Using Covariates to Improve the Efficacy of CUSUM Bubble Monitoring Procedures

Robert Taylor

University of Essex

(Based on joint work with Sam Astill and Yang Zu)

Sheffield Workshop on Bubbles and Crashes
15th May 2024

Outline of Presentation

Table of Contents

- 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), usually followed by a subsequent destruction of value through a sharp and catastrophic price collapse.
- As such, 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 reacted by considering new rules for macroprudential regulation and intervention.
- Crucial to the effectiveness of such approaches 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

- To date the literature has mainly focused on tests for (and dating of) the presence of historic periodically collapsing bubbles in asset price series. Seminal contributions include Phillips, Wu and Yu (2011) [PWY] and Phillips, Shi and Yu (2015) [PSY].
- A drawback of these historical bubble detection methods is that they are based on one-shot tests and it is not obvious how they could be validly applied in a sequential manner (as needs to be done in a real-time monitoring exercise), such that the theoretical false positive rate [FPR] of the resulting procedure (defined as the probability of incorrectly declaring at least one bubble episode during the monitoring period) could be controlled.
- There have been some attempts to develop FPR controlled real-time monitoring procedures for asset price bubbles. These split the data into a *training period* and a *monitoring period*. Homm and Breitung (2012) use a CUSUM-based detector, while Astill *et al.* (2018) use a method based on comparing the maximum value of test statistics computed in the training and monitoring periods.

- In practice, it would seem likely that information *additional* to the asset price series under test could usefully be deployed in bubble detection methods.
- Indeed, the extant literature suggests several potential covariates that could 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) and speculation measures (Singleton, 2014), whilst for cryptocurrencies, investor sentiment (Chen *et al.*, 2019) may play an important role.
- Finally, given bubble behaviour in real estate may precede equity (Cabellero *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.

- We explore how information from potential covariates can be incorporated into the CUSUM-based real-time monitoring procedure for explosive asset price bubbles developed in Homm and Breitung (2012).
- We show that where dynamic covariates are present in the data generating process [DGP], the FPR of the basic CUSUM procedure, which is based on the assumption that prices follow a univariate DGP will not, in general, be theoretically controlled.
- More positively, we establish that including relevant covariates in the construction of the CUSUM statistics both controls the theoretical FPR of the resulting procedure and can lead to substantial increases in the true positive rate [TPR] to detect an emerging bubble episode.

- We additionally allow for time varying volatility in the innovations driving the model through the use of a kernel-based variance estimator. A similar approach has been developed in the univariate setting by Astill, Harvey, Leybourne, Taylor and Zu (2023) [AHLTZ]
- An empirical illustration to the dataset of Welch and Goyal (2008) demonstrates that, in a pseudo real-time monitoring exercise, our proposed covariate augmented CUSUM procedure detects explosive episodes associated with the Black Monday crash and dotcom bubble sooner than approaches based on the assumption of a univariate DGP.

Table of Contents

We consider a time series process $\{y_t\}$ generated according to the DGP,

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

$$u_t = \begin{cases} u_{t-1} + v_t & t = 1, \ldots, \lfloor \tau_1 T \rfloor \\ (1+\delta)u_{t-1} + v_t & t = \lfloor \tau_1 T \rfloor + 1, \ldots, \lfloor \tau_2 T \rfloor \\ u_{t-1} + v_t & t = \lfloor \tau_2 T \rfloor + 1, \ldots, \lfloor \lambda T \rfloor \end{cases} \tag{2}$$

where $1 \leq \tau_1 < \tau_2 \leq \lambda$, $\lambda > 1$, and where $\lfloor . \rfloor$ denotes the integer part of its argument.

The *initial condition*, u_0, does not affect the subsequent analysis provided it is stochastically bounded, and we therefore set it to zero for expositional brevity.

Under the specification in (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 $t = \lfloor \tau_1 T \rfloor + 1$, before reverting back to $\delta_t = 0$ at time $t = \lfloor \tau_2 T \rfloor + 1$.

In the context of the DGP in (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 period* 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.

The Bubble Model

Our model imposes that y_t follows a unit root process over the training period $t = 1, ..., T$, while over the monitoring period y_t again follows a unit root process over the sub-periods $t = T + 1, ..., \lfloor \tau_1 T \rfloor$ and $t = \lfloor \tau_2 T \rfloor + 1, ..., \lfloor \lambda T \rfloor$, but crucially is subject to potentially explosive behaviour in the period $t = \lfloor \tau_1 T \rfloor + 1, ..., \lfloor \tau_2 T \rfloor$ if $\delta > 0$. In total, at the (notional) end of the monitoring period, there will be $\lfloor \lambda T \rfloor$ observations.

When $\delta > 0$, if $\tau_1 = 1$ then the explosive regime will begin at the start of the monitoring period, while if $\tau_2 = \lambda$, the explosive regime will still be on-going at the end 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 given by, $H_1 : \delta > 0$.

With regard 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, w_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

$$\alpha(L)v_t = \beta(L)' w_t + \varepsilon_t, \quad \varepsilon_t = \sigma_t \eta_t \tag{4}$$

where $\alpha(z) := 1 - \sum_{k=1}^{p} \alpha_k z^k$, $\beta(z) := \sum_{k=0}^{q} \beta_k z^k$, and where the m-vector of covariates, $w_t := (w_{1,t}, ..., w_{m,t})'$, is non-stochastic with mean vector zero. 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.

Precise conditions placed on w_t, η_t, the time-varying volatility σ_t, and the polynomial $\alpha(L)$ will be detailed subsequently in Assumption 2. In common with Hansen (1995), these rule out unit root behaviour in w_t, and so in many applications it will be natural for the covariates specified in w_t in (4) to constitute the first differences of other relevant financial and/or macroeconomic time series.

In the context of (4), the lag polynomial $\beta(L)$ allows for, but does not require, lags of the covariate w_t to enter the DGP, but excludes the possibility of leads of the covariate entering (4). This is because lead variables would clearly be unavailable to the practitioner in real-time.

Notice that the variables in w_t are not relevant covariates if $\beta(L) = 0$.

