Transregio 391

TRR 391 Working Paper #8

November 2025

Real-Time Monitoring for Stock Return Predictability in Nonstationary Volatility Environments

Matei Demetrescu[a], Fabian Schmidt[a], A.M. Robert Taylor[b]

Suggested citation:
Demestrescu, M., Schmidt, F., and Taylor, A.M.R. (2025). "Real-Time Monitoring for Stock Return Predictability in Nonstationary Volatility Environments". *TRR 391 Working Paper #8.* DOI: 10.17877/DE290R-25840.

Version 1.0, November 2025

[a] Department of Statistics, TU Dortmund University
[b] Essex Business School, University of Essex
Corresponding author: robert.taylor@essex.ac.uk

Abstract

Stock return predictability, should it exist, is likely to be episodic in nature. In order to exploit such *pockets of predictability* it is essential that they are rapidly detected, in real-time, as the nascent predictive regime emerges. This will typically entail the repeated (sequential) application of one-shot end-of-sample predictability statistics, updated as new data become available. Consequently, in addition to dealing with the usual inference problems caused by unknown regressor persistence and predictive regression endogeneity, one must also account for the multiple testing issues inherent in such monitoring procedures. In addition, stock returns and/or the predictors commonly used typically exhibit time-varying volatility and it is known that ignoring such data features can result in the spurious detection of predictability. We propose real-time monitoring procedures which take account of these issues. Our preferred monitoring strategy uses a CUSUM-type procedure based on a specific moment condition related to the predictive power of the putative predictor. We implement nonparametric adjustment methods to allow for the possibility of time varying volatility which do not require the practitioner to assume any particular parametric model for volatility. Monte Carlo simulations confirm that our proposed detection procedures display well controlled false positive rate across a range of feasible volatility paths coupled with good power to rapidly detect an emerging predictive episode. The empirical relevance of our proposed monitoring strategy is illustrated in a pseudo real-time monitoring exercise using a well-known dataset of S&P 500 returns.

Key words: predictive regression; real-time monitoring; CUSUM; unknown regressor persistence; nonstationary volatility; kernel-based volatility estimation.

JEL classifications: C12, C22, G14.

Acknowledgments We gratefully acknowledge the computing time provided on the Linux HPC cluster at TU Dortmund University (LiDO3), partially funded in the course of the Large-Scale Equipment Initiative by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) as project 271512359. Demetrescu and Schmidt gratefully acknowledge financial support from the German Research Foundation (DFG) within TRR 391: Spatio-temporal Statistics for the Transition of Energy and Transport (520388526).

1 Introduction

The predictability of asset returns using publicly available data has received a great deal of attention in both the economics and finance literature, leading to a large number of published studies examining whether the lagged values of various financial and macroeconomic variables have in-sample predictive power for returns. Candidate financial predictor variables considered have included valuation ratios such as the dividend-price or earnings-price ratio, the dividend yield and a variety of interest rate measures. Macroeconomic variables such as inflation and industrial production have also been considered.

A common feature of many of the putative predictors employed in empirical studies is that they are both highly persistent and endogenous, with a strong negative correlation found between the innovations to the returns and those driving the predictor; see, in particular, Campbell and Yogo (2006) and Welch and Goyal (2008). For these highly persistent and endogenous predictors, basing inference on conventional tests can be misleading. As a result, a number of likelihood-based procedures have been developed for the case where the predictor is endogenous and displays strong persistence within the local-to-unity class of processes; see, in particular, Cavanagh *et al.* (1995), Campbell and Yogo (2006), Jansson and Moreira (2006) and Elliott *et al.* (2015). Excepting Elliott *et al.* (2015), a major practical drawback with these approaches is that they are not valid if the predictor is weakly persistent. An alternative approach is to base predictability tests on methods of estimating the predictive regression which are robust to the properties of the regressor. Arguably the most successful such approach is Kostakis *et al.* (2015) who estimate the predictive regression using the extended instrumental variable [IVX] procedure of Phillips and Magdalinos (2009). Here each predictor in the predictive regression has an associated stochastic instrument formed by constructing a mildly integrated variable from the first differences of the predictor. The IVX instrument, by construction, has lower persistence than a near-integrated variable and, as a consequence, delivers predictability statistics with (asymptotically pivotal) standard limiting null distributions.

The methods discussed above are designed to test for in-sample return predictability and are based on the assumption of a constant parameter predictive regression model. However, there are several reasons to suspect that if stock returns are predictable, then it is likely to be a time-varying phenomenon. For example, significant changes in monetary policy and financial regulations could lead to shifts in the relationship between macroeconomic variables and the fundamental value of stocks, via the impact of these changes on economic growth and the growth rates of earnings and dividends. Timmermann (2008) argues that for most time periods stock returns are not predictable, but that there are 'pockets in time' where evidence of local predictability is seen, such that if a variable begins to have predictive power for stock returns then a short window of predictability might exist before investors learn about the new relationship between that variable and returns; see, in particular, Paye and Timmermann (2006) and Timmermann (2008). Along the same lines, Henkel *et al.* (2011) find that return predictability in the stock market appears to be closely linked to economic recessions with dividend yield and term structure variables displaying predictive power only during recessions. It is then clear that to exploit this predictive relationship one would need to monitor returns for predictability in a real-time setting.

In a recent paper, Harvey *et al.* (2021) develop statistical methods designed to detect the emergence of an end-of-sample predictive episode in real-time. Their approach is based on the sequential application of simple end-of-sample rolling regression t-statistics for the significance of the predictor variable calculated over the last m (where the theory assumes m is some finite constant) data points. Because this monitoring approach is based on a sequence of subsample statistics they avoid the issue of spurious detections highlighted by Inoue and Rossi (2005), allowing the practitioner to control the overall false positive detection rate [FPR] of the procedure (defined as the probability of at least one test in the monitoring sequence rejecting when the null hypothesis of no predictability is true) by using information obtained from a *training period*. They consider two such procedures. The first signals a predictive regime if and when a t statistic in the monitoring period exceeds the relevant extremal value of the t statistic computed over the training period. The second signals a predictability regime if and when the number of consecutive rejections (at a given marginal significance level) by the sequence of t tests observed in the monitoring period exceeds the longest run of such rejections in the training period. To the best of our knowledge, Harvey *et al.* (2021) provides the only FPR-controlled real-time prediction monitoring procedures currently available in the literature. Importantly, however, the results in Harvey *et al.* (2021) are based on the assumption that the innovations in the predictive regression model are strictly stationary which therefore does not allow for the possibility of unconditional volatility changes in either the predictive regression errors or the innovations driving the predictor, or for time-varying behaviour in the endogeneity correlation.

The general subsample testing approach developed in Harvey *et al.* (2021) has also been applied to develop real-time monitoring procedures for the emergence of asset price bubbles by Astill *et al.* (2018). They compare these methods against the CUSUM-based bubble monitoring procedure developed by Homm and Breitung (2012) and find that the CUSUM procedure tends to signal an emerging bubble slightly faster on average than the subsample tests. Moreover, Astill *et al.* (2023) have recently shown that the CUSUM bubble detection procedure of Homm and Breitung (2012) can be generalised to allow for changes in the unconditional volatility of the innovations driving prices, by incorporating a kernel-based estimator of the volatility process. It therefore seems worth exploring whether monitoring procedures based on CUSUM-type statistics could also deliver earlier detection of emerging predictive episodes than the subsample approach of Harvey *et al.* (2021), while also being robust to time-varying unconditional volatility.

To that end, in this paper we develop methods based on CUSUM-type, and also on fluctuations-type statistics, in each case designed to monitor for the emergence of a predictive regime in real-time. To allow for uncertain persistence and endogeneity in the putative predictors, our proposed monitoring procedures build on the IVX methodology of Kostakis *et al.* (2015). This allows for unit root, local-to-unity, and weakly stationary predictors, as in Harvey *et al.* (2021), but also allows for locally explosive predictors. In common with the conditions placed on the model innovations by Harvey *et al.* (2021), our proposed approaches allow for very general forms of conditional heteroskedasticity. While the presence of conditional heteroskedasticity, such as GARCH, in a wide range of financial variables, including asset returns, is a well-established stylised fact, a number of recent empirical studies have also questioned the reasonableness of assuming unconditional homoskedasticity. In particular, strong evidence of structural breaks in the unconditional

variance of asset returns is reported in, *inter alia*, Rapach *et al.* (2008), McMillan and Wohar (2011), Calvo-Gonzalez *et al.* (2010) and Vivian and Wohar (2012). By incorporating simple nonparametric adjustment methods to account for a wide class of time-varying behaviour in the volatility matrix, our procedures allow for both time-varying unconditional volatility and time-varying endogeneity correlations, neither of which are allowed under the assumptions made by Harvey *et al.* (2021) in connection with their monitoring procedures.

A key advantage of using IVX in a real-time monitoring context is the availability of standard asymptotic results for partial sums of IVX moment conditions (see, *inter alia*, Demetrescu *et al.*, 2023). As we will show, this allows us to develop sequences of CUSUM- and fluctuation-type statistics that in the unconditionally homoskedastic case can be compared with the familiar asymptotic boundary functions of Chu *et al.* (1996), with exceedances signalling the rejection of the null of no predictive component to the returns series. To do so, rather than generically monitoring the cumulated returns, we monitor the behaviour of the partial sums of the IVX moment conditions under the null, which state that the stock returns should be orthogonal to the lagged instruments. This allows one to detect *specific* departures from the null of no predictability; in particular, predictability induced by the putative predictor(s) under consideration. Moreover, this approach can be readily adapted to allow for nonstationary volatility and time-varying endogeneity. This can be done in two ways, both of which we explore. The first, based on the theoretical crossing probabilities of a variance-transformed Brownian motion, uses a modification of the boundary function based on real-time estimation of the variance profile. This modification is implicitly motivated by the variance-transformation approach of Cavaliere and Taylor (2008) and is computationally simple to implement because, unlike the unit root testing setting of Cavaliere and Taylor, no inversion of an empirical variance profile is required. The second, motivated by the work of Beare (2018), adopts a local standardisation approach around kernel-based estimation of the volatility process, such that the conventional (homoskedastic) boundary function of Chu *et al.* (1996) can validly be used.

The remainder of the paper is organised as follows. Section 2 describes the predictive regime model we consider together with the assumptions we will work under. The conventional full-sample IVX predictability tests are briefly reviewed in section 3. Our proposed real-time monitoring procedures, based on IVX moment conditions, are introduced in section 4. Here we also establish their large sample validity. In section 5 we use Monte Carlo simulation to investigate the finite-sample performance of our proposed monitoring procedures and compare these with the monitoring procedures from Harvey *et al.* (2021). We demonstrate the usefulness of our proposed monitoring procedures with an application to a measure of the U.S. equity premium in section 6. Section 7 concludes. Proofs of our main theoretical results, together with additional results relating to our simulation study and empirical study are contained in an online supplementary appendix.

2 The Predictive Regime Model and Assumptions

Following Demetrescu *et al.* (2022) we consider an *episodic* predictive regression model for stock returns, y_t, allowing for potential time-variation in the slope coefficient on the (putative) predictor variable, of the form

$$y_t = \mu_y + \beta_t x_{t-1} + u_t, \qquad t = 1, \ldots, T, ..., E \ (E > T) \tag{1}$$

where the (putative) predictor, x_t, satisfies the data generating process [DGP]

$$x_t = \mu_x + \xi_t, \qquad t = 0, \ldots, E \tag{2a}$$
$$\xi_t = \rho\, \xi_{t-1} + w_t, \qquad t = 1, \ldots, E \tag{2b}$$

with ξ_0 a mean zero $O_p(1)$ variate. The innovations u_t are assumed to be a martingale difference [MD] sequence, while w_t is allowed to exhibit weak serial dependence. For lucidity we will outline our procedure for the case of a single putative predictor, x_t, assumptions on which will be given below. Generalisations to the case where the predictive regression contains multiple predictors will be discussed at the end of section 4.4.

The DGP in (1) allows the slope coefficient on x_{t-1} to vary over time, thereby allowing the predictive content of the regressor, x_{t-1}, to change over time. Our interest in this paper focuses on developing real-time monitoring procedures to detect an emerging predictive regime. To that end, we envisage a practitioner standing at time $t = T$ who wishes to start monitoring for predictability of y_t by x_{t-1}, in real-time, from time $T+1$ up until some time pre-defined point E in the future, where the horizon E is such that E/T converges to a fixed number, ω, with $\omega > 1$, as $T \to \infty$. Accordingly, we split the data into two periods: the *training period* for $t = 1, ..., T$, and the *monitoring period* for $t = T+1, ..., E$. For the purposes of our analysis, we will assume that no predictive regime occurs within the training period, such that $\beta_t = 0$ for $t = 1, ..., T$; further discussion relating to the consequences of violating this assumption is given in section 5.

As discussed in section 1, it is important to allow for the possibility that the (putative) predictor variable, x_t, is either strongly or weakly persistent, and to allow the shocks driving the predictor, w_t in (2b), to be correlated with the unpredictable component of stock returns, u_t in (1). As regards the latter, we will allow u_t and w_t to be contemporaneously correlated and heteroskedastic; exact conditions will be detailed later in this section. For the former, we let ρ in (2b) satisfy the following assumption.

Assumption 1 *Exactly one of the two following conditions holds true:*

1. ***Strongly persistent predictors:*** *The autoregressive parameter ρ in* (2b) *is local-to-unity with $\rho := 1 - \frac{c}{T}$ where c is a fixed constant.*

2. ***Weakly persistent predictors:*** *The autoregressive parameter ρ in* (2b) *is fixed and bounded away from unity, $|\rho| < 1$.*

Remark 1 Assumption 1 follows the bulk of the predictive regression literature in allowing the putative predictor, x_t, to be either near-integrated (Assumption 1.1) or to follow a stable (weakly dependent) process (Assumption 1.2). In the former case X_t is strongly persistent, exhibiting sums of sample autoregressive coefficients which are close to unity, allowing for pure unit root predictors, locally stable predictors, and locally explosive predictors, without assuming knowledge of which of these holds. Here, near-integrated asymptotics apply to the large sample behaviour of statistics from the predictive regression in (1); see, *inter alia*, Elliott and Stock (1994). In the latter case x_t is weakly persistent and standard asymptotic methods hold on statistics from (1). □

We will develop monitoring procedures for testing the null hypothesis that y_t is not predictable by x_{t-1} at any point during the monitoring period, which do not require the practitioner to know which of Assumption 1.1 or Assumption 1.2 holds in (2b), nor indeed what the precise value of ρ is in either case. Some structure needs to be placed on the alternative hypothesis and we follow Demetrescu *et al.* (2022) in this regard and specify β_t in (1) to satisfy the following assumption.

Assumption 2 *In the context of* (1)*, let* $\beta_t := b\left(t/T\right)$*, where* $b^*(\cdot)$ *is a piecewise Lipschitz-continuous real function on* $[0, E/T]$*.*

Using the framework of Assumption 2 we can then, under the maintained hypothesis of no predictability in the training period, write our null hypothesis of no predictability in the monitoring period, that $\beta_t = 0$, $T < t \leq E$, as

$$H_0: \text{ The function } b^*(\tau) \text{ is identically zero for all } \tau \in [0, E/T] \tag{3}$$

and formally specify the alternative hypothesis as,

$$H_{1,b^*(\cdot)}: \quad \text{The function } b^*(\tau) \text{ is identically zero for all } \tau \in [0,1], \text{ but is non-zero over at least one non-empty open interval contained in } (1, E/T]. \tag{4}$$

Remark 2 Under H_0, $(y_t - \mu_y)$ is a MD sequence and, hence, y_t is not predictable by x_{t-1}, in either the training or monitoring periods. Under $H_{1,b^*(\cdot)}$ at least one subset of the sample observations in the monitoring period (this need not be a strict subset, so it could potentially contain all of the observations in the monitoring period) comprising contiguous observations exists for which $\beta_t \neq 0$, and where the size of this subset is proportional to the training period size, T. Notice that, under $H_{1,b^*(\cdot)}$, $b^*(\cdot)$ may be zero in certain parts of its domain and it may also change magnitude and/or sign over its domain; the former corresponds to data points where $\beta_t = 0$, while the latter reflects observations for which β_t does not have a fixed magnitude and/or sign across the monitoring period. □

We conclude this section by detailing in Assumption 3 the conditions that we will place on the disturbances u_t and w_t in (1) and (2b), respectively.

