.0$. FPR of Procedures. Mean Shift at $t = 230$ in a Relevant Covariate entered in Levels. Left Panel = Upward Shift, Right Panel = Downward Shift.

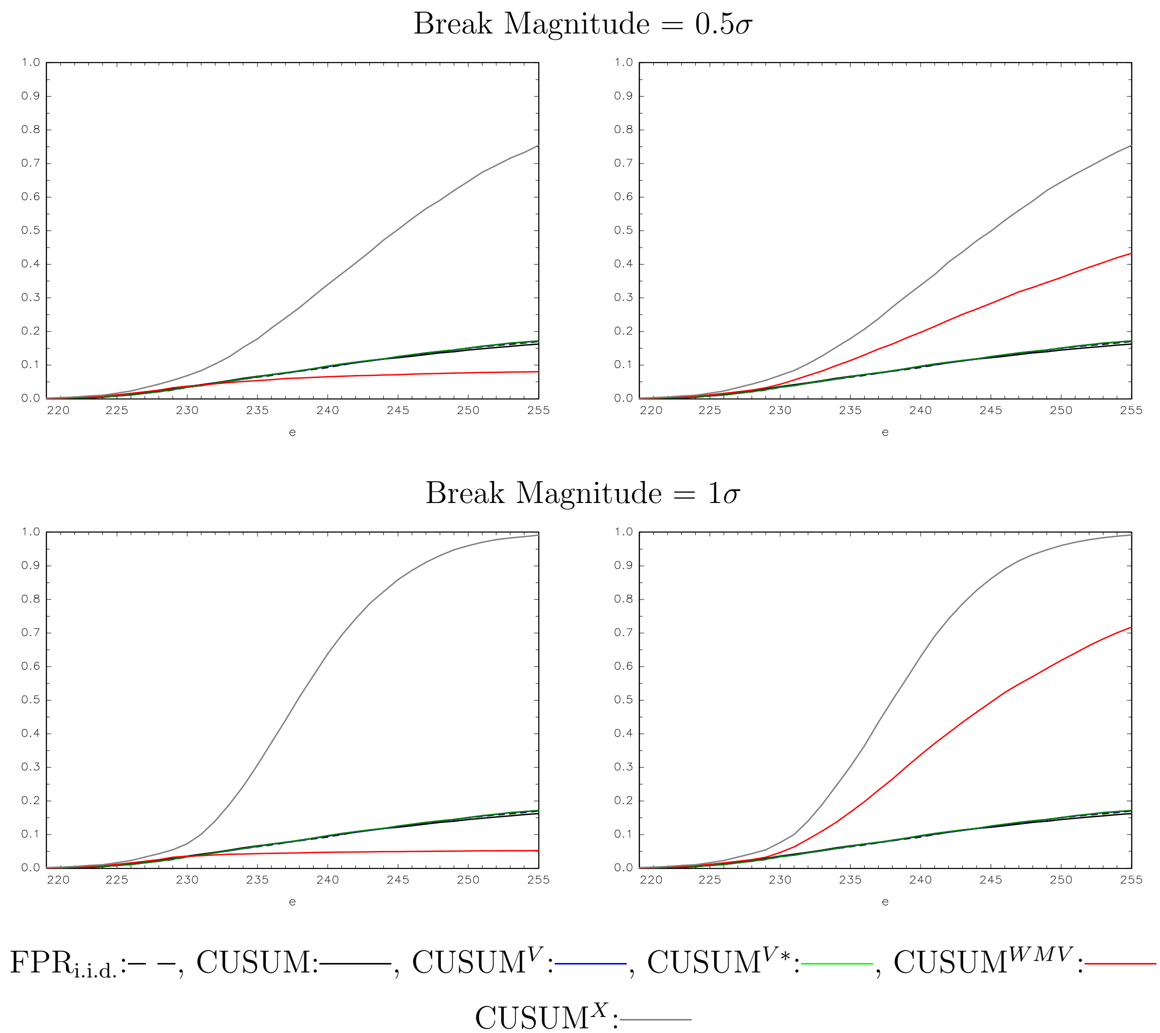
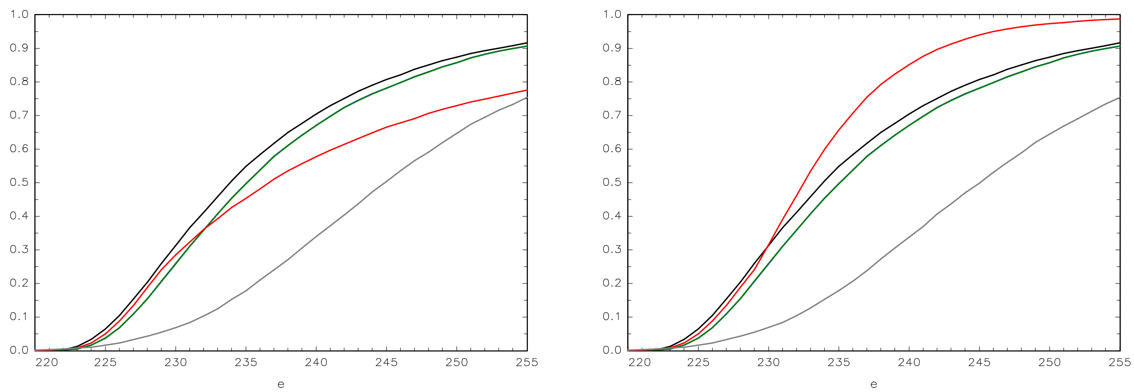
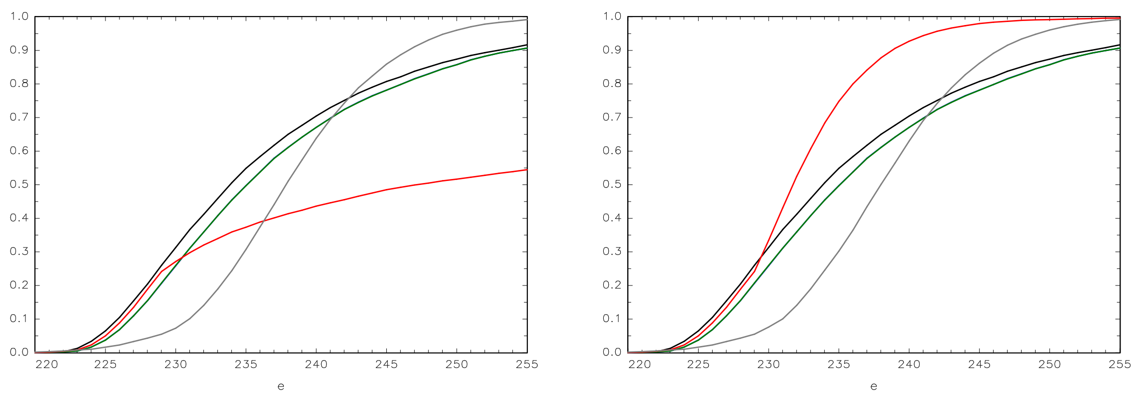