Under $H_0 : \delta = 0$, 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 have that

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

This is a heteroskedastic autoregressive model in Δy_t augmented by the level and (up to) q lags of the m covariates in w_t. Defining $z_t := (\Delta y_{t-1}, \ldots, \Delta y_{t-p}, w_t', w_{t-1}', ..., w_{t-q}')'$ and $\phi := (\alpha_1, \ldots, \alpha_p, \beta_0', \beta_1', \ldots, \beta_q')'$, (5) can be written more compactly as

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

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:

1. *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 $[0, \lambda]$, and is strictly positive, differentiable and uniformly bounded by a constant M. Furthermore, the derivative of $\sigma(\cdot)$ is Lipschitz continuous over $(0, \lambda)$.*
2. *Let η_t be a martingale difference sequence 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 $E(h_t) = \operatorname*{plim}_{T\to\infty} (1/\lfloor T\lambda \rfloor) \sum_{t=1}^{\lfloor T\lambda \rfloor} h_t = 1$.*
3. *$\{\eta_t\}$ is a strong mixing process with mixing coefficients of size $-r/(r-2)$, for some $r > 2$, and $E|\eta_t|^{2r} < \infty$.*
4. *$\alpha(z) \neq 0$ for all $|z| \leqslant 1$.*
5. *For all $0 \leq \kappa \leq \lambda$, $\operatorname*{plim}_{T\to\infty} (1/\lfloor T\kappa \rfloor) \sum_{t=1}^{\lfloor T\kappa \rfloor} z_t z_t' \sigma_t^2 h_t = \lim_{T\to\infty} E(1/\lfloor T\kappa \rfloor) \sum_{t=1}^{\lfloor T\kappa \rfloor} z_t z_t' \sigma_t^2 h_t = \Omega(\kappa)$, where $\Omega(\kappa)$ is a positive definite matrix with all the elements being finite and continuous in κ.*
6. *The vector of covariates 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 $\delta > 0$.*

Overall, our specification for the covariates is considerably more general than is imposed by Kramer, Ploberger and Alt (1988) in the context of their CUSUM test, or by Hansen (1995), Chang *et al.* (2017) [CSS] and Astill, Taylor, Kellard, and Korkos (2023) [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.

Hansen (1995), CSS and ATKK also require that the covariates are weakly dependent for their large sample results to hold. Although we do not need to impose the condition that the covariates are weakly dependent, the strength of the dependence allowed in the covariates is restricted by Assumption 2 which, for example, rules out covariates with (near-) unit roots.

Table of Contents

The Homm and Breitung (2012) CUSUM test

Using a training period of $t = 1, ..., T$, Homm and Breitung (2012) propose testing for explosive behaviour in the monitoring period based on the CUSUM statistic:

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

where $T < t \leq \lfloor T\lambda \rfloor$ is the monitoring observation, and $\tilde{\sigma}_t^2$ is a consistent estimate of σ^2 based on sample observations up to and including time t.

Homm and Breitung (2012) show that if the CUSUM statistic, 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, the joint limiting ($T \to \infty$) distribution of this sequence is given by,

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

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

The Homm and Breitung (2012) CUSUM test

Using Theorem 3.4 of Chu *et al.* (1996), Homm and Breitung (2012) show that under H_0, the result in (8) implies that, for any $\lambda > 1$,

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

where $c_t := \sqrt{b_\alpha + log(t/T)}$. The CUSUM monitoring procedure proposed in Homm and Breitung (2012) 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.

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).

It should be noted that these asymptotic settings for b_α assume a monitoring period of theoretically 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.

Homm and Breitung (2012) 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 a pure unit process driven by NIID(0,1) innovations.

The Astill *et al.* (2023) CUSUM test

Astill *et al.* (2023) [AHLTZ] show that the CUSUM procedure of Homm and Breitung (2012) does not have a theoretically controlled FPR if the errors display non-stationary volatility; that is, if $v_t = \sigma_t \eta_t$ with the volatility function, σ_t, displaying time-variation of the form specified by Assumption 2.1.

To correct for this, AHLTZ propose a modified version of the CUSUM procedure of Homm and Breitung (2012) where the S_T^t statistic in (7) is replaced by 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 based estimate for the spot variance, $\sigma_j^2 := \sigma^2(j/T)$.

The Astill *et al.* (2023) CUSUM test

The kernel smoothing estimator is defined as follows for $j \geqslant N+1$:

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

where the kernel function, $K(\cdot)$, and bandwidth, N, satisfy some technical conditions outlined in AHLTZ.

AHLTZ establish that the analogous CUSUM monitoring procedure based on SV_T^t rather than S_T^t is able to control the FPR when the innovations v_t in (2) exhibit time varying volatility, while retaining finite sample TPRs ("power") close to those of the standard CUSUM procedure of Homm and Breitung (2012) when the innovations are homoskedastic.

Extant Tests

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 it is also, in general, violated (even if $p = 0$) when $\beta(L) \neq 0$ if, for example, either the covariates, w_t, are serially correlated, or $q > 0$, or both. As a result the large sample results in (8) and (9) will not necessarily hold in such cases for S_T^t and SV_T^t.

Consequently, implementing CUSUM and CUSUM^V using the critical values from Homm and Breitung (2012) would result in monitoring procedures where the (theoretical) FPR would not necessarily be at the level expected by the practitioner.

Table of Contents

In the paper, we initially develop monitoring procedures that assume that the vector of covariates, w_t, has zero mean. This is restrictive in practice, so in the interest of time we will move straight to our preferred monitoring procedure that allows for a non-zero mean in the vector of covariates.

In order to allow for covariates with non-zero mean, we replace Assumption 1, with Assumption 3

Assumption 3

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

$$\alpha(L)v_t = \beta(L)'(x_t - c_x) + \varepsilon_t, \quad \varepsilon_t = \sigma_t \eta_t, \tag{12}$$

where c_x denotes the mean vector of the covariates, x_t. All other aspects of (12) remain as previously specified for w_t, appropriately redefined to the non-deterministic component of the covariate, $x_t - c_x =: w_t$, where relevant.

Under Assumption 3, in order to obtain CUSUM statistics which are exact invariant to c_x we need to augment the null regression (5) by a constant term, viz,

$$\Delta y_t = \mu + \sum_{k=1}^{p} \alpha_k \Delta y_{t-k} + \sum_{k=0}^{q} \beta_k' x_{t-k} + \varepsilon_t, \quad \varepsilon_t = \sigma_t \eta_t \tag{13}$$

Notice that inclusion of the constant term in (13) also entails that we can allow for the possibility of a non-zero mean in Δy_t, and, hence, a deterministic linear trend in y_t.

For ease of notation, re-write the null regression in (13) as

$$\Delta y_t = \varphi' g_t + \varepsilon_t, \quad \varepsilon_t = \sigma_t \eta_t, \tag{14}$$

where $\varphi := (\mu, \alpha_1, \ldots, \alpha_p, \beta_1', \ldots, \beta_q')'$, $g_t := (1, \Delta y_{t-1}, \ldots, \Delta y_{t-p}, x_t', x_{t-1}', \ldots, x_{t-q}')'$. Notice the regression has $1 + p + qm$ regressors.