Assumption 3 *Let the vector* $(u_t, w_t)'$, $t = 1, ..., T, ..., E$, *admit the decomposition,*

$$\begin{pmatrix} u_t \\ w_t \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & B(L) \end{pmatrix} \mathbf{H}\left(\frac{t}{T}\right) \boldsymbol{a}_t,$$

where:

1. $B(L) := \sum_{j\geq 0} b_j L^j$, *where* L *denotes the conventional lag operator, is a lag polynomial with* $b_0 = 1$ *and 1-summable coefficients, such that* $\sum_{j\geq 0} j\,|b_j| < \infty$

2. $\mathbf{H}(\cdot) := \begin{pmatrix} h_{11}(\cdot) & h_{12}(\cdot) \\ h_{21}(\cdot) & h_{22}(\cdot) \end{pmatrix}$ *is a matrix of piecewise Lipschitz-continuous bounded, non-stochastic functions on* $(-\infty, \omega]$, *invertible at all but a finite number of points*

3. $\boldsymbol{a}_t$ *is a uniformly* L_4*-bounded stationary and ergodic MD sequence satisfying* $\mathrm{E}\left(\boldsymbol{a}_t \boldsymbol{a}_t'\right) = \mathbf{I}_2$, *where* $\mathbf{I}_k$ *denotes the* $k \times k$ *identity matrix, and* $\mathrm{E}\left(\left\|\mathrm{E}_0\left(\sum_{t=1}^{E}(\boldsymbol{a}_t\boldsymbol{a}_t' - \mathbf{I}_2)\right)\right\|^2\right) = O(T^{2\epsilon})$ *for some* $\epsilon < \frac{1}{2}$, *with* $\mathrm{E}_0(\cdot)$ *denoting expectation conditional on the* σ*-algebra generated by* $\{\boldsymbol{a}_{-i}\}_{i=0}^{\infty}$.

Remark 3 Assumption 3 is very similar to Assumption 3 of Demetrescu *et al.* (2023), differing only in that Demetrescu *et al.* (2023) make the further assumption that $B(L)$ in Assumption 3.1, which determines the structure of the weak dependence in w_t, is the inverse of a finite-order autoregressive lag polynomial. This condition is used for the residual wild bootstrap implementation of the IVX predictive regression tests of Kostakis *et al.* (2015) they develop, but is not needed in the context of the methods we propose in this paper. By adopting essentially the same set of assumptions as Demetrescu *et al.* (2023) we are able to make use of the weak convergence results established in that paper which are needed to establish the asymptotic distribution of the IVX-based monitoring procedure we propose in section 4. As discussed in Demetrescu *et al.* (2023), Assumption 3 is much weaker than the assumptions under which Kostakis *et al.* (2015) work, in that we explicitly allow for both conditional heteroskedasticity of generic (rather than a specific parametric GARCH) form and unconditional heteroskedasticity. □

Remark 4 A detailed discussion of the conditions imposed by Assumption 3 is given in Demetrescu *et al.* (2023, pp. 6–7). Briefly, Assumption 3 allows for quite general forms of heteroskedasticity in $(u_t, v_t)' := \mathbf{H}(t/T)\boldsymbol{a}_t$ and, hence, in u_t and w_t. Assumption 3.2 allows for unconditional time heteroskedasticity in the innovations through the matrix $\mathbf{H}(\tau)$. Where $\mathbf{H}(\tau)$ is diagonal for all $\tau \in [0, \omega]$ the innovations $(u_t, v_t)'$ can display time-varying variances but are contemporaneously uncorrelated. Importantly, the off-diagonal elements of $\mathbf{H}(\tau)\mathbf{H}(\tau)'$ (*i.e.*, the covariance matrix of $(u_t, v_t)'$) are neither imposed to be zero nor to be independent of τ, thus allowing for contemporaneous and time-varying correlation among the innovations. The structure placed on $\mathbf{H}(\tau)$ by Assumption 3.2 allows for a wide class of models for the time-variation in the variance matrix of the innovations including single or multiple (co-) variance shifts, variances which follow a broken trend, and smooth transition variance shifts. As discussed in Breitung and Demetrescu (2015,

p. 360) such patterns are plausible with macro and financial data and it is therefore important to use tests which are robust to such behaviour to avoid the possibility of spurious rejection of the null because of non-constancy in the variance matrix rather than genuine predictability from x_{t-1}. Assumption 3 allows for cases where time-variation in the variance matrix occurs only in the training period, only in the monitoring period, or in both periods. The MD structure placed on $\boldsymbol{a}_t$ by Assumption 3.3 allows for conditional heteroskedasticity of a very general form obviating the need to choose a specific parametric model by instead adopting an explicit assumption of martingale approximability whereby $\mathrm{E}(\|\mathrm{E}_0(\sum_{t=1}^{E}(\boldsymbol{a}_t\boldsymbol{a}_t' - \mathbf{I}_2))\|^2) = O(T^{2\epsilon})$ for some $\epsilon < \frac{1}{2}$, where ϵ controls the degree of persistence permitted in the conditional variances. Stationary vector GARCH processes with finite fourth-order moments satisfy this condition with $\epsilon = 0$, although Assumption 3.3 is considerably more general as it also allows for asymmetric effects in the conditional variance. Stationary autoregressive stochastic volatility processes as, for example, are assumed in Johannes *et al.* (2014) are also permitted. □

3 Review of IVX Predictability Testing

Consider, for the present, the case where a practitioner standing at time T wishes to perform an *in-sample* test for predictability based on the training period, $t = 1, ..., T$. Under the maintained hypothesis that the slope parameter in (1) is constant across $t = 1, ..., T$, we obtain the standard constant parameter predictive regression for the training period

$$y_t = \mu_y + \beta x_{t-1} + u_t, \qquad t = 1, \ldots, T. \tag{5}$$

A number of procedures have been developed for testing the null hypothesis that $\beta = 0$ in (5) against the alternative that $\beta \neq 0$. Of these, the simplest is the standard ordinary least squares [OLS] Wald test for the significance of x_{t-1} in (5). While this test has a standard $\chi^2(1)$ limiting null distribution under Assumption 1.2, provided the errors are homoskedastic (although this can be weakened by using heteroskedasticity-robust standard errors), it does not under Assumption 1.1 where the limiting null distribution of the Wald statistic is nonstandard and depends on the local-to-unity parameter c, unless x_t is strictly exogenous with respect to u_t; see, for example, Campbell and Yogo (2006), and the references therein.

A popular and widely-applied solution to these inference problems in the predictive regression have been developed in Kostakis *et al.* (2015) who apply the IVX procedure of Phillips and Magdalinos (2009) to the estimation of (5). The basic idea underlying the IVX procedure is to instrument the regressor x_{t-1} by a variable of controlled persistence, constructed as

$$z_0 = 0 \quad \text{and} \qquad z_t = (1 - \varrho L)_+^{-1} \Delta x_t := \textstyle\sum_{j=0}^{t-1} \varrho^j \Delta x_{t-j}, t = 1, \ldots, T, \tag{6}$$

where $\Delta := (1 - L)$, and where $\varrho := 1 - aT^{-\eta}$ with $a > 0$ and $0 < \eta < 1$. The IVX scale and exponent parameters, a and η respectively, are tuning parameters set by the practitioner; Kostakis *et al.* (2015) recommend $a = 1$ and $\eta = 0.95$. Where x_t is near-

integrated satisfying Assumption 1.1, z_t is approximately a mildly integrated process and therefore of lower persistence than x_t. Moreover, where x_t is weakly dependent satisfying Assumption 1.2, we have that $z_t \approx \xi_t$.

Based on the training period data, the IVX-based Wald statistic (using Eicker-White standard errors) for testing the null hypothesis that $\beta = 0$ in (5) such that there is no (in-sample) predictability, is then defined as,

$$\mathcal{W} := \frac{\left(\sum_{t=1}^{T} z_{t-1}\left(y_t - \bar{y}_T\right)\right)^2}{\sum_{t=1}^{T} z_{t-1}^2 \hat{u}_t^2} \tag{7}$$

with $\bar{y}_T := \frac{1}{T}\sum_{t=1}^{T} y_t$. For $\hat{u}_t$, Kostakis *et al.* (2015) suggest using the OLS residuals from the estimation of (5). An asymptotically equivalent (under both the null and local alternatives) choice for $\hat{u}_t$, that we will use in the procedures we develop in this paper, is to use the OLS residuals obtained under the null; that is, $\hat{u}_t := y_t - \bar{y}_T$. As discussed in Kostakis *et al.* (2015, p. 1514) the IVX instrument, z_{t-1}, does not need to be demeaned because the slope estimator in the predictive regression is invariant to whether z_{t-1} is demeaned or not. Demetrescu *et al.* (2023) demonstrate that, under Assumption 3, the limiting null distribution of $\mathcal{W}$ is $\chi^2(1)$, regardless of whether Assumption 1.1 or 1.2 holds.

Of most relevance to the testing problem considered in this paper, Demetrescu *et al.* (2023) develop subsample implementations of the full-sample predictability tests of Kostakis *et al.* (2015), while Demetrescu *et al.* (2022) develop subsample implementations of the full-sample 2SLS-based predictability tests of Breitung and Demetrescu (2015). To illustrate, and denoting the integer part of a by $\lfloor a \rfloor$, suppose it was known that a pocket of predictability might occur over the subsample $t = \lfloor \tau_1 T \rfloor + 1, \ldots, \lfloor \tau_2 T \rfloor$, $0 \leq \tau_1 < \tau_2 \leq 1$, then it would be logical to compute the subsample IVX statistic

$$\mathcal{W}(\lfloor \tau_1 T \rfloor, \lfloor \tau_2 T \rfloor) := \frac{\left(\sum_{t=\lfloor \tau_1 T \rfloor + 1}^{\lfloor \tau_2 T \rfloor} z_{t-1}\left(y_t - \bar{y}_{\tau_1,\tau_2}\right)\right)^2}{\sum_{t=\lfloor \tau_1 T \rfloor + 1}^{\lfloor \tau_2 T \rfloor} z_{t-1}^2 (y_t - \bar{y}_{\tau_1,\tau_2})^2} \tag{8}$$

where $\bar{y}_{\tau_1,\tau_2} := (\lfloor \tau_2 T \rfloor - \lfloor \tau_1 T \rfloor)^{-1} \sum_{s=\lfloor \tau_1 T \rfloor + 1}^{\lfloor \tau_2 T \rfloor} y_s$. Demetrescu *et al.* (2023, Proposition 2, p. 12) demonstrate that, under Assumption 3, $\mathcal{W}(\lfloor \tau_1 T \rfloor, \lfloor \tau_2 T \rfloor)$ has a $\chi^2(1)$ limiting null distribution, regardless of whether Assumption 1.1 or 1.2 holds on ρ, and so a test for predictability in this specific subsample with asymptotic size α could be obtained by comparing $\mathcal{W}(\lfloor \tau_1 T \rfloor, \lfloor \tau_2 T \rfloor)$ with the $(1-\alpha)100\%$ critical value from the $\chi^2(1)$ distribution, say cv_α. In practice, it is unlikely the practitioner would know which specific subsample(s) of the data might admit predictive regimes, and so Demetrescu *et al.* (2022) and Demetrescu *et al.* (2023) develop asymptotically size-controlled tests based on the maxima of rolling, forward recursive, and backward recursive sequences of the $\mathcal{W}(\cdot,\cdot)$ statistics which do not assume knowledge of the location of any predictive episodes which might exist in the data. It is important to note, however, that these are one-shot in-sample predictability tests and not out-of-sample monitoring procedures.

In the predictive regression context, the only FPR controlled real-time monitoring approach in the literature we are aware of which is theoretically valid without knowing

which of Assumption 1.1 or 1.2 holds on the predictor has been developed in Harvey *et al.* (2021).[1] They base their monitoring procedure on the sequential application of rolling implementations of a standard regression t-statistic for the significance of x_{t-1} in (1). Their leading method delivers a procedure with a theoretically controlled FPR by rejecting H_0 if a t-statistic in the monitoring period exceeds the largest of the analogous t-statistics calculated from the training period. The approach of Harvey *et al.* (2021) is based on the assumption that w_t and u_t are both strictly stationary processes, which therefore does not allow for unconditional heteroskedasticity of the form specified in Assumption 3.

In the literature on asset price bubbles, Homm and Breitung (2012) develop two real-time monitoring procedures for explosive episodes. The first, labelled FLUC by Homm and Breitung (2012), is based on the sequential application of subsample ADF statistics. In the predictive regression context considered here, the analogue of FLUC based on IVX Wald statistics would be a procedure that rejected H_0 of (3) in favour of $H_{1,b^*(\cdot)}$ of (4) as soon as $\mathcal{W}(1,t)$, $t > T$, exceeded a given critical value. The second applies the methodology of Chu *et al.* (1996), using a CUSUM statistic applied to the first difference of prices (*i.e.*, simple returns) calculated sequentially across the monitoring period with a decision rule designed to control the theoretical FPR of the procedure. This decision rule uses a critical value which increases over the monitoring period, obtained from the boundary crossing probabilities of a Brownian motion process. While Homm and Breitung (2012) assume the innovations driving asset returns are homoskedastic, Astill *et al.* (2023) adapt the CUSUM approach to allow for conditionally and unconditionally heteroskedastic innovations, using a nonparameteric variance estimator.

In the next section we will develop analogues of the FLUC and CUSUM monitoring approaches from Homm and Breitung (2012) to obtain real-time monitoring procedures for predictability based on sequential statistics that have theoretically controlled FPRs. We will show how these sequential statistics can be self-normalised using the sample quadratic variation (analogously to the IVX Wald statistics) such that they can be validly applied without knowing which of Assumption 1.1 or Assumption 1.2 holds on the predictor, and how they can be implemented such that the theoretical FPR remains controlled in the presence of nonstationary volatility.

[1] In the case where the predictor is strictly stationary, the forecast monitoring approach of Inoue and Rossi (2005) could be used to achieve FPR control. Inoue and Rossi develop sequential testing procedures of parameter stability in out-of-sample forecasting setups, which amount to the use of specific boundary functions for Wald, LM and LR type tests in a GMM framework. Unfortunately, it is not clear from their work how to choose operational critical values such that one would obtain a monitoring procedure with a theoretically controlled FPR in our setup with time-varying volatility and possibly strongly persistent predictors.

4 Real-Time Predictability Monitoring

4.1 A FLUC-type Monitoring Procedure

To develop our first real-time predictability monitoring procedure, in the spirit of the FLUC approach of Homm and Breitung (2012), we consider an approach based on the sequential application of the Wald statistic, $\mathcal{W}(1,t)$, outlined in section 3 (which for simplicity we will denote henceforth by $\mathcal{W}_t$). Specifically, we will consider a monitoring approach based on the sequence of statistics,

$$\mathcal{W}_t := \frac{\left(\sum_{j=1}^{t} z_{j-1}\left(y_j - \bar{y}_j\right)\right)^2}{\sum_{j=1}^{t} z_{j-1}^2\left(y_j - \bar{y}_j\right)^2} \qquad \text{for } T+1 \leq t \leq E, \tag{9}$$

where $\bar{y}_j := j^{-1}\sum_{s=1}^{j} y_s$, and in the computation of the IVX instrument, z_t, in (6) the time index is extended through to E.