Figure A.43: $\beta = 0.5$, $\rho = 0.0$, $\sigma_{12} = 0.0$, $\alpha_1 = 0.0$. TPR of Procedures. Mean Shift at $t = 230$ in a Relevant Covariate entered in Levels. Left Panel = Upward Shift, Right Panel = Downward Shift.

Break Magnitude = 0.5σ



Break Magnitude = 1σ



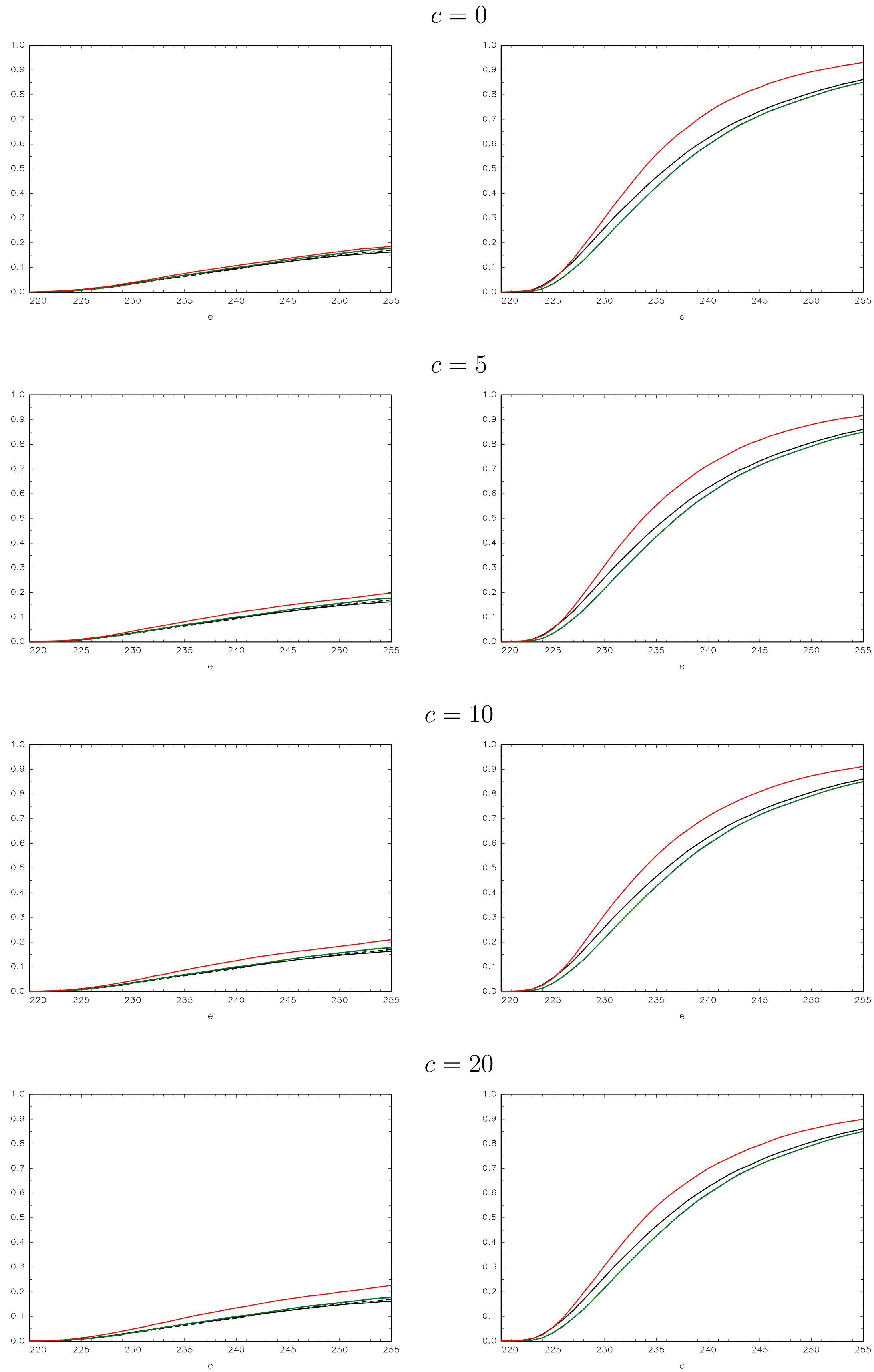
CUSUM:——, CUSUM^V:——, CUSUM^{V*}:——, CUSUM^{WMV}:——

CUSUM^X:——

A.4.9 Additional Simulations - Incorrectly Differenced Covariate

Figure A.44 reports the FPR and TPR of the procedures in the case where x_t is an unobserved relevant covariate, but what we actually observe is z_t, a strongly persistent (local-to-unity) process formed from x_t, and in order to remove the strong persistence we incorrectly take first differences of z_t, Δz_t, and then consider this covariate for inclusion in the CUSUM^{WMV} procedure. Data on x_t were therefore generated according to (1)-(2) and (23)-(24) with $\beta = 0.8$, $\rho = \sigma_{12} = \alpha_1 = 0.0$ and $\sigma^2_{1,t} = \sigma^2_{2,t} = 1\ \forall t$. We then generate $z_t = (1 - c/T)z_{t-1} + x_t$ as the observed series. The first differenced series, Δz_t, is then used as a covariate in the CUSUM^{WMV} procedure. Results are reported for $c \in \{= 0, 5, 10, 20\}$. We see that for $c = 0$ the FPR and TPR of the procedures is identical to those reported in Figure 2 panel(a), as would be expected given that in this case $\Delta z_t = x_t$. For the other values of c we see that the FPR of the CUSUM^{WMV} procedure is slightly inflated, and the TPR of the procedure is slightly reduced, with both of these effects increasing in the value of c. It should be noted, however, that while the TPR of the CUSUM^{WMV} procedure is decreasing in c, it is still significantly higher than the TPRs of all of the univariate procedures. These findings mirror those reported for covariate augmented unit root tests in Hansen (1995, pp.1159-1160) for this scenario. It is, however, worth noting that while the limiting null distribution of the covariate unit root tests proposed in Hansen (1995) depend in this scenario on c (when $c > 0$), in our context over-differenced covariates do not violate the regularity conditions given in Assumption 2 and, hence, the asymptotic null distribution of the sequence of $SWMV^t_T$, $t = T+1, ..., \lfloor \lambda T \rfloor$, statistics in this case is as given in Theorem 1, regardless of the value of c, such that the theoretical FPR of the resulting CUSUM^{WMV} procedure remains controlled according to the result in Corollary 1.

Figure A.44: $\beta = 0.5$, $\rho = 0.5$, $\sigma_{12} = 0.4$, $\alpha_1 = 0.2$ - Left Panel=FPR, Right Panel =TPR. Impact of Incorrectly Differenced Covariate



$\text{FPR}_{\text{i.i.d.}}$:– –, CUSUM:——, CUSUM^{V}:——, CUSUM^{V*}:——, CUSUM^{WMV}:——