Consider the infeasible re-weighting of (14), based on the true volatility function σ_t, given by

$$\frac{\Delta y_t}{\sigma_t} = \varphi' \frac{g_t}{\sigma_t} + \eta_t, \quad t = 1, \ldots, T, \ldots, \lfloor \lambda T \rfloor. \tag{15}$$

The associated (infeasible) recursive WLS estimator for φ at each point in the monitoring sample is then given by

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

and the corresponding (infeasible) recursive residuals based on the recursive WLS estimates are defined as

$$e_t^W := \Delta y_t - (\tilde{\varphi}_{t-1}^W)' g_t, \quad t = T+1, \ldots, \lfloor \lambda T \rfloor. \tag{16}$$

Following the proof strategy given for Theorem 1 of Kramer, Ploberger and Alt (1988), we show that, under the no bubble null hypothesis, the associated (infeasible) sequence of CUSUM statistics

$$SWM_T^t := \sum_{j=T+1}^{t} e_j^W / \sigma_j, \quad t > T$$

satisfies

$$T^{-1/2} SWM_T^{\lfloor Tr \rfloor} \Rightarrow \mathbb{W}(r) - \mathbb{W}(1), \;\; 1 < r \leq \lambda$$

where $\mathbb{W}(\cdot)$ is a standard Brownian motion on $[0, \lambda]$, such that we recover the usual Brownian motion-based limiting null distribution as obtained by the original CUSUM statistic of Homm and Breitung (2012).

To obtain a feasible version of SWM_T^t, we need to replace the true volatilities by estimates thereof.

To that end, we define the double array of OLS residuals from (14) as,

$$f_{i,t}^* := \Delta y_i - (\hat{\varphi}_t)' g_i, \quad i = \max(p+2, q+1), \ldots, t, \ t = T+1, \ldots, \lfloor \lambda T \rfloor. \tag{17}$$

Based on these residuals we consider the following spot variance estimator

$$\tilde{\sigma}_{j,N,t}^2 := \sum_{s=0}^{N} k_s (f_{j-s,t}^*)^2, \quad k_s := \frac{K\left(\frac{s}{N}\right)}{\sum_{s=0}^{N} K\left(\frac{s}{N}\right)}, \tag{18}$$

for some kernel function, $K(\cdot)$, and bandwidth parameter, N.

Using these residuals we can then define the *feasible* recursive WLS estimator

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

where $\tilde{\sigma}_{j,N,t}^2$ is defined applying the formula in (18) to the $\{f_{i,t}^*\}$ residuals defined in (17).

A feasible version of the sequence of SWM_T^t statistics is then defined as,

$$SWMV_T^t := \sum_{j=T+1}^{t} \frac{\hat{e}_j^W}{\tilde{\sigma}_{j,N,j}}, \qquad t > T \tag{19}$$

where $\hat{e}_j^W := \Delta y_j - (\hat{\varphi}_{j-1}^W)' g_j$.

In order to establish the the limiting null distribution of the sequence of $SWMV_T^t$ statistics we need to replace Assumption 2.5 on the regressors in the OLS regression, (6), with a corresponding condition on the WLS regression with a constant included, (15). Analogously to Assumption 2.5, this excludes the possibility of asymptotic collinearity between the regressors.

Assumption 4

For all $0 \leq \kappa \leq \lambda$, $\text{plim}_{T\to\infty}(1/\lfloor T\kappa \rfloor)\sum_{s=1}^{\lfloor T\kappa \rfloor} g_s g_s' h_s/\sigma_s^2 = \lim_{T\to\infty}(1/\lfloor T\kappa \rfloor)$ $E(\sum_{s=1}^{\lfloor T\kappa \rfloor} g_s g_s' h_s/\sigma_s^2) = \Theta(\kappa)$, *where* $\Theta(\kappa)$ *is a positive definite matrix with all the elements being finite and continuous in* κ.

Theorem 1

Let the data be generated according to (1)-(3), (12) under the null hypothesis $H_0 : \delta = 0$. If Assumptions 2-4 hold, excluding Assumption 2.5, then, as $T \to \infty$, it follows that

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

where $\mathbb{W}(\cdot)$ denotes a standard Brownian motion on $[0, \lambda]$, and

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

Theorem 1 establishes that the large sample properties of the CUSUM^{WMV} procedure again coincide with those which obtain for the original CUSUM procedure of Homm and Breitung (2012).

The practical implementation of $SWMV_T^t$ requires choices to be made for both the kernel function and bandwidth used in constructing the nonparametric volatility estimator.

In both our simulations and empirical exercise we employed a Gaussian kernel with the bandwidth, N, selected using the local cross validation criteria recommended by AHLTZ.

If it were known that the volatility function $\sigma_t = \sigma < \infty$, for all t, then one could 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 recursive LS estimator of the coefficient vector φ in (14).

Notice that the constancy of the volatility function means that the WLS transformation is no longer needed when computing the recursive residuals so that the numerator of this statistic is based on the recursive residuals e_j^* rather than $\hat{e}_j^W$.

Table of Contents

We generated data 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 we assume monitoring ends at time $\lambda T = 255$. Under the null we set $\delta = 0$, whereas under the alternative we set $\delta = 0.005$, $\lfloor \tau_1 T \rfloor = 220$ and $\lfloor \tau_2 T \rfloor = \lambda T$, such that y_t follows a unit root process during the training period, before switching to an explosive regime which starts when monitoring commences and continues until the end of the monitoring period.

For the error term v_t and the covariate w_t, we use an unconditionally heteroskedastic extension of the simulation DGP detailed in section 5.1 on page 143 of Chang *et al.* (2017) [CSS]:

$$v_t = \alpha_1 v_{t-1} + \beta w_t + \varepsilon_{1,t}, \qquad (20)$$

$$w_{t+1} = \rho w_t + \varepsilon_{2,t}, \qquad (21)$$

with the covariate initialised at $w_0 = 0$.

The variance matrix of the innovation vector, $(\varepsilon_{1,t}, \varepsilon_{2,t})'$, was generated according to:

$$\begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} \sim NIID(0, \Sigma_t), \qquad \Sigma_t := \begin{bmatrix} \sigma^2_{1,t} & \sigma_{12,t} \\ \sigma_{12,t} & \sigma^2_{2,t} \end{bmatrix} \tag{22}$$

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

$$\sigma_{1,t} := 1 + (\sqrt{4} - 1)\,[1 + \exp(-\theta(t - 219))]^{-1} \tag{23}$$
$$\sigma_{2,t} := 1 + (\sqrt{4} - 1)\,[1 + \exp(-\theta(t - 219))]^{-1} \tag{24}$$

with $\theta = 0.25$; that is, a logistic smooth transition in volatility from 1 to $\sqrt{4}$ centred on the end of the training period.

In the paper 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}. This corresponds to the homoskedastic model for $(v_t, w_t)'$ considered by CSS.

(b) $\sigma_{1,t}$ and $\sigma_{2,t}$ satisfy (23) and (24), respectively, 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 (23), $\sigma_{2,t} = 1$, 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 (23), $\sigma_{2,t} = 1$, 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 $\sigma^2_{1,t}$.

For the purposes of the talk, I'll focus on some examples from cases (a) and (b).