While this is the natural extension of the full-sample Wald predictability statistic to a monitoring setup, some nontrivial adjustments are required to ensure FPR control under the null of no predictability when allowing for time-varying volatility and uncertain persistence in the (putative) predictor, x_t. If both u_t and x_t were known to be strictly stationary then, following Inoue and Rossi (2005), one could achieve a controlled FPR, α, by comparing $\mathcal{W}_t$ with asymptotic (upper) bounds given by

$$b_t^2 = a_\alpha^2 + \ln(t/T) \tag{10}$$

where a_α^2 depends on the desired FPR α via $\alpha := 2\left(1 - \Phi(a_\alpha) + a_\alpha\phi(a_\alpha)\right)$. The use of IVX estimation will be shown to side-step the issue that x_t may be strongly persistent. However, accounting for time-varying volatility in the predictive regression poses an additional challenge, as we now discuss.

The limiting boundary function in (10) is derived from the crossing probabilities for standard Brownian motion processes; see, for example, Chu *et al.* (1996, p. 1052). As such, it is appropriate only where the sequence of statistics in (9) has a (squared) Brownian motion limit under the no predictability null hypothesis. As the results we now present in Lemma 1 show, this is not the case under Assumptions 1–3, and so comparison of the sequence of $\mathcal{W}_t$ statistics in (9) with the bounds in (10) will not, in general, deliver a monitoring procedure with a FPR of α.

Lemma 1 *Let Assumptions 1–3 hold. Then, uniformly in $\tau \in [0,\omega]$, for some $\omega > 1$, we have that, as $T \to \infty$,*

(i) Under Assumption 1.1 (strong persistence), with $\epsilon < \min\{1-\eta; \eta/2\}$ in Assumption 3, and defining $\bar{u}_j := j^{-1}\sum_{s=1}^{j} u_s$,

$$\frac{1}{T^{1/2+\eta/2}}\sum_{j=1}^{\lfloor \tau T \rfloor} z_{j-1}\left(u_j - \bar{u}_j\right) \Rightarrow M_1(\tau), \qquad \frac{1}{T^{1+\eta}}\sum_{j=1}^{\lfloor \tau T \rfloor} z_{j-1}^2\left(u_j - \bar{u}_j\right)^2 \Rightarrow \left[M_1\right](\tau)$$

(ii) Under Assumption 1.2 (weak persistence),

$$\frac{1}{\sqrt{T}}\sum_{j=1}^{\lfloor \tau T \rfloor} z_{j-1}\left(u_j - \bar{u}_j\right) \Rightarrow M_2(\tau), \qquad \frac{1}{T}\sum_{j=1}^{\lfloor \tau T \rfloor} z_{j-1}^2\left(u_j - \bar{u}_j\right)^2 \Rightarrow [M_2](\tau),$$

with the convergence results in (i) and (ii) holding jointly with

$$\frac{1}{\sqrt{T}}\sum_{j=1}^{\lfloor \tau T \rfloor} \begin{pmatrix} u_j \\ v_j \end{pmatrix} \Rightarrow \begin{pmatrix} M_u(\tau) \\ M_v(\tau) \end{pmatrix}$$

where "$\Rightarrow$" denotes weak convergence in the space of càdlàg functions, and where $M_1(\tau)$, $M_2(\tau)$, $M_u(\tau)$ and $M_v(\tau)$ are (scaled) variance-transformed Brownian motions and where, for a generic stochastic process $S(\tau)$, $[S](\tau)$ denotes the quadratic variation process of $S(\tau)$. Formal definitions of $M_1(\tau)$, $M_2(\tau)$, $M_u(\tau)$ and $M_v(\tau)$ are provided in section S.1.9 of the supplementary appendix.

For future reference, we note that the quadratic variation processes $[M_1](\tau)$, $[M_2](\tau)$, $[M_u](\tau)$ and $[M_v](\tau)$ are differentiable almost everywhere (and nonlinear in general), and we denote their (piecewise) derivatives by $\sigma_1^2(\tau)$, $\sigma_2^2(\tau)$, $\sigma_u^2(\tau)$ and $\sigma_v^2(\tau)$, respectively. Furthermore, under strong persistence, a consequence of Lemma 1 is that

$$\frac{1}{\sqrt{T}}\xi_{\lfloor \tau T \rfloor} \Rightarrow J_{c,\mathrm{H}}(\tau) := B(1)\int_0^\tau e^{-c(\tau-s)}\mathrm{d}M_v(s) \tag{11}$$

where $B(1)$ is the lag polynomial from Assumption 3.1 evaluated at 1.

Remark 5 A straightforward consequence of the results in Lemma 1 is that, regardless of whether the predictor is weakly or strongly persistent, under the null hypothesis (so that $y_t - \bar{y}_t = u_t - \bar{u}_t$), the sequence of $\mathcal{W}_t$ statistics in (9) does not converge to the square of a standard Brownian motion process. Rather, it converges to the (normalised) square of a Gaussian process with a nonlinear *variance profile*, where we define the variance profile of a generic stochastic process $S(\tau)$, $\tau \in [0, \omega]$, as $[S](\tau)/[S](1)$, collapsing to the square of a standard Brownian motion process only in the case where the covariance matrix of $(u_t, v_t)'$ is constant. Moreover, the form of the limiting Gaussian process involved depends on whether the predictor is weakly or strongly persistent; in particular, it is a function of $M_1(\tau)$ in the strongly persistent case and of $M_2(\tau)$ in the weakly persistent case. As noted in Demetrescu *et al.* (2023, p. 8), the variance profiles of $M_1(\tau)$ and $M_2(\tau)$ do not, in general, coincide under Assumption 3. □

Remark 6 Interestingly, although the partial sums considered in Lemma 1 include demeaned components, their limiting behaviour can still be expressed in terms of Brownian processes rather than in terms of the commonly encountered Brownian bridge processes in such contexts (see, for example, Ploberger and Krämer, 1992). This is due to the fact that the demeaned u_j do not enter the partial sums directly, but rather are combined *multiplicatively* with the instruments z_{j-1}, which are, however, centred at zero, by construction. □

Although, the results in Lemma 1 imply that we cannot obtain a controlled FPR in the presence of unconditional heteroskedasticity by comparing the sequence of $\mathcal{W}_t$ statistics in (9) with the asymptotic bounds in (10), they do, however, hold the key to obtaining a monitoring procedure that does have a controlled FPR in unconditionally heteroskedastic environments. To that end, notice first from Cavaliere (2004), and the references therein, that we can equivalently represent the processes $M_1(\tau)$ and $M_2(\tau)$ as

$$M_i(\tau) \stackrel{d}{=} \sqrt{[M_i](1)} \cdot W_i\left(\pi_i(\tau)\right), \quad i = 1, 2, \tag{12}$$

where $\pi_i(\tau) := \int_0^\tau \sigma_i^2(s)ds / \int_0^1 \sigma_i^2(s)ds$, $i = 1, 2$, denotes the variance profile of M_1 and M_2, respectively, and where W_1 and W_2 are standard Brownian motion processes. As noted in Remark 5, $\pi_1(\tau)$ and $\pi_2(\tau)$ will not, in general, coincide. Although $W_1(\tau)$ and $W_2(\tau)$ are not the same objects, their characterisation as standard Brownian motion processes suffices for our purposes and so, provided no confusion arises, we will use $W(\tau)$ generically to denote both.

Under the conditions of Lemma 1 and the no predictability null hypothesis, H_0 of (3), we therefore have from the results in Lemma 1, and using (12), that

$$\mathcal{W}_{\lfloor \tau T \rfloor} \Rightarrow \frac{M_i^2(\tau)}{[M_i](\tau)} \stackrel{d}{=} \frac{[M_i](1)\, W^2\left(\pi_i(\tau)\right)}{[M_i](\tau)} = \frac{1}{\pi_i(\tau)} W^2\left(\pi_i(\tau)\right). \tag{13}$$

on any interval $[\varepsilon, \omega]$ for $0 < \varepsilon < \omega$. The limiting null behaviour of the sequence of $\mathcal{W}_t$ statistics is therefore seen to depend on the persistence of x_t but, crucially, because of the implicit self-normalization of $\mathcal{W}_t$, does so only through the variance profile. The large sample behaviour of $\mathcal{W}_t$ is seen to be driven by its numerator, which converges to either $M_1^2(\tau)$ or $M_2^2(\tau)$, depending on whether x_t is strongly or weakly persistent. In order to derive a monitoring procedure with a controlled FPR, we must first determine an expression for the crossing probabilities of a variance-transformed Brownian motion. This we now provide in Lemma 2.

Lemma 2 *Let Assumptions 1–3 hold, and consider the variance-transformed Brownian motions $M_1(\cdot)$ and $M_2(\cdot)$ from Lemma 1 extended to the real positive half-line. Then, for $i = 1, 2$ and any $\omega > 1$,*

$$\mathrm{P}\left(M_i^2(\tau) > [M_i](\tau)\left(a_\alpha^2 + \ln \pi_i(\tau)\right) \text{ for some } 1 \leq \tau \leq \omega\right) \leq 2\left(1 - \Phi(a_\alpha) + a_\alpha \phi(a_\alpha)\right)$$

where ϕ and Φ are the pdf and cdf of the standard normal distribution.

Using the result in Lemma 2, which provides an asymptotic upper bound for the exceedance probability, it is seen that we can account for time-varying volatility in a relatively straightforward manner. Specifically, the result in Lemma 2 coupled with (13) implies that a monitoring procedure with a controlled FPR can be obtained by comparing $\mathcal{W}_t$ with the *variance-transformed* boundary function,

$$b_{t,\pi}^2 = a_\alpha^2 + \ln \pi\left(\frac{t}{T}\right), \qquad t > T, \tag{14}$$

where $\pi(\cdot) = \pi_1(\cdot)$ under strong persistence and $\pi(\cdot) = \pi_2(\cdot)$ under weak persistence. Notice that (14) reduces to the boundary function in (10) under unconditional homoskedasticity, as would be expected, given that $\pi(t/T) = t/T$ in this case.

Notice therefore that there is no need to variance-transform the data to account for time-varying volatility, unlike in the unit root testing scenario considered in Cavaliere and Taylor (2008), or to resort to a wild bootstrap implementation as, for example, in Harvey *et al.* (2016). All that is required is a consistent estimate of the variance profile (together with numerical computation of a_α; see, *inter alia*, Chu *et al.*, 1996; Inoue and Rossi, 2005). At first sight this might appear to be infeasible in that it would seem to require knowledge of whether x_t is weakly or strongly persistent, given that $\pi_1(\cdot)$ and $\pi_2(\cdot)$ do not, in general, coincide; see Remark 5. However, this issue can be avoided if we can find an estimate that converges in large samples to $\pi_1(\cdot)$ under strong persistence and to $\pi_2(\cdot)$ under weak persistence. This can be achieved by computing, as you go, an estimate formed from the ratio of (unsmoothed) estimates of the quadratic variation at times t and T,

$$\hat{\pi}\left(\frac{t}{T}\right) := \frac{\sum_{j=1}^{t} z_{j-1}^2 \left(y_j - \bar{y}_j\right)^2}{\sum_{j=1}^{T} z_{j-1}^2 \left(y_j - \bar{y}_j\right)^2}, \quad t = T+1, T+2, \ldots. \tag{15}$$

The uniform consistency of this estimator on $\tau \in [0, \omega]$, under the no predictability null hypothesis, follows directly from the results in Lemma 1, and holds regardless of whether the predictor is weakly or strongly persistent; we will subsequently show in section 4.3 that this result also holds under local alternatives. A feasible version of the bounds in (14), can then be obtained by using

$$b_{t,\hat{\pi}}^2 = a_\alpha^2 + \ln \hat{\pi}\left(\frac{t}{T}\right), \quad t > T. \tag{16}$$

Notice that the constant term a_α in (14) coincides with that in (10), and so the tabulated values for this constant given in Table 1 of Inoue and Rossi (2005) can also be used in evaluating $b_{t,\hat{\pi}}^2$.

The key implication of the foregoing results is now provided in Proposition 1 which establishes the asymptotic validity of the monitoring procedure based on the sequence of FLUC-type statistics, $\mathcal{W}_t$, in the sense that it has an FPR which, for large T, can be controlled by the practitioner at a given level, α, regardless of whether the predictor is weakly or strongly persistent.

Proposition 1 *Let Assumptions 1–3 hold. Then under the no predictability null hypothesis, H_0 of* (3)*, we have that*

$$\lim_{T\to\infty} \mathrm{P}\left(\mathcal{W}_t \geq b_{t,\hat{\pi}}^2 \text{ for some } T+1 \leq t \leq E\right) \leq \alpha.$$

4.2 A CUSUM-type Monitoring Procedure

Our second approach to developing a real-time monitoring procedure for predictability is based on the well-known CUSUM approach for detecting parameter change in a regression

model. In particular, we will consider a monitoring procedure based on the sequence of CUSUM-type statistics,

$$\mathcal{M}_t = \frac{\left(\sum_{j=T+1}^{t} z_{j-1}\left(y_j - \bar{y}_j\right)\right)^2}{\sum_{j=1}^{T} z_{j-1}^2\left(y_j - \bar{y}_j\right)^2}, \quad T+1 \leq t \leq E. \tag{17}$$

The statistic $\mathcal{M}_t$ can be interpreted as an Anderson-Rubin-type statistic for the moment condition $\mathrm{E}\left(z_{t-1}u_t\right) = 0$, of the kind employed in the predictive regression setting by Breitung and Demetrescu (2015) for one-shot testing. It is also worth noting that although we refer to the $\mathcal{M}_t$ statistic as a CUSUM statistic, it differs in two significant ways from a conventional CUSUM statistic. First, while the latter is designed to detect parameter change in the underlying regression model during the monitoring period, $\mathcal{M}_t$ is specifically designed to detect changes in the coefficient on x_{t-1} in the predictive regression in (1) from zero during the monitoring period; as such, it will in theory reject if β_t is constant, but non-zero, across the monitoring period. Second, conventional CUSUM statistics are based around recursive residuals, whereas OLS residuals are used in constructing $\mathcal{M}_t$; cf. Remark 6.

There are two key differences between the sequence of $\mathcal{M}_t$ statistics in (17) and the sequence of $\mathcal{W}_t$ statistics in (9). First, the starting point of the cumulative sum in the numerator of $\mathcal{M}_t$ is $T+1$ and, second, the $\mathcal{M}_t$ statistics are normalised using only data from the training period. In contrast, both the numerator and denominator of the $\mathcal{W}_t$ statistics mix data from the training and monitoring periods. The $\mathcal{M}_t$ statistics therefore offer a much greater degree of separation between the training and monitoring period - this separation would be complete if the mean estimate, $\bar{y}_j$, in the numerator of $\mathcal{M}_t$ was based only on data from the monitoring period; see the later discussion in Remark 11. An implication of this is that the FLUC-type method is not able to distinguish between in-sample and out-of-sample predictability: any systematic deviation from the null that occurs in the training period will affect monitoring based on $\mathcal{W}_t$. As a result, the FLUC procedure may reject whenever a pocket of predictability exists in the training period even if no predictability is present in the monitoring period. We will quantify the impact of the presence of a predictive episode in the training sample on the monitoring procedures based on $\mathcal{W}_t$ and $\mathcal{M}_t$ in section 5. These results confirm the foregoing intuition that $\mathcal{W}_t$ does not display a controlled FPR when predictability is present in the training period. Some theoretical discussion relevant to this issue will also be given following Lemma 4 which characterises the behaviour of the sequences of monitoring statistics under local alternatives.