We report rejection rates for the CUSUM^{WMV} procedure, together with those of the standard CUSUM procedure of Homm and Breitung (2012) 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 but where the null regression is given by (6) but excludes the covariate terms. The rationale behind including this is that including only lags of Δy_t will still lead to a procedure that is able to deal with any serial correlation in Δy_t induced by the presence of the covariate in the DGP without being able to exploit the covariate for power gains under the alternative. It therefore provides an FPR controlled benchmark against which to quantify the power gains from including the covariate.

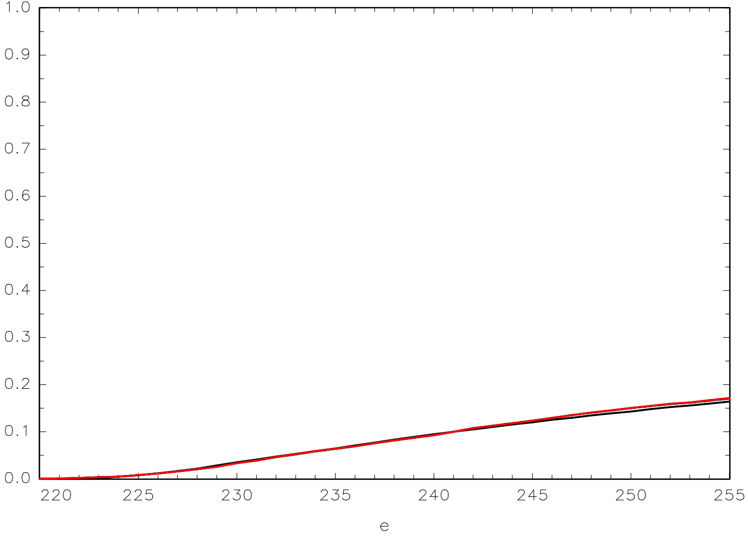
The null regression we use in connection with CUSUM^{V*} will not contain an intercept as, when excluding the covariate from the regression, an intercept is only required if we wished to allow for a deterministic linear trend in y_t.

When performing the CUSUM^{WMV} procedure we use the BIC applied to models estimated by OLS to select the values of p and q in (13) based on data available up to the first monitoring observation, setting the maximum permitted values of p and q to $p_{\max} = 4$ and $q_{\max} = 2$, respectively. For the CUSUM^{V*} procedure we also set the maximum permitted value of p to $p_{\max} = 4$.

If the minimum value of the BIC from the regression underlying the CUSUM^{V*} procedure is below that from (13) we determine that the candidate covariate is irrelevant and so inference for the CUSUM^{WMV} procedure is instead based on the CUSUM^{V*} procedure. Additionally, if the BIC selects $p = 0$ in the regression underlying the CUSUM^{V*} procedure the CUSUM^{V*} procedure reduces to the CUSUM^{V} procedure of AHLTZ which is based only on Δy_t.

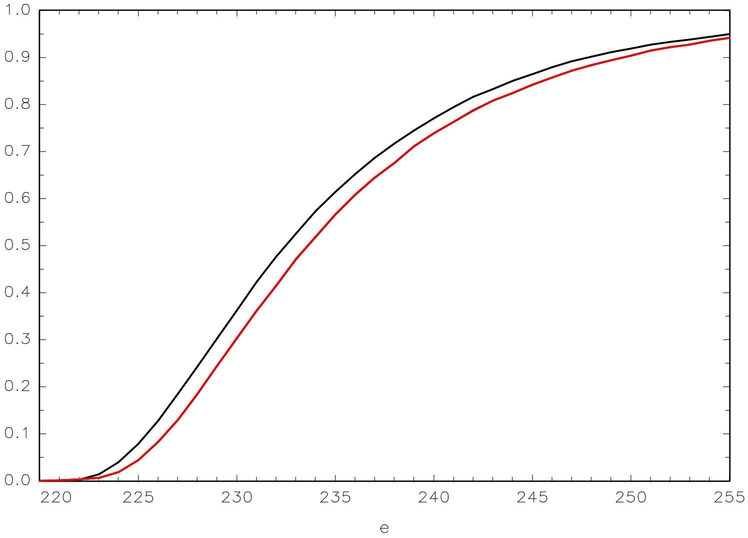
As in Homm and Breitung (2012), we select a value of b_α such that the FPR of the monitoring procedures is equal to 0.10 by time $t = 241$ when the series y_t is a pure unit root process driven by $NIID(0, 1)$ innovations and the covariate is an irrelevant white noise process, i.e. $\beta = \rho = \alpha_1 = 0$ and $\sigma_{12} = 0, \sigma^2_{1,t}, \sigma^2_{2,t} = 1$, for all t.

FPR - Univariate DGP $\beta = \rho = \sigma_{12} = \alpha_1 = 0$. Homoskedastic



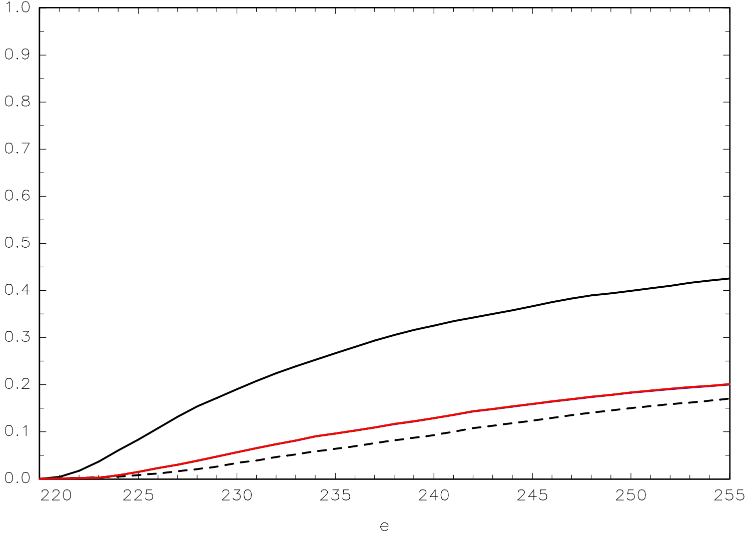
$\text{FPR}_{\text{i.i.d.}}$:– –, CUSUM:——, CUSUM^{V}:——, CUSUM^{V*}:——, CUSUM^{WMV}:——

TPR - Univariate DGP $\beta = \rho = \sigma_{12} = \alpha_1 = 0$. Homoskedastic



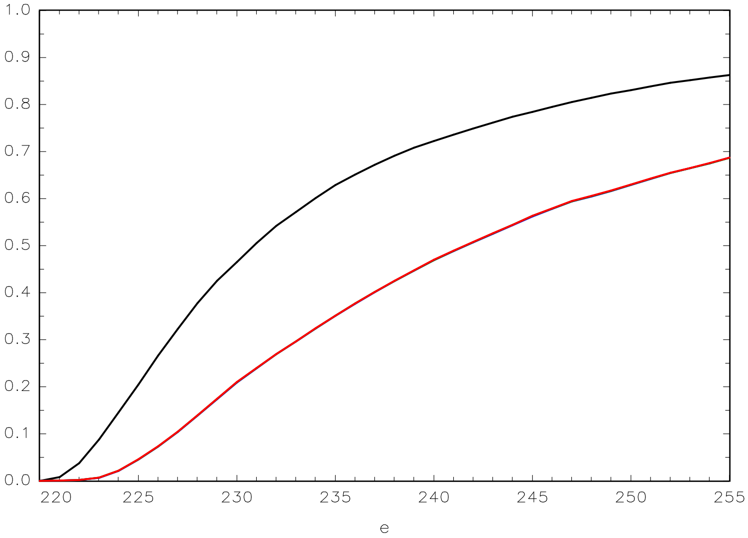
CUSUM:——, CUSUM^{V}:——, CUSUM^{V*}:——, CUSUM^{WMV}:——