In order to develop the relevant asymptotic boundary function, such that a real-time monitoring procedure based on the sequence of $\mathcal{M}_t$ statistics in (17) with a controlled FPR can be obtained, we first note, using the results in Lemma 1, that the limiting distribution of this sequence of statistics under the null, H_0 of (3), is given by

$$\mathcal{M}_{\lfloor \tau T \rfloor} \Rightarrow \frac{\left(M_i(\tau) - M_i(1)\right)^2}{\left[M_i\right](1)}, \qquad 1 \leq \tau \leq \omega, \tag{18}$$

where it is recalled that $i = 1$ corresponds to the the strongly persistent case (Assumption

1.1) and $i = 2$ to the weakly persistent case (Assumption 1.2).

Observe that the limiting null distribution in (18) is different from that of the sequence of $\mathcal{W}_t$ statistics given in (13), but like the latter also depends on which of Assumptions 1.1 and 1.2 holds through the differing variance profiles, $\pi_1(\cdot)$ and $\pi_2(\cdot)$. As a consequence, the boundary function in (14) is not appropriate for use in connection with the sequence of statistics in (17); in particular, it cannot be used as the basis for delivering a monitoring procedure with FPR controlled at α.

Under unconditional homoskedasticity the limit in (18) reduces to $(W(\tau)-W(1))^2$ (where it is recalled that $W(\cdot)$ generically denotes a standard Brownian motion), regardless of whether Assumption 1.1 or Assumption 1.2 holds. This limiting distribution coincides with (the square of) that given for the CUSUM-type sequence of statistics used to monitor for a price bubble in Homm and Breitung (2012, p. 211). Consequently, we have from Chu *et al.* (1996) that comparison with the boundary function

$$c_t^2 = \frac{t}{T}\left(\tilde{a}_\alpha + \ln\left(\frac{t}{T}\right)\right), \quad t > T \tag{19}$$

where $\tilde{a}_\alpha$ is such that $\alpha = e^{-\tilde{a}_\alpha/2}$ will yield a monitoring procedure with FPR α.

Analogously to Lemma 2, in order to develop a monitoring procedure with a controlled FPR under unconditional heteroskedasticity we first need to develop an expression for the crossing probabilities of the (heteroskedastic) Gaussian limiting processes featuring in (18). We now state this in Lemma 3.

Lemma 3 *Let Assumptions 1–3 hold, and consider the variance-transformed Brownian motions M_1 and M_2 from Lemma 1 extended to the real positive half-line. Then, for $i = 1, 2$, and any $\omega > 1$,*

$$\mathrm{P}\left(\left(M_i(\tau) - M_i(1)\right)^2 > [M_i](1)\,\pi_i(\tau)\left(\tilde{a}_\alpha + \ln \pi_i(\tau)\right) \text{ for some } 1 \le \tau \le \omega\right) \le e^{-\tilde{a}_\alpha/2}.$$

By analogy to the development in section 4.1 for the W_t sequence of statistics, the result in Lemma 3 implies that we can develop a feasible monitoring procedure by comparing $\mathcal{M}_t$ with the (upper) bounds given by

$$c_{t,\hat{\pi}}^2 = \hat{\pi}\left(\frac{t}{T}\right)\left(\tilde{a}_\alpha + \ln \hat{\pi}\left(\frac{t}{T}\right)\right), \quad t > T \tag{20}$$

where $\hat{\pi}(\cdot)$ is as defined in (15) and where $\tilde{a}_\alpha = -2\ln\alpha$. Doing so ensures that, for large T, the monitoring procedure will have a FPR controlled at level α, regardless of whether the predictor is weakly or strongly persistent. This result is now formally stated in Proposition 2.

Proposition 2 *Let Assumptions 1–3 hold. Then under the no predictability null hypothesis, H_0 of (3), we have that*

$$\lim_{T\to\infty} \mathrm{P}\left(\mathcal{M}_t \ge c_{t,\hat{\pi}}^2 \text{ for some } T+1 \le t \le E\right) \le \alpha.$$

Remark 7 Although we have outlined our monitoring procedures in terms of testing against two-sided alternatives (positive or negative predictability) of the form specified in $H_{1,b(\cdot)}$ of (4), we may also wish to test against the one-sided alternatives that $\beta_t > 0$ ($\beta_t < 0$) in some subset(s) of the data can be considered simply by defining $b(\cdot)$ to be a non-negative (non-positive) function. Several studies have found that imposing an economically motivated structure, such as a known slope sign, on the predictive regression model can lead to better and more accurate findings of return predictability; for example, Campbell and Thompson (2008) find that, among other things, imposing positive predictability (so that the sign of the predictor is imposed to be positive under the alternative) almost always improves the out-of-sample predictability obtained for the predictors considered for equity returns in Welch and Goyal (2008). Monitoring against one-sided alternatives can be easily implemented by simple modifications of the FLUC and CUSUM approaches outlined above, given the symmetry of the limiting null distributions of both. Taking the monitoring procedure based on $\mathcal{M}_t$ to illustrate, we would compare the sequence of statistics

$$\frac{\sum_{j=T+1}^{t} z_{j-1}\left(y_j - \bar{y}_j\right)}{\sqrt{\sum_{j=1}^{T} z_{j-1}^2 \left(y_j - \bar{y}_j\right)^2}}, \quad T+1 \leq t \leq E$$

with the (one-sided) upper or lower bound, as appropriate, given by $\pm|c_{t,\hat{\pi}}|$ (where these are obtained from the two-sided band outlined as above, but for a nominal level of 2α). The same applies to the procedures outlined below in section 4.4. □

4.3 Behaviour under Local Alternatives

We next turn our attention to an analysis of the large sample behaviour of the $\mathcal{W}_t$ and $\mathcal{M}_t$ sequences of statistics under departures from the no predictability null hypothesis. Predictive regressions for stock returns typically exhibit small R^2 and low signal-to-noise ratios (see, *inter alia*, Campbell, 2008, Phillips and Lee, 2013, and Phillips, 2015) and so departures from the null, should predictability be present, are likely to be small. We will therefore conduct our theoretical analysis of the large sample properties of $\mathcal{W}_t$ and $\mathcal{M}_t$ under local alternatives such that the slope parameter β_t in (1) is local-to-zero for an asymptotically non-vanishing set of the sample observations. The localisation rate (or Pitman drift) is such that β_t is specified to lie in a neighbourhood of zero which shrinks with the training sample size, T, at a rate which depends on which of Assumptions 1.1 and 1.2 holds in (1).

Accordingly, we will analyse the large sample behaviour of the $\mathcal{W}_t$ and $\mathcal{M}_t$ sequences of statistics under local alternatives of the form

$$H_{1,\eta_T,b(\cdot)} : \beta_t = \frac{1}{n_T} b(\cdot) \tag{21}$$

where the function $b(\cdot)$ is a piecewise Lipschitz-continuous real function on $[0, E/T]$ which is non-zero over at least one non-empty open interval contained in $[0, E/T]$, and the scaling constant, n_T, depends on the degree of persistence of the predictor such that $n_T = \sqrt{T}$ under weak persistence, and $n_T = T^{1/2+\eta/2}$ under strong persistence.

Remark 8 The different localisation rates in $H_{1,\eta_T,b(\cdot)}$ under weak and strong persistence reflect the fact that near-integration implies a much stronger signal from the predictor x_{t-1} than when x_{t-1} is weakly persistent. The function $b(\cdot)$ corresponds to those subsets of contiguous observations for which the slope coefficient β_t in the predictive regression in (1) is non-zero, such that the lagged regressor x_{t-1} has predictive power. This need not be a strict subset, so it could contain all of the observations in the training and monitoring periods, and nor does the function $b(\cdot)$ have to be constant over these subsets. Notice also that if $b(s)$ is identically zero for all $s \in [0,1]$, then it coincides with the function $b^*(\cdot)$ defined in $H_{1,b^*(\cdot)}$ of (4), which imposes the maintained hypothesis that no predictive regimes exist within the training period. We will derive limiting power results under $H_{1,\eta_T,b(\cdot)}$ rather than $H_{1,b^*(\cdot)}$ as it allows us to examine what, if any, impact predictability in the training sample has on the large sample behaviour of the $\mathcal{W}_t$ and $\mathcal{M}_t$ sequences of statistics. □

In Lemma 4, we next establish the large sample behaviour of the $\mathcal{W}_t$ and $\mathcal{M}_t$ sequences of statistics under the local alternative, $H_{1,\eta_T,b(\cdot)}$ of (21).

Lemma 4 *Let Assumptions 1–3 hold. Then under* $H_{1,\eta_T,b(\cdot)}$ *of* (21)*, as* $T \to \infty$*, uniformly for* $\tau \in [1,\omega]$*, it holds that:*

(i) Under Assumption 1.1 (strong persistence),

(a) $\mathcal{W}_{\lfloor \tau T \rfloor} \Rightarrow \frac{1}{[M_1](\tau)} \left(M_1(\tau) + \sqrt{\frac{2B(1)^2}{a}} \left(\int_0^\tau b(s) \mathrm{d}[M_v](s) + \int_0^\tau \overline{b\, J}^{rec}(s) \mathrm{d}J_{c,\mathrm{H}}(s) \right) \right)^2$*, with* $\overline{b\, J}^{rec}(\tau) := b(\tau) J_{c,\mathrm{H}}(\tau) - \frac{1}{\tau} \int_0^\tau b(s) J_{c,\mathrm{H}}(s) \mathrm{d}s$*, the recursively demeaned product of* $b(\tau)$ *and* $J_{c,\mathrm{H}}(\tau)$*, and*

(b) $\mathcal{M}_{\lfloor \tau T \rfloor} \Rightarrow \frac{1}{[M_1](1)} \left(M_1(\tau) - M_1(1) + \sqrt{\frac{2B(1)^2}{a}} \left(\int_1^\tau b(s) \mathrm{d}[M_v](s) + \int_1^\tau \overline{b\, J}^{rec}(s) \mathrm{d}J_{c,\mathrm{H}}(s) \right) \right)^2$

(ii) Under Assumption 1.2 (weak persistence),

(a) $\mathcal{W}_{\lfloor \tau T \rfloor} \Rightarrow \frac{1}{[M_2](\tau)} \left(M_2(\tau) + \kappa^2 \int_0^\tau b(s) \mathrm{d}[M_v](s) \right)^2$*, where* κ^2 *denotes the sum of the squared coefficients of the filter* $(1-\rho L)^{-1} B(L)$*, and*

(b) $\mathcal{M}_{\lfloor \tau T \rfloor} \Rightarrow \frac{1}{[M_2](1)} \left(M_2(\tau) - M_2(1) + \kappa^2 \int_1^\tau b(s) \mathrm{d}[M_v](s) \right)^2$.

Remark 9 A by-product of the proof of Lemma 4 is that the uniform consistency of $\hat{\pi}(\cdot)$ for $\pi_1(\cdot)$ under strong persistence and for $\pi_2(\cdot)$ under weak persistence, shown previously to hold under the null hypothesis, H_0 of (3), also holds under $H_{1,\eta_T,b(\cdot)}$. Consequently, the estimated boundary functions for the monitoring procedures based on the $\mathcal{W}_t$ and $\mathcal{M}_t$ sequences of statistics, (16) and (20), respectively, are also asymptotically valid under $H_{1,\eta_T,b(\cdot)}$ of (21).

Remark 10 The results in Lemma 4 reduce to those given previously for the limiting null distributions of the $\mathcal{W}_t$ and $\mathcal{M}_t$ statistics when $b(s)$ is identically zero for all $s \in [0, \omega]$. Under alternatives of the form considered in $H_{1,b^*(\cdot)}$ of (4), such that $b(s)$ is non-zero in at least one non-empty open interval in $(1, \omega]$, the results in Lemma 4 establish that both statistics display non-trivial offsets from their limiting null distributions and will therefore have power to detect local departures from the null hypothesis. Analytic expressions of the form given in Propositions 1 and 2 could be established based on the results in Lemma 4, but would not be especially informative. Consequently, we will use Monte Carlo simulation methods in section 5 to gauge the empirical rejection rates of the monitoring procedures based on $\mathcal{W}_t$ and $\mathcal{M}_t$. □

Remark 11 A further consequence of the results in Lemma 4 is that, in the weakly persistent case, the limiting distribution of the sequence of $\mathcal{M}_t$ statistics in cases where predictability is present in the training period (i.e. where $b(s)$ is non-zero over at least one non-empty interval in $[0, 1]$) is unchanged from what it is under the null hypothesis, H_0 of (3). In contrast, the limiting distributions of the $\mathcal{W}_t$ statistics show an offset from their null distribution when predictability is present in the training period, in both the weakly and strongly persistent cases. In the strongly persistent case, invariance to training period predictability no longer holds for the $\mathcal{M}_t$ statistics, because of the presence of the recursively demeaned component, $\overline{bJ}^{rec}(\tau)$. This effect can, however, be eliminated by re-initialising the recursive mean estimate of y_t used in computing the partial sums of $z_{t-1}(y_t - \bar{y}_t)$ at the start of the monitoring period. It is straightforward to show that the limiting null distribution of the resulting sequence of re-initialised $\mathcal{M}_t$ statistics is again given by the right member of (18), such that the result in Proposition 2 still holds. Moreover, the limiting distribution of these re-initialised $\mathcal{M}_t$ statistics will also be invariant to the presence of any predictability in the training period of the form of non-zero (open) intervals in $b(s)$ for $s \in [0, 1]$. We will explore these issues further in section 5. □

4.4 Alternative Procedures based on Smoothed Volatility Estimation

A key aspect of the monitoring procedures based on the $\mathcal{W}_t$ and $\mathcal{M}_t$ statistics is the estimate of the variance profile, $\hat{\pi}(t/T)$ in (15). Recall that this estimate is constructed as the ratio of unsmoothed estimates of the quadratic variation at time t and T, at each point in the monitoring period, $t = T + 1, ..., E$. An obvious alternative to this approach is to use an estimate formed as the ratio of estimates of the quadratic variation at time t and T, based on (the appropriate cumulation of) locally weighted variance estimates; that is,

$$\tilde{\pi}\left(\frac{t}{T}\right) := \frac{\sum_{j=1}^{t} \tilde{\sigma}_j^2}{\sum_{j=1}^{T} \tilde{\sigma}_j^2}, \quad t = T+1, T+2, \ldots, \tag{22}$$

where

$$\tilde{\sigma}_j^2 := \sum_{s=0}^{\min\{h,j\}} \psi_s \, z_{j-1-s}^2 (y_{j-s} - \bar{y}_{j-s})^2 \quad \text{with} \quad \psi_s := \frac{K(s/h)}{\sum_{i=0}^{\min\{h,j\}} K(i/h)}, \tag{23}$$

and where $K(\cdot)$ and h denoting the kernel function and bandwidth, respectively.[2]

Observe that $\tilde{\sigma}_j^2$ is (upon proper normalisation) a locally smoothed estimate of $\sigma_1^2(j/T)$ under strong persistence and of $\sigma_2^2(j/T)$ under weak persistence, at all points of continuity. Consequently, cumulating $\tilde{\sigma}_j^2$ provides an estimate of the antiderivative of $\sigma_1^2(\cdot)$ and $\sigma_2^2(\cdot)$ under strong and weak persistence, respectively; that is, of the quadratic variation processes of $M_1(\cdot)$ and $M_2(\cdot)$. The ratio of this cumulation formed at time t with that formed at time T therefore provides an estimate of the variance profile.

The estimate in (22) can be interpreted as a shrinkage estimator of the variance profile, designed to reduce variability relative to the estimate based on unsmoothed estimates, $\hat{\pi}\left(\frac{t}{T}\right)$. This variability may be large and can result in critical bounds that are too conservative; see, for example, the simulation results in section 5, where the monitoring procedure based on comparing $\mathcal{M}_t$ with a boundary function calculated using $\hat{\pi}(\cdot)$ tends to detect predictability later than its analogue using a boundary function calculated using $\tilde{\pi}(\cdot)$.

We make the following technical assumptions on the kernel function $K(\cdot)$ used in (22).

Assumption 4 *The kernel function $K(\cdot)$ is continuously differentiable over the interval $(0,1)$ and is such that $K(x) = 0$ for $x \leq 0$ and $x \geq 1$, and $\int_0^1 K(x)\mathrm{d}x = 1$.*

Remark 12 Assumption 4 imposes that $K(\cdot)$ is left-sided, such that the smoothed quadratic variation estimator computed at time $t > T$ does not include any future observations. This is, however, not restrictive in a real-time monitoring context where the practitioner does not have access to future data. Notice also that the conditions on the kernel function impose that $K(0) = 0$ and $K(1) = 0$, which implies that the current observation is left out when estimating the variance. Examples of kernels which satisfy Assumption 4 include the rectangular, Epanechnikov, Bartlett, and truncated Gaussian kernels (with their boundary values adjusted to 0 in each case). Further details relating to implementation, including bandwidth selection, are provided in section 5.1. □

In Lemma 5 we now establish that, under suitable conditions on the kernel function and bandwidth parameter, the smoothed variance profile estimator, $\tilde{\pi}(t/T)$ of (22), shares the same uniform consistency properties as the unsmoothed estimator, $\hat{\pi}(t/T)$ of (15).

Lemma 5 *Let Assumptions 1–4 hold. Then we have that under either the null hypothesis, H_0 of (3), or the local alternative, $H_{1,\eta_T,b(\cdot)}$ of (21), uniformly for $\tau \in [0,\omega]$:*

(i) Under Assumption 1.1 (strong persistence), $\tilde{\pi}(\lfloor \tau T \rfloor/T) \Rightarrow \pi_1(\tau)$, and

[2]Strictly speaking the weights ψ_s also depend on j, but, since this is only the case for $j < h$ at the beginning of the training period, we choose not to complicate notation and, hence, drop the additional index, j.

(ii) Under Assumption 1.2 (weak persistence), $\tilde{\pi}(\lfloor \tau T \rfloor / T) \Rightarrow \pi_2(\tau)$,

provided that $h, T \to \infty$ *such that* $h/\sqrt{T} \to 0$.

An immediate consequence of Lemma 5 is that the results given in Propositions 1 and 2 relating to the estimated boundary functions based on $\hat{\pi}(\cdot)$, also hold when the estimated boundary functions are based on $\tilde{\pi}(\cdot)$. Consequently, the monitoring procedures based on comparing $\mathcal{W}_t$ with the bounds $b^2_{t,\tilde{\pi}} = a^2_\alpha + \ln \tilde{\pi}(t/T)$ and $\mathcal{M}_t$ with the bounds $c^2_{t,\tilde{\pi}} = \tilde{\pi}(t/T)(\tilde{a}_\alpha + \ln \tilde{\pi}(t/T))$, in each case for $t = T+1, ..., E$, will both have FPRs controlled at level α in large samples, regardless of whether the predictor is weakly or strongly persistent.

The approach outlined above, based on constructing boundary functions using a kernel-based (smoothed) estimate of the variance path, also suggests a different approach based on standardised statistics. Here, in the spirit of Beare (2018), we would directly standardise the increments $z_{j-1}(y_j - \bar{y}_j)$ used in the construction of the $\mathcal{W}_t$ and $\mathcal{M}_t$ statistics by a consistent estimate of their variance at time j, prior to their cumulation in the numerators and denominators of the statistics. Monitoring procedures can then be formed by comparing the resulting standardised $\mathcal{W}_t$ and $\mathcal{M}_t$ statistics with the homoskedastic boundary functions in (10) and (19), respectively.

Taking the $\mathcal{M}_t$ statistic to illustrate, this approach is therefore based on monitoring using the standardised version of this statistic, given by

$$\widetilde{\mathcal{M}}_t := \frac{\left(\sum_{j=T+1}^{t} \frac{1}{\hat{\sigma}_j} z_{j-1}(y_j - \bar{y}_j)\right)^2}{\sum_{j=1}^{T} \frac{1}{\hat{\sigma}_j^2} z_{j-1}^2 (y_j - \bar{y}_j)^2}, \quad T < t \leq E, \tag{24}$$

where $\hat{\sigma}_j$ is any estimator which satisfies a suitable set of high-level uniform consistency assumptions. These conditions require careful formulation in the setup considered in this paper, given that the increments, $z_{j-1}(y_j - \bar{y}_j)$, have different properties under weak and strong persistence. In particular, under weak persistence the IVX instrument, z_{j-1} in (24), is bounded in probability, while under strong persistence it is of $O_p(T^{\eta/2})$. At the same time, the variance functions $\sigma_1^2(j/T)$ and $\sigma_2^2(j/T)$ are both bounded, regardless of the persistence of the predictor.

We can impose uniform consistency on $\hat{\sigma}_j$ in our setting by appropriately normalising z_j by its order of magnitude, resulting in the following requirements. Under strong persistence, we assume that the following condition holds

$$\sup_{T+1 \leq j \leq E} \left| \frac{\hat{\sigma}_j^2}{T^\eta} - \sigma_1^2(j/T) \right| \xrightarrow{p} 0, \tag{25}$$

while, under weak persistence, we assume that

$$\sup_{T+1 \leq j \leq E} \left| \hat{\sigma}_j^2 - \sigma_2^2(j/T) \right| \xrightarrow{p} 0. \tag{26}$$

We will also require the additional (standard) conditions that $\sigma_1^2(\cdot)$ and $\sigma_2^2(\cdot)$ are both

continuous functions which are bounded away from zero; recall that Assumption 3 only implies the weaker condition that they are differentiable almost everywhere.

In Lemma 6 we now state the limiting behaviour of the components which comprise the numerator and denominator of the standardised $\widetilde{\mathcal{M}}_t$ statistics in (24) when H_0 of (3) holds.

Lemma 6 *Let Assumptions 1–3 hold. Additionally assume that: (i) under strong persistence, $\sigma_1(\cdot)$ is continuous and bounded away from zero, and the uniform consistency condition in* (25) *holds; (ii) under weak persistence, $\sigma_2(\cdot)$ is continuous and bounded away from zero, and the uniform consistency results in* (26) *holds; and (iii) $\hat{\sigma}_j$ is adapted to the sequence of σ-algebras generated by $\{u_i, v_i\}_{i=-\infty}^{j-1}$. Then, as $T \to \infty$, uniformly for $\tau \in [1, \omega]$, it holds that:*

$$\frac{1}{\sqrt{T}} \sum_{j=T+1}^{\lfloor \tau T \rfloor} \frac{1}{\hat{\sigma}_j} z_{j-1} \left(u_j - \bar{u}_j\right) \quad \Rightarrow \quad W(\tau) - W(1) \tag{27}$$

$$\frac{1}{T} \sum_{j=1}^{T} \frac{1}{\hat{\sigma}_j^2} z_{j-1}^2 \left(u_j - \bar{u}_j\right)^2 \quad \xrightarrow{p} \quad 1. \tag{28}$$

Corollary 1 *An immediate consequence of the results in Lemma 6 is that, under H_0 of* (3)*, the sequence of standardised $\widetilde{\mathcal{M}}_t$ statistics in* (24) *satisfies*

$$\widetilde{\mathcal{M}}_{\lfloor \tau T \rfloor} \Rightarrow \left(W(\tau) - W(1)\right)^2, \qquad 1 \leq \tau \leq \omega, \tag{29}$$

regardless of whether Assumption 1.1 or Assumption 1.2 holds. Observe that the limiting distributions in (29) *are invariant to any heteroskedasticity present which satisfies Assumption 3.*

Remark 13 The result in (29) implies that a monitoring procedure based on comparing $\widetilde{\mathcal{M}}_t$, for $t > T$, with the (homoskedastic) upper bounds in (19) will have a FPR which is controlled at α in large samples, regardless of whether the predictor is weakly or strongly persistent. □

Remark 14 Notice that although the denominator of $\widetilde{\mathcal{M}}_t$ in (24) converges in probability to 1 under both weak and strong persistence, its presence in the statistic is nonetheless still vital in self-normalising the statistic, such that it has a well defined limiting distribution for both weakly and strongly persistent predictors. □

Remark 15 The additional conditions required to hold on $\sigma_1^2(\cdot)$ and $\sigma_2^2(\cdot)$ in Lemma (6) can be imposed by assuming the matrices $\mathbf{H}(\cdot)$ in Assumption 3 to be uniformly Lipschitz-continuous on their domain with eigenvalues that are bounded away from zero. □

To conclude this section, we briefly mention the issue of simultaneously monitoring with multiple potential predictors. Along the lines of the proof of Lemma 1, it is not difficult to establish weak convergence results for multiple partial sums of the form $\sum_{t=1}^{[rT]} \boldsymbol{z}_{t-1}(y_t - \bar{y}_t)$ under the null, where $\boldsymbol{z}_t$ is a k-dimensional vector each element of which is the IVX instrument associated with the corresponding element of the k-dimensional vector of (putative) predictors, $\boldsymbol{x}_t$. These limits will be multivariate Gaussian processes, possibly with non-linear variance profiles and time-varying correlations. Although it is straightforward in the homoskedastic case, as is established in Inoue and Rossi (2005), to extend the results of Chu *et al.* (1996) to the multivariate case, Lemmata 2 and 3 can only be extended to the multivariate case if each element of the multivariate Gaussian limiting processes has the same variance profile and the correlation between them is constant. This restriction would seem unreasonable in practice.

A simple, albeit conservative, solution is to monitor for predictability using each of the k predictors separately, using the methods for bivariate predictive regressions outlined in this section, and use a Bonferroni-based correction to asymptotically control the joint FPR across these k monitoring procedures. An alternative would be to explore the use a local standardisation approach based on a uniformly consistent estimate of the covariance matrix at time t which is then used to orthogonalise the elements $\boldsymbol{z}_{t-1}(y_t - \bar{y}_t)$, leading to asymptotics based on independent standard Brownian motion processes. In this case comparison can be made with the multivariate versions of the relevant (homoskedastic) boundary functions. However, kernel-based estimation of covariance matrices is a non-trivial task in its own right, particularly if the number of regressors is large. Moreover, the monitoring procedures we develop in this paper use the IVX methods developed in Kostakis *et al.* (2015) which are based on the assumption that all k predictors belong to the same persistence class (so that the predictors are either all strongly persistent or are all weakly persistent). Because of these complications we will not pursue multivariate monitoring procedures further in this paper but leave it as a suggestion for further research.

5 Finite Sample Simulations

We now investigate the finite sample behaviour of our proposed real-time monitoring procedures from section 4, comparing these with the procedures developed in Harvey *et al.* (2021). We will investigate both the empirical FPRs in cases where no predictability is present in the monitoring period and empirical TPRs (where TPR is the true probability rate – defined as the probability of at least one test in the monitoring sequence rejecting when the null hypothesis is false) for cases where a predictive window is present in the monitoring period. We will use the terms size and power interchangeably to refer to FPR and TPR, respectively. As shorthand notation, we will denote the FLUC-type monitoring procedure from section 4.1 as $\mathcal{W}^{\hat{\pi}}$, and the CUSUM-type monitoring procedure from section 4.2 as $\mathcal{M}^{\hat{\pi}}$. The modification of the latter from section 4.4 using a locally smoothed estimate of the variance profile will be denoted by $\mathcal{M}^{\tilde{\pi}}$. Finally, the procedure from section 4.4 based on the locally standardised statistic, $\widetilde{\mathcal{M}}_t$, based on the variance estimates defined in (23), will be denoted $\widetilde{\mathcal{M}}$.

5.1 Implementation Issues

Before outlining our simulation DGPs, we first discuss two implementation issues. The first involves the generation of critical values from the boundary functions used in implementing the monitoring procedures outlined in section 4. Second, in connection with the monitoring procedures outlined in section 4.4, we detail the choice of kernel function and bandwidth used.

Critical Values

In the context of their CUSUM-based real-time monitoring procedure for detecting asset price bubbles, Homm and Breitung (2012) find that basing their procedure on the asymptotic critical values, $|c_t|$, from (19), can be very conservative in practice, where T and E are both finite. We found the same to be true of our proposed CUSUM based procedures, $\mathcal{M}^{\hat{\pi}}$, $\mathcal{M}^{\tilde{\pi}}$ and $\widetilde{\mathcal{M}}$. This is perhaps not surprising given that the asymptotic boundary functions are designed such that the FPR of the procedure in large samples approaches α as ω tends to infinity. In the light of this, Homm and Breitung (2012) recommend running the procedure using finite sample critical values, generated by Monte Carlo simulation of a Gaussian random walk for the given value of the training sample size, T, for a monitoring period that terminates at time E. Unfortunately, we cannot generate critical values in the same way as Homm and Breitung (2012) because we have to allow for the fact that the predictor can be either weakly or strongly persistent under our null hypothesis. We therefore take a different approach, and use direct simulation of the limiting null distributions of the statistics, based on a discretisation of $T = 1000$, for a large number of possible values of $\omega := E/T$ and α, to obtain an estimated response surface which is a polynomial function of ω and the desired FPR, α. Finite-horizon critical values can then be obtained using the response surface for any desired ω and α. Full details on how these were computed together with the estimated response surface can be found in section S.2 of the supplementary appendix.

For the FLUC-based procedure, as we will see below, basing the procedure on the asymptotic critical values, $b^2_{t,\hat{\pi}}$, results in the empirical FPR of $\mathcal{W}^{\hat{\pi}}$ lying *above* the nominal level; that is, the procedure is liberal in practice. We found that using the critical value approach outlined above for the CUSUM procedures rendered the FLUC procedure even more liberal, and so for $\mathcal{W}^{\hat{\pi}}$ we report results based on asymptotic critical values. This behaviour is not entirely unexpected given that it is well documented that the t-ratio version of the conventional IVX statistic discussed in section 3 exhibits highly non-Gaussian behaviour, even for relatively large sample sizes, under strong persistence; see, *inter alia*, Demetrescu *et al.* (2023), Hosseinkouchack and Demetrescu (2021), and Kostakis *et al.* (2015). This phenomenon is partly due to the dependence between the numerator and denominator of the statistic which only vanishes slowly as the sample size diverges. By largely separating the data from the training and monitoring periods, the $\mathcal{M}_t$ statistic is far less impacted by this effect.

Kernel and Bandwidth Choices

In general, the choice of kernel function tends to be much less crucial for the finite sample

performance of kernel-based estimators than is the bandwidth, and we found this general rule to hold true for our procedures. We experimented with a number of possible choices of kernel in connection with the procedures outlined in section 4.4 but found little difference across these choices. The reported results are for the truncated Gaussian kernel.