FPR - Univariate DGP $\beta = \rho = \sigma_{12} = \alpha_1 = 0$. Scenario (b)



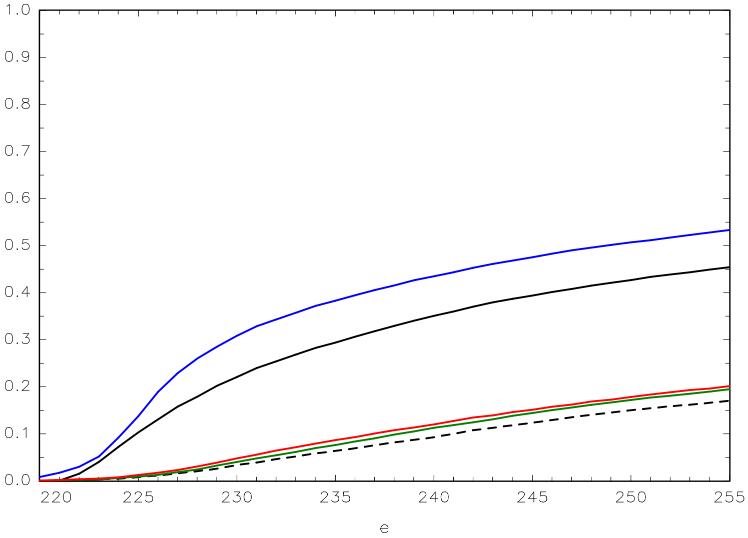
$\text{FPR}_{\text{i.i.d.}}$:– –, CUSUM:——, CUSUM^{V}:——, CUSUM^{V*}:——, CUSUM^{WMV}:——

TPR - Univariate DGP $\beta = \rho = \sigma_{12} = \alpha_1 = 0$. Scenario (b)



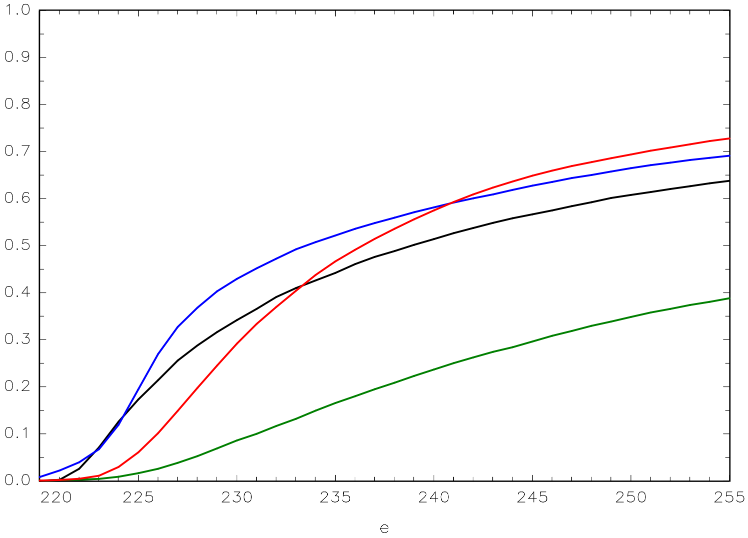
CUSUM:——, CUSUM^V:——, CUSUM^{V*}:——, CUSUM^{WMV}:——

FPR - $\beta = 0.8$, $\rho = 0.8$, $\sigma_{12} = 0.4$, $\alpha_1 = 0.2$. Homoskedastic



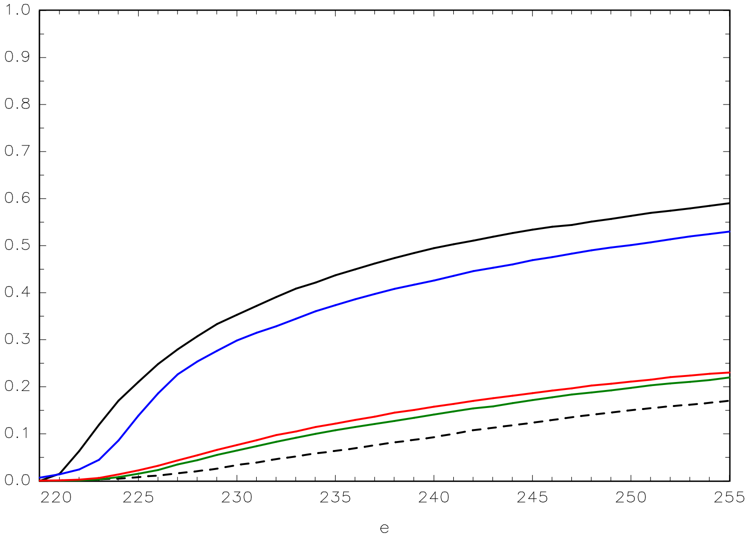
$\text{FPR}_{\text{i.i.d.}}$:– –, CUSUM:——, CUSUM^{V}:——, CUSUM^{V*}:——, CUSUM^{WMV}:——

TPR - $\beta = 0.8$, $\rho = 0.8$, $\sigma_{12} = 0.4$, $\alpha_1 = 0.2$. Homoskedastic



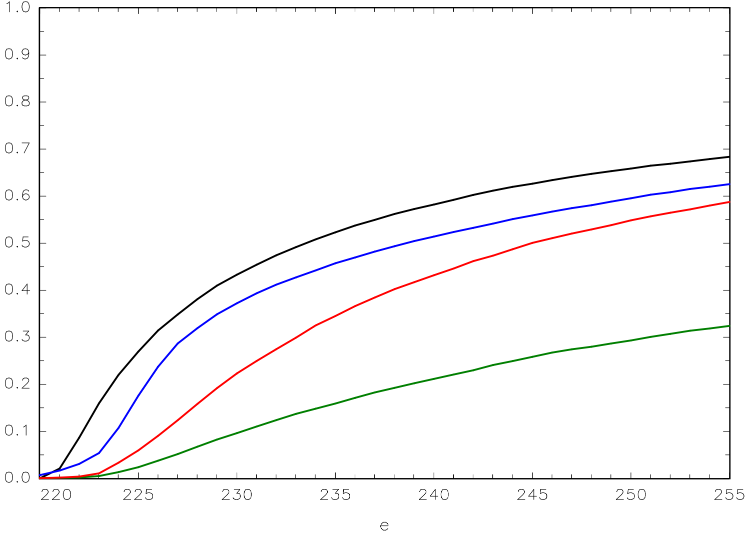
CUSUM:——, CUSUM^V:——, CUSUM^{V*}:——, CUSUM^{WMV}:——

FPR - $\beta = 0.8$, $\rho = 0.8$, $\sigma_{12} = 0.4$, $\alpha_1 = 0.2$. Scenario (b)



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

TPR - $\beta = 0.8$, $\rho = 0.8$, $\sigma_{12} = 0.4$, $\alpha_1 = 0.2$. Scenario (b)



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

Many, many more simulations are available in the paper!