In practice, it is the choice of bandwidth that is crucial to the performance of kernel-based volatility estimates. Other things being equal, adopting too large a bandwidth will result in over-smoothing which leads to increased bias in the volatility estimator, while using too small a bandwidth results in under-smoothing which leads to an increased variance in the resulting volatility estimator, both of which will have a detrimental impact on the empirical FPR and TPR of the resulting procedure. In order to automate the decision on how to trade off the bias and variance of the estimator we adopt the cross-validation approach employed by Astill *et al.* (2023) who suggest picking a locally (in terms of t) optimal bandwidth h_t^* as

$$h^* := \underset{h=h_{\min},\ldots,h_{\max}}{\arg\min} \; CV_t(h), \quad CV_t(h) := \frac{1}{h_{\max}} \sum_{j=t-h_{\max}+1}^{t} \left(\tilde{\sigma}_{j,h}^2 - z_{j-1}^2 \left(y_j - \bar{y}_j\right)^2\right)^2,$$

for each $t = 2, \ldots, \lfloor \omega T \rfloor$ in a rolling window-type manner.[3]

Despite our data-driven bandwidth choice, the set of bandwidths our cross-validation approach is allowed to choose from has to be specified in advance. Based on preliminary experiments, for a training period of 60 observations we found that setting $h_{\min} = 6, h_{\max} = 30$ provides the best trade-off between controlling FPR and maintaining good TPR for the procedure based on $\mathcal{M}_t^{\tilde{\pi}}$, while for the procedure based on $\widetilde{\mathcal{M}}_t$, this was achieved setting $h_{\min} = 15, h_{\max} = 45$. Our reported results are therefore based on the following bandwidth choices:

$$\begin{aligned} \mathcal{M}^{\tilde{\pi}} &: h_t^* = \arg\min_h CV_t(h),\ h \in [6, 30],\ t = 2, \ldots, \lfloor \omega T \rfloor, \\ \widetilde{\mathcal{M}} &: h_t^* = \arg\min_h CV_t(h),\ h \in [15, 45],\ t = 2, \ldots, \lfloor \omega T \rfloor. \end{aligned}$$

5.2 Simulation Design

In all of our experiments, we simulated data according to the DGP in (1) –(2b), setting $\mu_y = \mu_x = 0$ without loss of generality, for a range of values for the autoregressive parameter ρ, given by $\{1, 0.985, 0.965, 0.9\}$. Our reported simulations generate samples of size $E = 361$ with the first $T = 301$ observations forming the training period and the observations $t = 302, ..., 361$ forming the monitoring period. Our choices of T and E match those used in the simulation setup of Harvey *et al.* (2021) because we will compare the performance of our proposed monitoring procedures with those of the (two-tailed versions) of the *MAX* and *SEQ* monitoring procedures developed in Harvey *et al.* (2021). These choices of T and E result in $\omega \approx 1.2$. As discussed previously, these procedures

[3]We note that this criterion requires a slight adjustment when $j < h$. Because smoothing starts at $t = 2 < h_{\max}$, there are fewer than $h_{\max}$ observations available that $CV_t(h)$ can use to average across. Therefore, we adjust $CV_t(t)$ at the beginning of our training period so that it makes use of all available observations rather than the most recent $h_{\max}$ observations.

tend to be increasingly conservative the smaller is ω. Fortunately, preliminary simulations show that our methods can be well calibrated with relatively few training observations. This allows us to use only the last 60 observations ($t = 242, ..., 301$) from the training period, increasing the effective ω for our methods to 2, and we pick our simulated critical bounds accordingly, as outlined in section 5.1, to match the implied FPR of $\alpha = 0.2$ of the *MAX* and *SEQ* procedures from Harvey *et al.* (2021) discussed next.

The *MAX* and *SEQ* approaches from Harvey *et al.* (2021) achieve FPR control by subsampling t-statistics for the significance of the (putative) predictor from the training period data using a rolling-window approach. The resulting sequence of t-statistics is then used to obtain empirical critical values. Specifically, *MAX* indicates predictability whenever a t-statistic from the monitoring period exceeds the largest t-statistic obtained in the training period. Alternatively, *SEQ* utilizes rejection sequences to detect predictability. Beginning in the training period, the longest sequence of t-statistics exceeding the $1-\alpha_{SEQ}$ quantile of the training period t-statistics is determined. Then, *SEQ* signals predictability whenever one observes a sequence of null rejections in the monitoring period that is longer than the longest rejection sequence observed in the training period. The corresponding theoretical FPRs of both methods then deterministically depend on the chosen training and monitoring period lengths, and on the width of the rolling window.[4] For $T = 301$ and $E = 361$, choosing an internal rolling window width of 30 observations, $\alpha_{SEQ} = 0.05$, and ruling out any overlap in the data used to calculate the training period t-statistic and the data used to obtain the monitoring t-statistics, will result in theoretical FPRs of approximately $\alpha = 0.2$ for both *MAX* and *SEQ*. The removal of data overlap between training and monitoring t-statistics is intended to preserve power by creating two subsets of training data for *MAX* and *SEQ*. The first part ($t = 1, ..., 271$) is used to subsample t-statistics to construct the critical values, while the second part ($t = 272, ..., 301$) directly preceding the monitoring period forms part of the monitoring statistics.

We report results for the empirical FPR of the procedures under the null hypothesis, H_0, that $\beta_t = 0$, $t = 1, ..., 361$. We will also examine how quickly the procedures can detect predictability in the monitoring period. The first six simulation DGPs we consider are for cases where a single predictability episode is present starting at time $t = 302$ which persists until the end of the monitoring period. For these DGPs we therefore analyse the empirical TPRs of the procedures where $\beta_t = 0$, $t = 1, ..., 301$, and $\beta_t = \beta$, $t = 302, ..., 361$, reporting results for $\beta \in \{0.1, 0.2\}$. The final three simulation DGPs we consider are cases where predictability is either also present in the training period or occurs later within the monitoring period.

In particular, we will consider the following DGPs in our simulation study:

> **DGP1 (baseline model)**: As a baseline homoskedastic DGP we consider the case where the innovation vector $(u_t, v_t)'$ is generated as an i.i.d. sequence

[4] For more details, see Harvey *et al.* (2021, Eq. 9) and its modification where any overlap between the data used for training period t-statistics and the data used for the monitoring t-statistics is excluded (Harvey *et al.*, 2021, p. 55).

of bivariate normal distributions with mean vector zero and covariance matrix

$$\Sigma = \begin{pmatrix} 1 & \phi \\ \phi & 1 \end{pmatrix}$$

setting $\phi = -0.9$.[5]

DGP2 (conditional heteroskedasticity): This extends the baseline setup of DGP1 to allow u_t and v_t to follow independent GARCH(1, 1) processes. Keeping the unconditional correlation structure of DGP1, the innovations u_t and v_t are re-scaled using standard deviations obtained from the variance processes

$$\sigma^2_{u,t} = (1 - a_{u,1} - b_{u,1}) + a_{u,1}\varepsilon^2_{u,t-1} + b_1\sigma^2_{u,t-1}$$
$$\sigma^2_{v,t} = (1 - a_{v,1} - b_{v,1}) + a_{v,1}\varepsilon^2_{v,t-1} + b_1\sigma^2_{v,t-1}.$$

Choosing $a_{u,1} = a_{v,1} = 0.1, b_{u,1} = b_{v,1} = 0.85$ results in stationary GARCH(1,1) processes with $\mathrm{E}[u_t^2] = \mathrm{E}[v_t^2] = 1$ since $\mathrm{E}[u_t^2] = \mathrm{E}[v_t^2] = \frac{(1-a_1-b_1)}{1-a_1-b_1} = 1$ if $a_1, b_1 \geq 0$ and $a_1 + b_1 < 1$. Hence, the unconditional (co)variance structure remains unaffected by the rescaling.

DGP3 & DGP4 (smooth volatility break in monitoring period): To explore behaviour when a change in unconditional volatility occurs in the monitoring period, we consider two smooth transition volatility models. In the first, DGP3, σ_u^2 and σ_v^2 both triple over the monitoring period via the logistic transition function (using σ_t generically to denote both volatilities) $\sigma_t := 1 + \sigma_{\text{scale}}\left(1 + \exp\left[-\varphi\left(t - t_{\text{pivot}}\right)\right]\right)^{-1}$ where φ is the transition speed and t_{pivot} denotes the mid point of the transition. Picking $\sigma_{\text{scale}} = \sqrt{3} - 1$, $\varphi = 0.25$ and $t_{\text{pivot}} = 332$ results in a smooth upward transition centred around the midpoint of the monitoring period. In DGP4, the volatilities transition in the opposite direction, so that the variance of the innovations decreases from $\sigma_u^2 = \sigma_v^2 = 3$ to $\sigma_u^2 = \sigma_v^2 = 1$, via the transition function $\sigma_t := \sqrt{3} + \sigma_{\text{scale}}\left(1 + \exp\left[-\varphi\left(t - t_{\text{pivot}}\right)\right]\right)^{-1}$ with $\sigma_{\text{scale}} = 1 - \sqrt{3}$, $\varphi = 0.25$ and $t_{\text{pivot}} = 332$. The covariance matrix then becomes

$$\Sigma_t = \begin{pmatrix} \sigma^2_{u,t} & \phi\sigma_{u,t}\sigma_{v,t} \\ \phi\sigma_{u,t}\sigma_{v,t} & \sigma^2_{v,t} \end{pmatrix}$$

where $\phi = -0.9$. The other aspects of DGP1 are maintained in each case.[6]

DGP5 (smooth volatility break in the training period): As for DGP3 except that $t_{\text{pivot}} = 272$, such that the variances trebles centred around $t = 272 < T$.

DGP6 (smooth endogeneity break in training period): To explore the behaviour when an endogeneity change occurs in the training period, we consider a setup where the correlation parameter ϕ in Σ changes from weak ($\phi =$

[5]Valuation ratios like the dividend-price ratio or the earnings-price ratio exhibit strongly negatively correlated error terms when used as regressors in predictive regressions for the equity premium.

[6]In the case of DGP3, DGP4 and DGP5, we also considered the case where the variance breaks occurred instantaneously. In each case the results were qualitatively similar to those for the smooth transition case and, hence, are not reported.

-0.3) to strong ($\phi = -0.9$) negative correlation while the unconditional variances from DGP1 remain unchanged. Similar to DGP3, a logistic transition function controls the change so that $\phi_t := -0.3 + \phi_{\text{scale}} \left(1 + \exp\left[-\varphi \left(t - t_{\text{pivot}}\right)\right]\right)^{-1}$ where φ and t_{pivot} still denote the transition speed and the mid point of the transition, respectively. Picking $\phi_{\text{scale}} = -0.9 - (-0.3)$, $\varphi = 0.25$ and $t_{\text{pivot}} = 272$ results in a smooth upward transition centred around the point separating the data that is used to obtain the critical values from the data that is used for monitoring by *MAX* and *SEQ*. The other aspects of DGP1 are maintained.

DGP7 & DGP8 (predictability in the training period): DGP7 and DGP8 vary the baseline DGP1 to investigate the effect on the monitoring procedures when the maintained hypothesis of no predictability in the training period is violated. To that end, for DGP7 (DGP8.1) we additionally set $\beta_t = \beta^* = 0.1$ for $t = 272, ..., 286$ ($t = 152, ..., 166$). Subsequently, we explore the case where $\beta_t = \beta^* = 0.3$ for $t = 152, ..., 166$ in DGP8.2.

DGP9 (delayed predictability): DGP9 investigates the performance of the monitoring procedures when predictability begins after the start of the monitoring period. To that end, DGP9 varies from DGP1 in that we consider $\beta_t = 0$, $t = 1, ..., 331$ and $\beta_t = \beta$, $t = 332, ..., 361$, with $\beta \in \{0.1, 0.2\}$.

All simulations were performed on MATLAB 2024b using the Threefry 4x64 random number generator. In the context of the $\mathcal{W}^{\hat{\pi}}$, $\mathcal{M}^{\hat{\pi}}$, $\mathcal{M}^{\tilde{\pi}}$ and $\widetilde{\mathcal{M}}$ procedures, we follow the recommendation of Kostakis *et al.* (2015) and set $a = 1$ and $\eta = 0.95$ in (6) when generating IVX instruments.

5.3 Discussion of Results

Consider first Figures 1 and 2 which graph the empirical FPRs and TPRs of the monitoring procedures under the baseline DGP1. An obvious feature of these is that the $\mathcal{W}^{\hat{\pi}}$ (FLUC) procedure does not provide good FPR control under H_0 in cases where the (putative) predictor is (very) strongly persistent ($\rho = 1, \rho = 0.985$ and $\rho = 0.965$), though it is acceptable for less persistent predictors ($\rho = 0.9$). This observation is in line with our earlier discussion in section 5.1. As a result, we will not consider $\mathcal{W}^{\hat{\pi}}$ further in our reported results, other than for DGP7 where we show that its FPR is also very badly affected by the presence of predictability in the part of the training period used with these procedures (recall that this is the last 60 observations in the available training period). However, for the other procedures we propose in this paper, $\mathcal{M}^{\hat{\pi}}$, $\mathcal{M}^{\tilde{\pi}}$ and $\widetilde{\mathcal{M}}$, the results in Figures 1 and 2 show that they all provide good FPR control under H_0, regardless of the degree of predictor persistence. The minor differences that are observed between the empirical FPRs of these three procedures results from the fact that the finite-horizon critical bounds used for them are neither tailored to a specific method nor to this specific simulation DGP. The *MAX* and *SEQ* procedures of Harvey *et al.* (2021) are also seen to achieve good FPR control under DGP1 for all of the values of ρ considered.

Where predictability occurs in the baseline DGP1, $\beta = 0.1, 0.2$, Figures 1 and 2 highlight that significant differences can exist between the efficacies of those monitoring procedures

which display good FPR control, with these differences more marked the larger the value of ρ. The best performing of the $\mathcal{M}^{\hat{\pi}}$, $\mathcal{M}^{\tilde{\pi}}$ and $\widetilde{\mathcal{M}}$ procedures generally outperforms the best performing of the *MAX* and *SEQ* procedures, very significantly so in the case of $\rho = 1$. This is likely because of the backward looking rolling window structure of the latter which necessarily causes a delay in them detecting predictability. Among the $\mathcal{M}^{\hat{\pi}}$, $\mathcal{M}^{\tilde{\pi}}$ and $\widetilde{\mathcal{M}}$ procedures, $\mathcal{M}^{\tilde{\pi}}$ tends to detect predictability the earliest (on average) with $\mathcal{M}^{\hat{\pi}}$ marginally behind, while $\widetilde{\mathcal{M}}$ shows the slowest reaction among these three procedures to predictability. For all of the procedures, and as would be expected from the discussion in Remark 8, detection power is lower, other things equal, the smaller the value of ρ.

At this stage, we also note that the results under DGP9 (see Figures S.17 and S.18 in section S.3 of the supplementary appendix), where the predictive window does not start until $t = 332$, are qualitatively similar to those for DGP1 save that the empirical TPR curves of the procedures do not start to take off until about 30 observations later than they do for DGP1, as would be expected. In particular, our CUSUM-based statistics still in general outperform the *MAX* and *SEQ* procedures, again most notably so when $\rho = 1$.

For all of the monitoring procedures considered, the results for DGP2, where the innovations both follow GARCH$(1, 1)$ processes, are very similar throughout to the results for DGP1 in Figures 1 and 2. These results are reported in Figures S.1 and S.2 in section S.3 of the supplementary appendix.

Consider next the simulation DGPs which display unconditional heteroskedasticity. First, consider Figures 3 and 4 which relate to DGP3 where an upward shift in volatility occurs during the monitoring period. The empirical FPR curves of the $\mathcal{M}^{\tilde{\pi}}$ procedure displays a very marked upward kink at around the time the volatility change occurs. As a result, $\mathcal{M}^{\tilde{\pi}}$ has an empirical FPR well in excess of 0.2 by the end of the monitoring procedure, regardless of the value of ρ. The kink occurs around the time of the upward variance shift where more "large" increments will occur which will raise the values of the monitoring statistic before the critical value properly adjusts. Although $\mathcal{M}^{\hat{\pi}}$ and $\widetilde{\mathcal{M}}$ also display very minor kinks, they both display decent FPR control for all the values of ρ considered. This is also the case for the *MAX* and *SEQ* procedures, in spite of these not having theoretical validity in unconditionally heteroskedastic environments. Where predictability occurs under DGP3, we see a qualitatively similar power ordering among the FPR-controlled procedures - $\mathcal{M}^{\hat{\pi}}$, $\widetilde{\mathcal{M}}$, *MAX* and *SEQ* - as was observed under DGP1.

We turn next to Figures 5 and 6 which relate to DGP4, where a downward shift in volatility occurs in the monitoring period. Here we see that all of our proposed monitoring procedures, $\mathcal{M}^{\hat{\pi}}$, $\mathcal{M}^{\tilde{\pi}}$ and $\widetilde{\mathcal{M}}$ display empirical FPR curves that grow more slowly over the monitoring period than was seen for DGP1, with their FPRs ending up well below α at the end of the monitoring period, regardless of the value of ρ. The *MAX* and *SEQ* procedures are less conservative. As a result they see a boost in their power to detect a predictive episode, relative to $\mathcal{M}^{\hat{\pi}}$, $\mathcal{M}^{\tilde{\pi}}$ and most notably $\widetilde{\mathcal{M}}$, compared to the relative power of these procedures under DGP1. Notwithstanding this effect, however, $\mathcal{M}^{\hat{\pi}}$ and $\mathcal{M}^{\tilde{\pi}}$ are still in general noticeably more powerful than *MAX* and *SEQ* for the larger values of ρ considered under DGP4.

In DGP5 we investigate the impact of an upward volatility transition centred approximately halfway through our CUSUM-based procedures' effective training period at $t = 272$. These results are graphed in Figures S.3 and S.4 in section S.3 of the supplementary appendix. These results are qualitatively similar to those observed under DGP1, except for $\widetilde{\mathcal{M}}$ which becomes notably more conservative than under DGP1.

Next, in DGP6, we explore the case where the endogeneity correlation changes from being relatively weak in the training period to strongly negative in the monitoring period. These results are graphed in Figures S.5 and S.6 in section S.3 of the supplementary appendix. It is clear from these results that the behaviour of the $\mathcal{M}^{\hat{\pi}}$, $\mathcal{M}^{\tilde{\pi}}$ and $\widetilde{\mathcal{M}}$ procedures is essentially the same as was seen under DGP1; cf. Figures 1 and 2. In contrast, the FPRs of the *MAX* and *SEQ* procedures are both inflated relative to DGP1. In the case of *SEQ* this only results in very mild over-sizing, but for *MAX* the 20% nominal level is exceeded considerably before the end of the monitoring period, regardless of the value of ρ.

The final set of results we discuss relate to DGP7 and DGP8 where a predictive episode is present in the training period. Consider first the results for DGP7, graphed in Figures 7 and 8, where the predictive episode is located such that it is contained in the training period used for the $\mathcal{W}^{\hat{\pi}}$, $\mathcal{M}^{\hat{\pi}}$, $\mathcal{M}^{\tilde{\pi}}$ and $\widetilde{\mathcal{M}}$ procedures, and is contained with the second part of the training period used for the *MAX* and *SEQ* procedures. It is immediately seen from the top panels of the figures that the FLUC-type monitoring procedure, $\mathcal{W}^{\hat{\pi}}$, does not have a controlled FPR in this case. For example, for $\rho = 1$ the empirical FPR of $\mathcal{W}^{\hat{\pi}}$ is already at 40% at the start of the monitoring period, and reaches nearly 60% by the end of the monitoring period. This occurs because of a noncentrality (which becomes larger the closer ρ is to one) in the expanding window mean estimates, $\bar{y}_j$, used in the construction of the $\mathcal{W}_t$ statistic in (9), caused by the inclusion of predictive data from the training period.

The other three monitoring procedures we propose, $\mathcal{M}^{\hat{\pi}}$, $\mathcal{M}^{\tilde{\pi}}$ and $\widetilde{\mathcal{M}}$, all continue to show FPR control under DGP7, except when the predictor is a pure unit root series, $\rho = 1$. Here the empirical FPR exceeds the nominal 20% level by about half way through the monitoring procedure. Again, this is attributable to the bias in the mean estimates, $\bar{y}_j$, in the numerator of the $\mathcal{M}_t$ statistic in (17), caused by the inclusion of data from the predictive episode in the training sample. The *MAX* and *SEQ* procedures are also seen to be badly affected by the presence of a predictive episode in the training period, with for example the *MAX* procedure already displaying an empirical FPR of almost 30% at the start of the monitoring period. In the case of *MAX* and *SEQ* this effect occurs because the first few t-statistics calculated in the monitoring period use data that comes from the training period, including data where predictability is present, and, hence, they will incorrectly reject the null hypothesis of no predictability in the monitoring period too often.

Results for DGP8, where a predictive episode is present in the training period used for *MAX* and *SEQ*, but not in the training period used for $\mathcal{M}^{\hat{\pi}}$ (as the latter are employed with a training period of 60 observations, see section 5.2), $\mathcal{M}^{\tilde{\pi}}$ and $\widetilde{\mathcal{M}}$, are depicted in Figures S.13 and S.14 for DGP8.1 where $\beta^* = 0.1$ (a relatively weak predictability signal), and in Figures S.15 and S.16 for DGP8.2 where $\beta^* = 0.3$ (a relatively strong predictability signal) in the supplementary appendix. These results show that, as might be expected,

the $\mathcal{M}^{\hat{\pi}}$, $\mathcal{M}^{\tilde{\pi}}$ and $\widetilde{\mathcal{M}}$ procedures are essentially unaffected, relative to DGP1, by this predictive episode. In contrast, *MAX* and *SEQ* are more conservative than under DGP1. This effect is seen to be more pronounced the stronger is the persistence of the predictor and the larger the value of β^*, other things equal. This occurs because the location of this predictive episode inflates many of the t-statistics calculated in the training period for *MAX* and *SEQ*, the more so the larger the value of β^*, thereby making it less likely that these procedures will reject in the monitoring period, other things equal.

To conclude this section it is worth briefly considering how one might mitigate against the impact of scenarios where the training period used contains observations with predictive content. First, and most obviously, in any empirical application of these methods, we would strongly advise running a standard IVX predictability test, as described in section 3, on the whole of the training period data before commencing with a monitoring exercise. If such a test failed, then we would advise against using either the FLUC-type monitoring procedure, $\mathcal{W}^{\hat{\pi}}$, proposed in this paper, or the *MAX* and *SEQ* procedures of Harvey *et al.* (2021) (recall that the training period used for the procedures proposed in this paper will not, in general, coincide with that used in connection with *MAX* and *SEQ*).

In the case of the $\mathcal{M}^{\hat{\pi}}$, $\mathcal{M}^{\tilde{\pi}}$ and $\widetilde{\mathcal{M}}$ procedures, a practical solution, following the discussion in Remark 11, is to re-initialise the $\bar{y}_j$ estimate used in the numerator of $\mathcal{M}_t$ in (17) using an expanding window which starts at $t = T + 1 - k$, using $k = 0$, rather than at $t = 1$. By doing so, the training period predictability signal will no longer affect the mean estimates used in the numerator of (17) and the tests would be expected to display good FPR control again. Figures S.7 and S.8 demonstrate this for the case of DGP7. Although using $k = 0$ is a straightforward way to remove the effect of training period predictability, this approach is, however, seen to negatively affect the TPR of our approaches, particularly for very strongly persistent predictors; cf. Figure S.7. Because predictability in DGP7 begins at time $t = T + 1$, the mean estimates in the numerator for the choice of $k = 0$ are based only on observations that include a predictability signal. Consequently, these estimates absorb some of the predictability signal causing a power loss. If instead one chooses $k > 0$, then so long as there is no predictability in the last k observations in the training period, this should have largely the same effect on the FPR as choosing $k = 0$. Figures S.9-S.12 summarise the results for $k = 6$ and $k = 12$, *i.e.* re-initialising $\bar{y}_j$ in the monitoring period using the most final six or twelve observations from the training period. As can be seen, the resulting empirical FPRs are very similar to those obtained for $k = 0$ in both cases, while the negative effects of $k = 0$ on the TPRs is mitigated, such that we almost regain the earlier power levels seen for DGP1, especially when working with $k = 12$; indeed, for weakly persistent predictors the TPRs are even seen to be improved, relative to those obtained under DGP1.

On the basis of the simulation results presented in this section, $\mathcal{M}^{\hat{\pi}}$ and $\widetilde{\mathcal{M}}$, implemented with the finite-horizon critical bounds developed in section S.2 of the supplementary appendix, appear to achieve a very useful practical trade-off between avoiding spurious rejections under the null yet maintaining good power to detect predictive episodes both in homoskedastic environments and when faced with changes in the unconditional variances, or the endogeneity correlation, occurring in either the monitoring or the training period. $\mathcal{M}^{\tilde{\pi}}$ is slightly more powerful than $\mathcal{M}^{\hat{\pi}}$ but does not control the FPR as well as $\mathcal{M}^{\hat{\pi}}$ in heteroskedastic environments, while $\widetilde{\mathcal{M}}$ is the most conservative of the proposed meth-

ods. Our procedures are generally more powerful than the *MAX* and *SEQ* procedures of Harvey *et al.* (2021), with this power advantage most obviously seen for predictors with an autoregressive root close to one. Where predictive episodes are present in the training period, our simulations show that the procedures based on the $\mathcal{M}_t$ statistic in (17) can still be reliably used, provided one appropriately re-initialises the mean estimates used in the numerator of the statistic.

6 Empirical Application

6.1 Dataset and Preliminary Data Analysis

In this section we apply the procedures developed in this paper to conduct a pseudo real-time monitoring exercise for the predictability of the equity premium. The data we use to do this stems from the datasets first used in Welch and Goyal (2008) and Goyal *et al.* (2024) in their 2024 vintages, available from Amit Goyal's homepage, https://sites.google.com/view/agoyal145. The data comprises monthly S&P 500 observations along with various candidate predictors. The sample data cover the period January 1965 to December 2024, which amounts to 720 sample observations in total (some of the candidate predictors have later start dates).

The equity premium is obtained from this dataset according to its standard definition in the literature using the month-end S&P 500 observations and the treasury bill rate; that is,

$$eq\_p_t = \ln(1 + r_{m,t}) - \ln(1 + r_{f,t}),$$

where $r_{m,t}$ denotes the return of the CRSP index and $r_{f,t}$ denotes the rate of the risk-free treasury bill. The eq_p_t series is graphed in Figure S.19 in section S.4 of the supplementary appendix.

The dataset of candidate predictors of eq_p_t we consider consists of a variety of variables including, *inter alia*, dividend- and earnings-related variables, market-related indicators such as volatility measures and technical indicators, bond rates and bond yields, and inflation data; see Table S.1 for a complete list of the candidate predictors used. The one period lag of each of these variables is employed in a bivariate predictive regression for eq_p_t, of the form given in (1). Some of the candidate predictors are entered in logs, and we refer the reader to the definitions in Welch and Goyal (2008) and Goyal *et al.* (2024) for detailed descriptions of the data. Time-series plots of these variables are provided in Figure S.20 of the supplementary appendix.

Before conducting our pseudo real-time monitoring exercise, we first perform some preliminary statistical analysis of the full sample of equity premium and the range of candidate predictors we consider. Starting with the equity premium, we note that the average monthly equity premium is 0.0046, or approximately 0.46 %, with a standard deviation of 0.044. The smallest equity premium is observed in October 1987, around the time

of Black Monday, at -0.2482, *i.e.* a negative excess return of approximately -22%. The largest equity premium is reached at 0.1492, an approximate excess return of 16.1%, in October 1974, towards the end of the 1973-1974 stock market crash.

Turning to the range of possible predictors considered in our application, we investigate their degrees of persistence and endogeneity. To get an idea of the former, AR(1) models were fitted to each individual variable using the full sample of available data. The fitted values of the AR(1) parameter from these regressions are reported in Table S.2 in the supplementary appendix. Although an AR(1) model will not necessarily accurately capture the true underlying DGP, it should suffice to get an indication of how strongly persistent these variables are. The results suggest that most of the candidate predictors can be considered strongly persistent with many, including the dividend-price ratio *dp*, the dividend yield *dy*, and the book-to-market ratio *bm*, returning an AR(1) coefficient very close to unity. Moreover, for many of the candidate predictors, strong negative endogeneity correlations are also estimated. These are reported in Table S.2 and are estimated as the sample correlation between the residuals from the fitted bivariate predictive regression model for the given predictor and the residuals from the AR(1) model fitted to that predictor. For example, for *dp* the estimated endogeneity correlation is -0.99, while for *bm* it is -0.75. These results accord with the earlier findings of, *inter alia*, Welch and Goyal (2008) and Goyal *et al.* (2024) that strong regressor persistence coupled with high endogeneity is regularly encountered when running predictive regression methods for returns, so that a method robust to these data features (such as the monitoring procedures proposed in this paper) is required for valid inference.

Another important feature of the monitoring procedures we propose is their robustness to unconditional heteroskedasticity in the data. In connection with this, Table S.3 in the supplementary appendix reports the results of applying the four stationary volatility tests ($\mathcal{KS}$, $\mathcal{R}$, $\mathcal{CVM}$ and $\mathcal{AD}$, using a Bartlett long run variance estimator with lag truncation parameter 4) proposed by Cavaliere and Taylor (2008, pp. 311–312) to our data. These tests are valid regardless of whether the variable is weakly or strongly persistent, and test the null hypothesis of unconditional homoksedasticity against the alternative of unconditional heteroskedasticity. In line with earlier findings in the literature, the tests reject the null hypothesis of stationary volatility for almost half of our candidate predictors, highlighting the importance of using monitoring procedures, such as those proposed in this paper, which are robust to the possibility of non-stationary volatility in the data.

Table S.2 also reports the outcomes of conventional IVX-based tests for predictability applied to the full available samples of data.[7] These are seen to provide no statistically significant evidence of predictability at the 5% level for any of the predictors, with the exception of the cross-sectional premium *csp*, the tail risk *tail*, the long-term corporate bond returns *corpr*, and the output gap *ogap*. Taken together with the weak explanatory power of the predictors, indicated by the low R^2 values, and in multiple cases negative adjusted R^2 values, these results suggest that if predictability is present for these candidate predictors, then intermittent spells of predictability would seem more likely than

[7]We do not report any in-sample OLS-based tests as these are likely to be unreliable in the light of the aforementioned evidence of strong persistence and endogeneity present in many of the candidate predictors.

predictability signals that persist throughout the whole sample.

Finally, following our recommendation in section 5.3, before conducting monitoring we first pre-test each predictor for predictive content in just the training period data. The IVX-based predictability tests indicate training period predictability at the 5% level only for two of the volatility measures, *impvar* and *rsvix*. Recall that neither of these variables is identified as a predictor based on the complete data sample. Moreover, the data for both of these series start in 1996, so that the initial five years of data include Russia's 1998 default, during which both series show a notable local peak. For both these series the estimated AR(1) coefficient is also under 0.80; as we saw in section 5.3, predictability in the training sample with ρ values below unity does not appear to significantly affect the empirical FPR of the monitoring procedures we will use in section 6.2. Based on these findings, we will treat the training data as having no predictive episode present for all of the candidate predictors considered.

6.2 Pseudo Real-Time Monitoring Exercise

In our monitoring exercise, we will mimic a practitioner applying our $\mathcal{M}_t^{\hat{\pi}}$, $\mathcal{M}_t^{\tilde{\pi}}$, and $\widetilde{\mathcal{M}}_t$ monitoring procedures, as if in real-time. We use the first 60 available observations of each time series as training data. For most of our predictor series, this then implies that observations from January 1965 to December 1969 serve as the training period, with monitoring beginning in January 1970, resulting in $\omega \leq 12$ for all of the candidate predictors. Using the value of ω relevant to each predictor, we parameterise the critical bound using the simulated $\tilde{a}_\alpha$ values, obtained as detailed in section S.2 of the supplementary appendix, to achieve nominal FPRs of approximately 10% at the end of our available pseudo-out-of-sample period. As in our simulation study, in connection with the $\mathcal{M}_t^{\tilde{\pi}}$ monitoring statistic we set $h_{\min} = 6, h_{\max} = 30$, while for the $\widetilde{\mathcal{M}}_t$ statistic we set $h_{\min} = 15, h_{\max} = 45$. Moreover, we continue to use $a = 1$ and $\eta = 0.95$ in (6) when generating the IVX instruments.