Overall, the BIC does a decent job of discarding irrelevant covariates and including relevant covariates.

Very little "power" is lost by using our proposed CUSUM^{WMV} procedure relative to the CUSUM^{V} procedure of AHLTZ under the bubble alternative when the covariate is irrelevant.

On the other hand, when a relevant covariate enters the DGP the standard CUSUM and CUSUM^{V} procedures display significant FPR distortions under the null, while the FPR of the CUSUM^{WMV} procedure is well controlled when the innovations are either homoskedastic or exhibit time varying volatility.

The TPR gains from including a relevant covariate under the alternative are clearly seen by examining the difference in rejection frequencies between CUSUM^{WMV} and CUSUM^{V*}, the latter being the only univariate procedure to exhibit decent FPR control across all scenarios considered.

Table of Contents

We now turn to an investigation into how our proposed monitoring procedures would behave 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 to be monitored for emerging bubble episodes is the price-dividend ratio (Index/D12). As candidate covariates we consider: 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 discussed earlier.

We use the same bandwidth selection rule as in the MC simulations and, again, use the BIC to select p and q, as well as whether to include the covariates at all, in the pre-whitening regression (13), 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 first 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

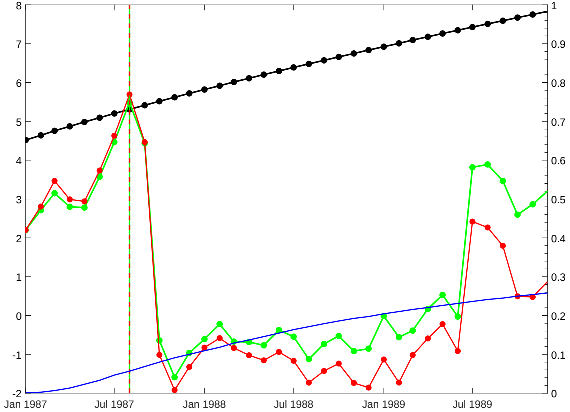
From this we found that the variables ltr, corpr and dfr can be used as possible covariates in levels, whereas rfree, infl, svar, tms, E12, b/m, tbl, GS10, AAA, BAA, lty, ntis, csp, de, ep and dfy need to enter in first differences.

At the start of the monitoring procedure applying the BIC 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.

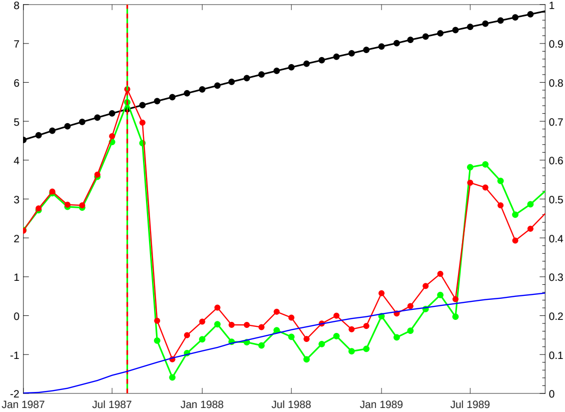
For the CUSUM^{V*} procedure the BIC selects a lag length of $p = 0$ so that this procedure is identical to the CUSUM^{V} procedure, we therefore report results only for the latter procedure.

We 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.

Black Monday - b/m

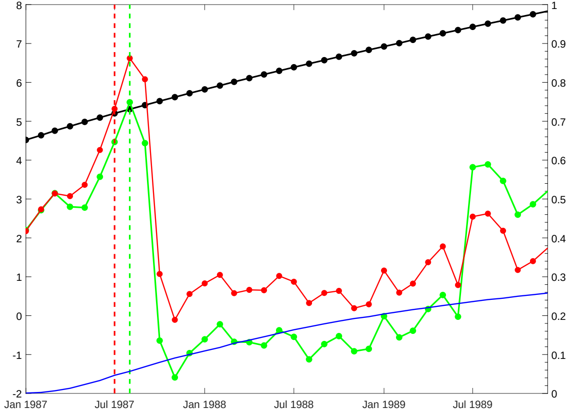


LEFT AXIS: Critical Value:——, CUSUM^V, CUSUM^{V*}:——,CUSUM^{WMV}:——
RIGHT AXIS: False Positive Rate:——



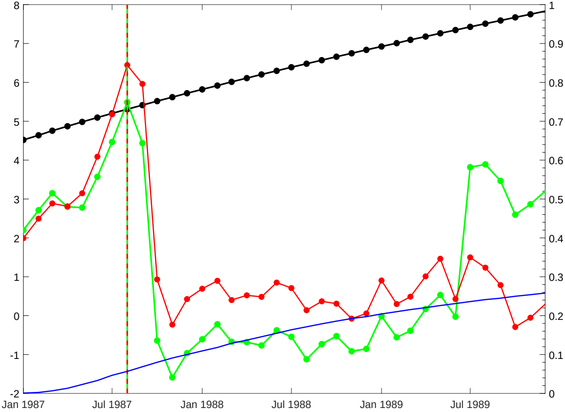
LEFT AXIS: Critical Value:——, CUSUM^V, CUSUM^{V*}:——,CUSUM^{WMV}:——
RIGHT AXIS: False Positive Rate:——

Black Monday - AAA



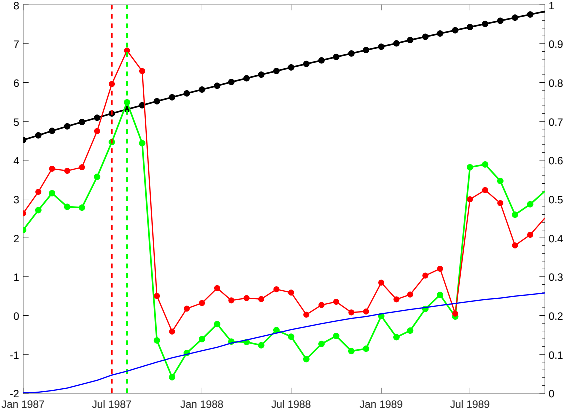
LEFT AXIS: Critical Value:——, CUSUM^{V}, CUSUM^{V*}:——,CUSUM^{WMV}:——
RIGHT AXIS: False Positive Rate:——

Black Monday - BAA



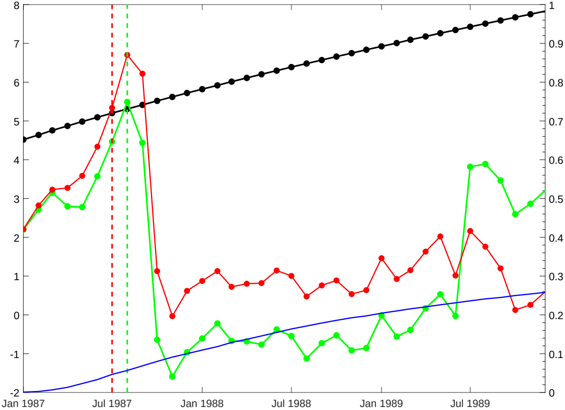
LEFT AXIS: Critical Value:——, CUSUM^V, CUSUM^{V*}:——,CUSUM^{WMV}:——
RIGHT AXIS: False Positive Rate:——

Black Monday - LTY

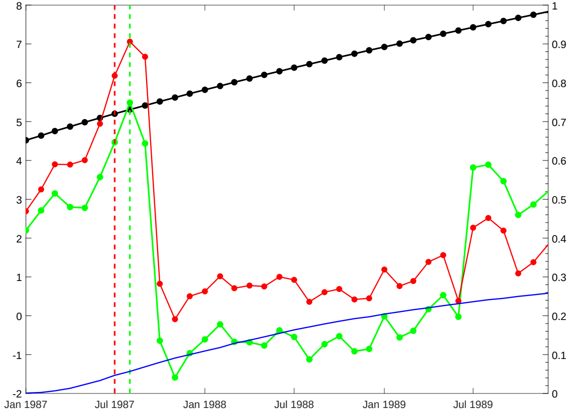


LEFT AXIS: Critical Value:——, CUSUM^{V}, CUSUM^{V*}:——,CUSUM^{WMV}:——
RIGHT AXIS: False Positive Rate:——

Black Monday - GS10

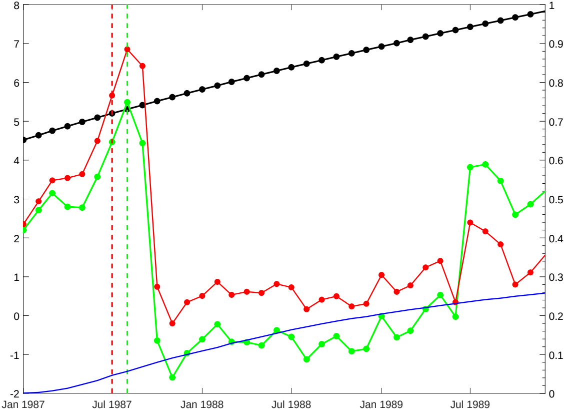


LEFT AXIS: Critical Value:——, CUSUM^V, CUSUM^{V*}:——,CUSUM^{WMV}:——
RIGHT AXIS: False Positive Rate:——



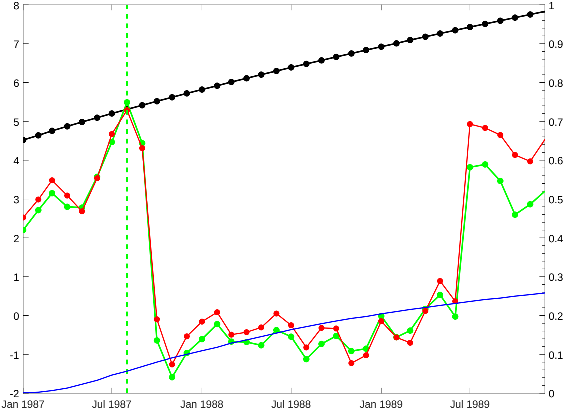
LEFT AXIS: Critical Value:——, CUSUM^{V}, CUSUM^{V*}:——, CUSUM^{WMV}:——
RIGHT AXIS: False Positive Rate:——

Black Monday - corpr



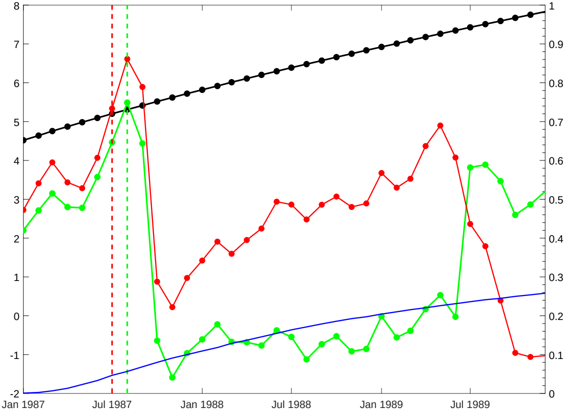
LEFT AXIS: Critical Value:——, CUSUM^{V}, CUSUM^{V*}:——,CUSUM^{WMV}:——
RIGHT AXIS: False Positive Rate:——

Black Monday - csp



LEFT AXIS: Critical Value:——, CUSUM^V, CUSUM^{V*}:——,CUSUM^{WMV}:——
RIGHT AXIS: False Positive Rate:——

Black Monday - ep



LEFT AXIS: Critical Value:——, $CUSUM^{V}$, $CUSUM^{V*}$:——, $CUSUM^{WMV}$:——
RIGHT AXIS: False Positive Rate:——

The plots of these CUSUM statistics suggest that the bubble episode prior to Black Monday was short lived, with only a small window of opportunity for detection before the collapse of the price-divided ratio.

Nonetheless, we see that the CUSUM^{WMV} procedure would have detected this bubble in July 1987 when using 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 the univariate CUSUM^{V} procedure.

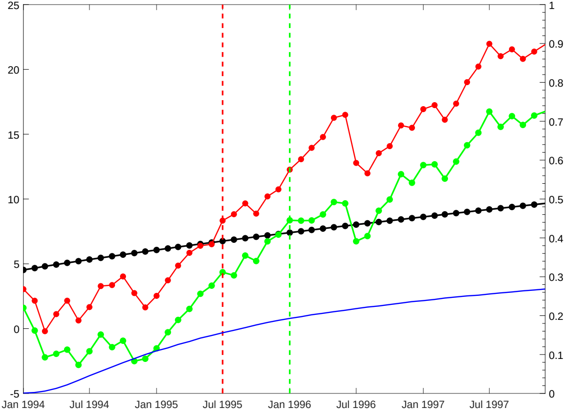
For the other covariates, the CUSUM^{WMV} procedure first rejects at the same time as CUSUM^{V}, excepting Δ(csp) where the CUSUM^{WMV} procedure marginally fails to reject in August 1987.

We next examine how a monitoring procedure that began in January 1994, ahead of the dotcom bubble, would have played out.

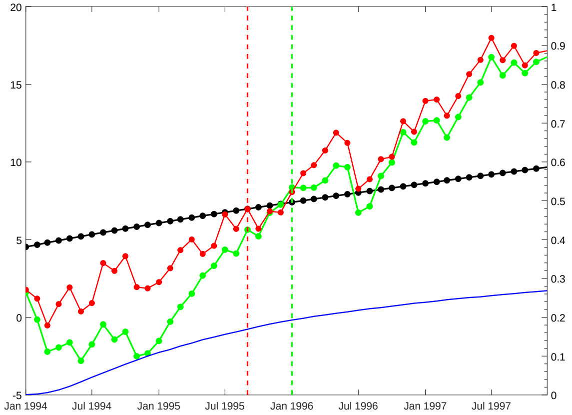
The monitoring procedures were performed exactly as for the Black Monday exercise, except that the training sample of data was of length $T = 72$, running from January 1988 to December 1993, in order to avoid the abrupt collapse in the price-dividend ratio witnessed at the end of 1987 following Black Monday.

This necessitated setting $b_\alpha = 0.2672$ to retain an FPR of 0.10 after 1 year, again assuming the price-dividend data were generated by a unit root process driven by NIID innovations. 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 report results only for the set of covariates that were just seen to be useful in the context of the bubble episode prior to Black Monday.

Dot Com - AAA



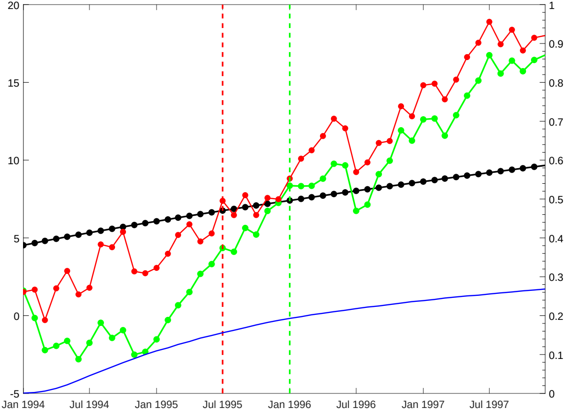
LEFT AXIS: Critical Value:——, CUSUM^{V}, CUSUM^{V*}:——,CUSUM^{WMV}:——
RIGHT AXIS: False Positive Rate:——



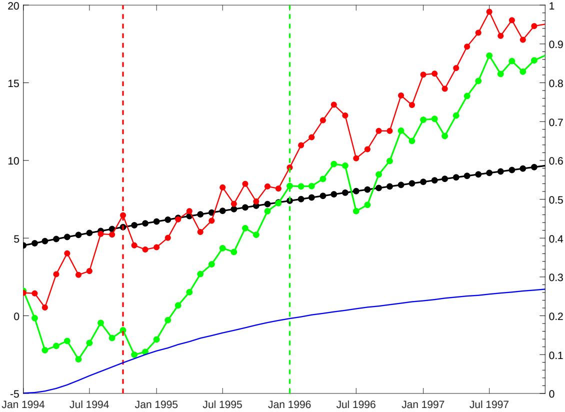
LEFT AXIS: Critical Value:——, CUSUM^V, CUSUM^{V*}:——,CUSUM^{WMV}:——

RIGHT AXIS: False Positive Rate:——

Dot Com - Itr



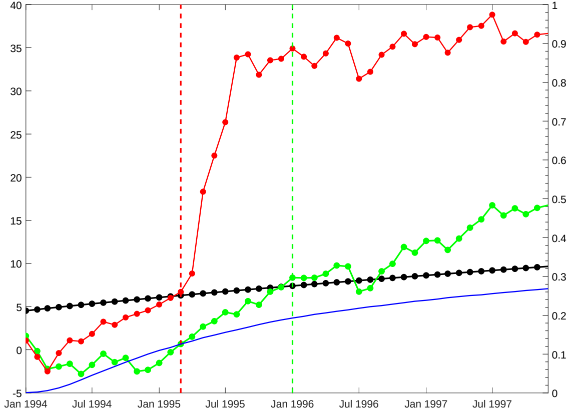
LEFT AXIS: Critical Value:——, CUSUM^V, CUSUM^{V*}:——,CUSUM^{WMV}:——
RIGHT AXIS: False Positive Rate:——



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

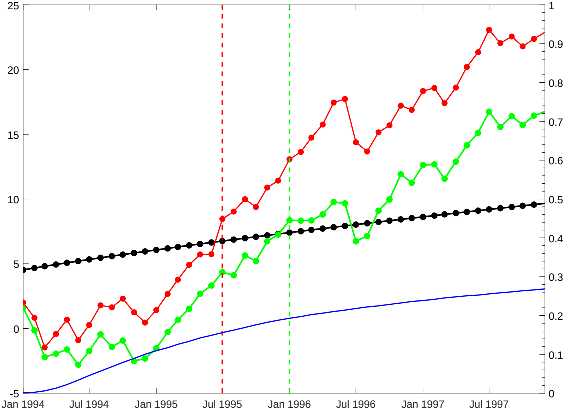
RIGHT AXIS: False Positive Rate:——

Dot Com - ep



LEFT AXIS: Critical Value:——, CUSUM^V, CUSUM^{V*}:——,CUSUM^{WMV}:——
RIGHT AXIS: False Positive Rate:——

Dot Com - GS10



LEFT AXIS: Critical Value:——, CUSUM^{V}, CUSUM^{V*}:——,CUSUM^{WMV}:——
RIGHT AXIS: False Positive Rate:——

The CUSUM^V procedure which incorporates no covariate augmentation first rejects the null of no bubble in January 1996.

The CUSUM^{WMV} procedure rejects earlier when using any of these 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).

Table of Contents

Conclusions

We develop an extension of the CUSUM real-time bubble monitoring procedure of Homm and Breitung (2012) which incorporates additional information from covariates, and is designed to be robust to time-varying volatility in the data.

We show that the sequence of extended CUSUM statistics underlying our monitoring procedure has the same joint limiting null distribution as the sequence of original CUSUM statistics in Homm and Breitung (2012) has (under the more restrictive conditions in their paper), so that a monitoring procedure based on the well-known boundary function used in Homm and Breitung (2012) remains asymptotically valid.

Monte Carlo simulation evidence shows that, in contrast to the univariate monitoring procedures of Homm and Breitung (2012) and AHLTZ, where a relevant covariate enters the DGP our proposed procedure has a controlled FPR under the null. Moreover, under the alternative the information from the covariate can lead to potentially significant gains in TPR over these univariate procedures.

An empirical application to the dataset of Welch and Goyal (2008) highlights the empirical usefulness of our covariate augmented monitoring procedure. In a pseudo real-time monitoring exercise, for several plausible covariates it would have led to an earlier rejection than the univariate procedure of AHLTZ for both the Black Monday and dotcom bubble episodes.