The procedures indicate predictability if and when a monitoring statistic is encountered which exceeds the critical value bound. In such cases we treat this as the start date of a predictive episode and also provide an informal end date of the predictive episode. For the latter we use the first time after detection that the outcome of the monitoring statistic falls back below the critical value bound. To stabilise this date in situations where the monitoring statistic only marginally exceeds the critical value bound, we skip at least six months before assigning an end date to the predictive episode. Consequently, all of the predictability windows we report are at least six months long. To resume monitoring after a predictive episode has been deemed to have finished, the critical bound is adjusted to the remaining number of available pseudo-out-of-sample observations; that is, $\tilde{a}_\alpha$ is recomputed to achieve an expected FPR of 10% at the end of the revised monitoring period.[8]

[8]Following the discussion in section 5.3 regarding predictive episodes in the training period, as a robustness check we also re-ran the monitoring procedures skipping another year of observations following the end of any detected predictability window and re-initialising $\bar{y}_j$ using only the most recent $k = 12$ observations. This is done to avoid any potential contamination of the test statistic resulting from the

Our monitoring procedures are conducted using two-tailed critical bounds with the results summarised in Table 1. Consider first the results for the $\mathcal{M}^{\hat{\pi}}$ procedure. This detects four windows of predictability around the stock market boom and bust in the early 1970s with long-term focused sovereign or corporate rates of return *ltr*, the default yield spread *dfy*, *corpr*, and *ogap* all identified as predictors. The end-point detection approach suggests that *ltr* predicts the equity premium from September 1974 until April 1975, *dfy* predicts it from September 1974 to August 1975, while *corpr* predicts it for even longer, namely from November 1973 until June 1975. For *ogap* the minimum predictive window length of six months is reported. The $\mathcal{M}^{\hat{\pi}}$ monitoring procedure additionally suggests that the term spread *tms* is a predictor of the equity premium from December 1985 to September 1987, the end coinciding with Black Monday in October 1987. A predictive window in the late 1970s associated with new orders and shipments of durable goods *ndrbl* was also detected. Two windows of predictability were also found for illiquidity measures *lrzt*: from January 2018 to September 2018, and in the second half of 2022.

Consider next the results relating to the $\mathcal{M}^{\bar{\pi}}$ monitoring procedure in Table 1. Perhaps not surprisingly in the light of the simulation results in section 5, which showed this procedure to be less conservative than $\mathcal{M}^{\hat{\pi}}$, this procedure indicates the presence of more predictability windows than $\mathcal{M}^{\hat{\pi}}$. Additional predictability windows are detected, among others, for the earnings-price ratio *ep*, the earnings-average price ratio *ep10*, and the long-term yield *lty*. With the exception of the predictability windows seen for the variance risk premium *vrp* from November 1983 to September 1984, *lzrt* from November 2008 to April 2009, and for *svar* and *tms* which display predictive content for the equity premium around the time of the stock market crash in 1987, all of these windows date to the 1974/75 period for which a number of detections were also seen with the $\mathcal{M}^{\hat{\pi}}$ monitoring procedure.

We next turn to the results for the $\widetilde{\mathcal{M}}$ procedure in in Table 1. Interestingly, this is the only one among the three procedures that detects a predictive signal from dividend-earnings, *de*, this being dated to end at the time of Black Monday. It also detects predictability from *AAA* and *BAA* in the 1970s neither of which were found by the other two procedures. The results for $\widetilde{\mathcal{M}}$ also find a longer window of predictability for *corpr* than the other two procedures, extending to December 2004, and also find a second window of predictability from October 2008 to April 2010 for this predictor. The results for $\widetilde{\mathcal{M}}$ also support the earlier detections obtained with the other two procedures for *ltr*, *tms*, *corpr*, *dfy*, *ogap*, and *ndrbl*. Moreover, $\widetilde{\mathcal{M}}$ finds further windows of predictability for several of the predictors; for example, for the Treasury bill rate *tbl*, it detects a second window of predictability starting in October 1987.

7 Conclusions

In this paper we have developed a number of novel econometric procedures used for real-time monitoring for the emergence of a predict episode in returns. Our preferred

previous predictability window featuring in the training period after restarting the procedure. These results (available upon request) are very similar to those reported.

approach is based on sequences of CUSUM-type statistics constructed from the relevant IVX instrument moment condition. The false probability rate is controlled by comparing the sequence of monitoring statistics with critical value bounds constructed using a real-time estimate of the variance profile, or by standardising the monitoring statistics using kernel-based variance estimates. The former approach is novel and has the potential to be applied beyond the current setting to other scenarios where exceedance probabilities of sequences of monitoring statistics need to be controlled for possible nonstationary volatility in the data in real-time. The methods we develop are robust to whether the predictors are strongly or weakly persistent, and to the possibility of time variation in the unconditional second moments (unconditional heteroskedasticity and/or time-varying endogeneity correlations) of the shocks in either the training or monitoring periods. In an empirical application, our proposed monitoring procedures uncover evidence supportive of the existence of periods where a measure of the U.S. equity premium is predictable for a number of (lagged) financial and macroeconomic variables.

References

Astill, S., Harvey, D. I., Leybourne, S. J., Sollis, R., & Taylor, A. M. R. (2018). Real-time monitoring for explosive financial bubbles. *Journal of Time Series Analysis*, *39*(6), 863–891.

Astill, S., Harvey, D. I., Leybourne, S. J., Taylor, A. M. R., & Zu, Y. (2023). CUSUM-based monitoring for explosive episodes in financial data in the presence of time-varying volatility. *Journal of Financial Econometrics*, *21*(1), 187–227.

Beare, B. K. (2018). Unit root testing with unstable volatility. *Journal of Time Series Analysis*, *39*(6), 816–835.

Breitung, J., & Demetrescu, M. (2015). Instrumental variable and variable addition based inference in predictive regressions. *Journal of Econometrics*, *187*(1), 358–375.

Calvo-Gonzalez, O., Shankar, R., & Trezzi, R. (2010). *Are commodity prices more volatile now? A long-run perspective* (Policy Research Working Paper Series No. 5460). The World Bank.

Campbell, J. Y., & Yogo, M. (2006). Efficient tests of stock return predictability. *Journal of Financial Economics*, *81*(1), 27–60.

Campbell, J. Y. (2008). Viewpoint: Estimating the equity premium. *Canadian Journal of Economics/Revue canadienne d'économique*, *41*(1), 1–21.

Campbell, J., & Thompson, S. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies*, *21*, 1509–1531.

Cavaliere, G. (2004). Unit root tests under time-varying variances. *Econometric Reviews*, *23*(3), 259–292.

Cavaliere, G., & Taylor, A. M. R. (2008). Time-transformed unit root tests for models with non-stationary volatility. *Journal of Time Series Analysis*, *29*(2), 300–330.

Cavanagh, C. L., Elliott, G., & Stock, J. H. (1995). Inference in models with nearly integrated regressors. *Econometric Theory*, *11*(5), 1131–1147.

Chu, C.-S. J., Stinchcombe, M., & White, H. (1996). Monitoring structural change. *Econometrica*, *64*(5), 1045–1065.

Demetrescu, M., Georgiev, I., Rodrigues, P., & Taylor, A. M. R. (2023). Extensions to IVX methods of inference for return predictability. *Journal of Econometrics*, *237*(2C), 105271.

Demetrescu, M., Georgiev, I., Rodrigues, P. M. M., & Taylor, A. M. R. (2022). Testing for episodic predictability in stock returns. *Journal of Econometrics*, *227*(1), 85–113.

Demetrescu, M., & Hillmann, B. (2022). Nonlinear predictability of stock returns? Parametric vs. nonparametric inference in predictive regressions. *Journal of Business & Economic Statistics*, *40*(1), 382–397.

Elliott, G., & Stock, J. H. (1994). Inference in time series regression when the order of integration of a regressor is unknown. *Econometric Theory*, *10*, 672–700.

Elliott, G., Müller, U. K., & Watson, M. W. (2015). Nearly optimal tests when a nuisance parameter is present under the null hypothesis. *Econometrica*, *83*(2), 771–811.

Giraitis, L., Koul, H. L., & Surgailis, D. (2012). *Large Sample Inference for Long Memory Processes*. Imperial College Press, London.

Goyal, A., Welch, I., & Zafirov, A. (2024). A comprehensive 2022 look at the empirical performance of equity premium prediction. *The Review of Financial Studies*, *37*(11), 3490–3557.

Harvey, D. I., Leybourne, S. J., Sollis, R., & Taylor, A. M. R. (2021). Real-time detection of regimes of predictability in the US equity premium. *Journal of Applied Econometrics*, *36*(1), 45–70.

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

Henkel, S. J., Martin, J. S., & Nardari, F. (2011). Time-varying short-horizon predictability. *Journal of Financial Economics*, *99*(3), 560–580.

Homm, U., & Breitung, J. (2012). Testing for speculative bubbles in stock markets: A comparison of alternative methods. *Journal of Financial Econometrics*, *10*(1), 198–231.

Hosseinkouchack, M., & Demetrescu, M. (2021). Finite-sample size control of IVX-based tests in predictive regressions. *Econometric Theory*, *37*(4), 769–793.

Inoue, A., & Rossi, B. (2005). Recursive predictability tests for real-time data. *Journal of Business & Economic Statistics*, *23*(3), 336–345.

Jansson, M., & Moreira, M. J. (2006). Optimal inference in regression models with nearly integrated regressors. *Econometrica*, *74*(3), 681–714.

Johannes, M., Korteweg, A., & Polson, N. (2014). Sequential learning, predictability, and optimal portfolio returns. *Journal of Finance*, *69*, 611–644.

Kostakis, A., Magdalinos, T., & Stamatogiannis, M. P. (2015). Robust econometric inference for stock return predictability. *Review of Financial Studies*, *28*(5), 1506–1553.

McMillan, D., & Wohar, M. (2011). Structural breaks in volatility: The case of UK sector returns. *Applied Financial Economics*, *21*(15), 1079–1093.

Paye, B. S., & Timmermann, A. (2006). Instability of return prediction models. *Journal of Empirical Finance*, *13*(3), 274–315.

Phillips, P. C. B., & Lee, J. H. (2013). Predictive regression under various degrees of persistence and robust long-horizon regression. *Journal of Econometrics*, *177*(2), 250–264.

Phillips, P. C. B. (2015). Pitfalls and possibilities in predictive regression. *Journal of Financial Econometrics*, *13*(3), 521–555.

Phillips, P. C. B., & Magdalinos, T. (2009). *Econometric inference in the vicinity of unity* (CoFie Working Paper No. 7). Singapore Management University.

Ploberger, W., & Krämer, W. (1992). The CUSUM test with OLS residuals. *Econometrica*, *60*(2), 271–285.

Rapach, D. E., Strauss, J. K., & Wohar, M. E. (2008). Forecasting stock return volatility in the presence of structural breaks. In D. E. Rapach & M. E. Wohar (Eds.), *Forecasting in the presence of structural breaks and model uncertainty*. Emerald Group Publishing Limited.

Timmermann, A. (2008). Elusive return predictability. *International Journal of Forecasting*, *24*(1), 1–18.

Vivian, A., & Wohar, M. (2012). Commodity volatility breaks. *Journal of International Financial Markets, Institutions and Money*, *22*(2), 395–422.

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

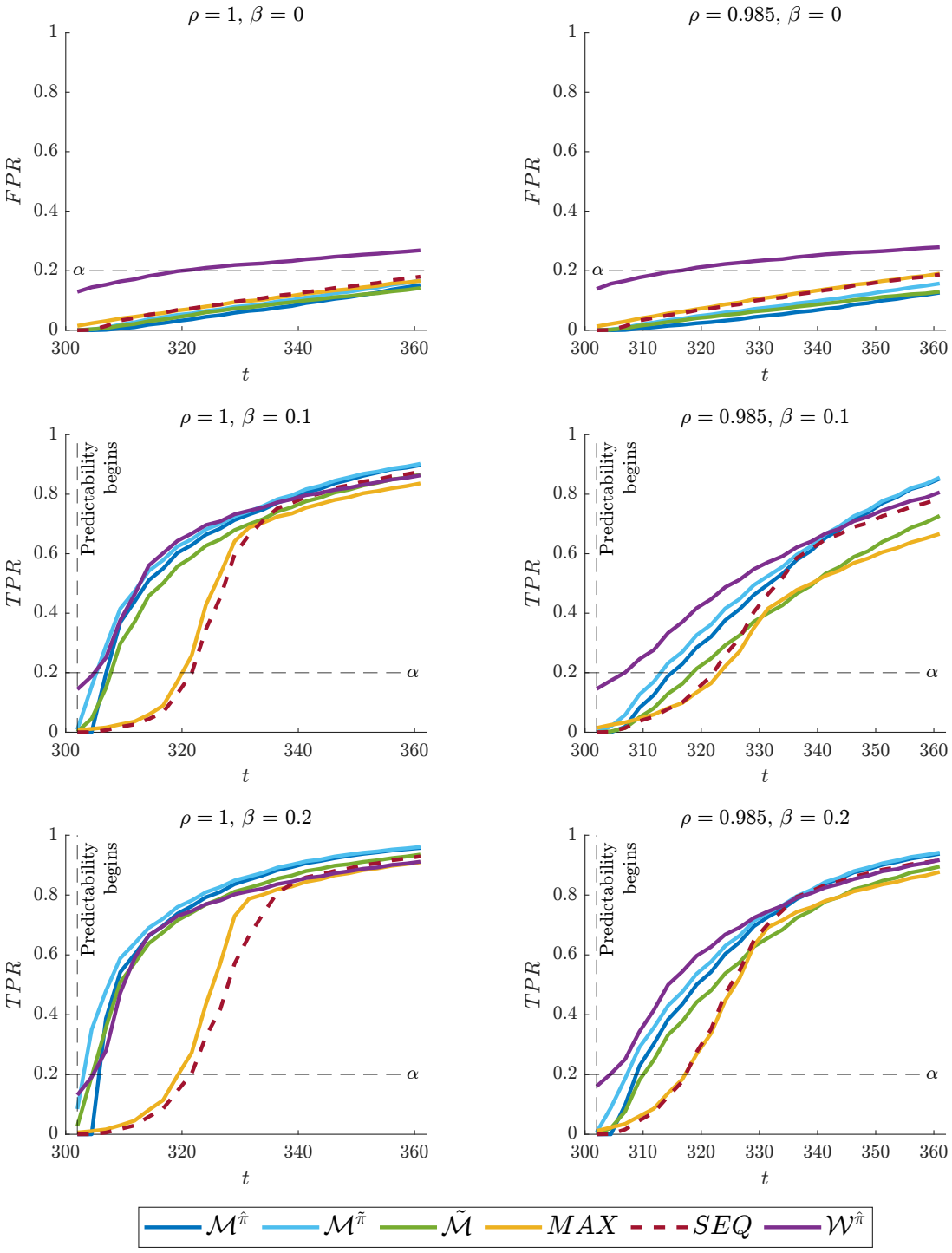

Figure 1: FPR/TPR curves of $\mathcal{M}^{\hat{\pi}}, \mathcal{M}^{\tilde{\pi}}, \widetilde{\mathcal{M}}, MAX, SEQ$ and $\mathcal{W}^{\hat{\pi}}$ for $\rho \in \{1, 0.985\}$ and $\beta \in \{0, 0.1, 0.2\}$ and data generated from DGP1 (baseline).

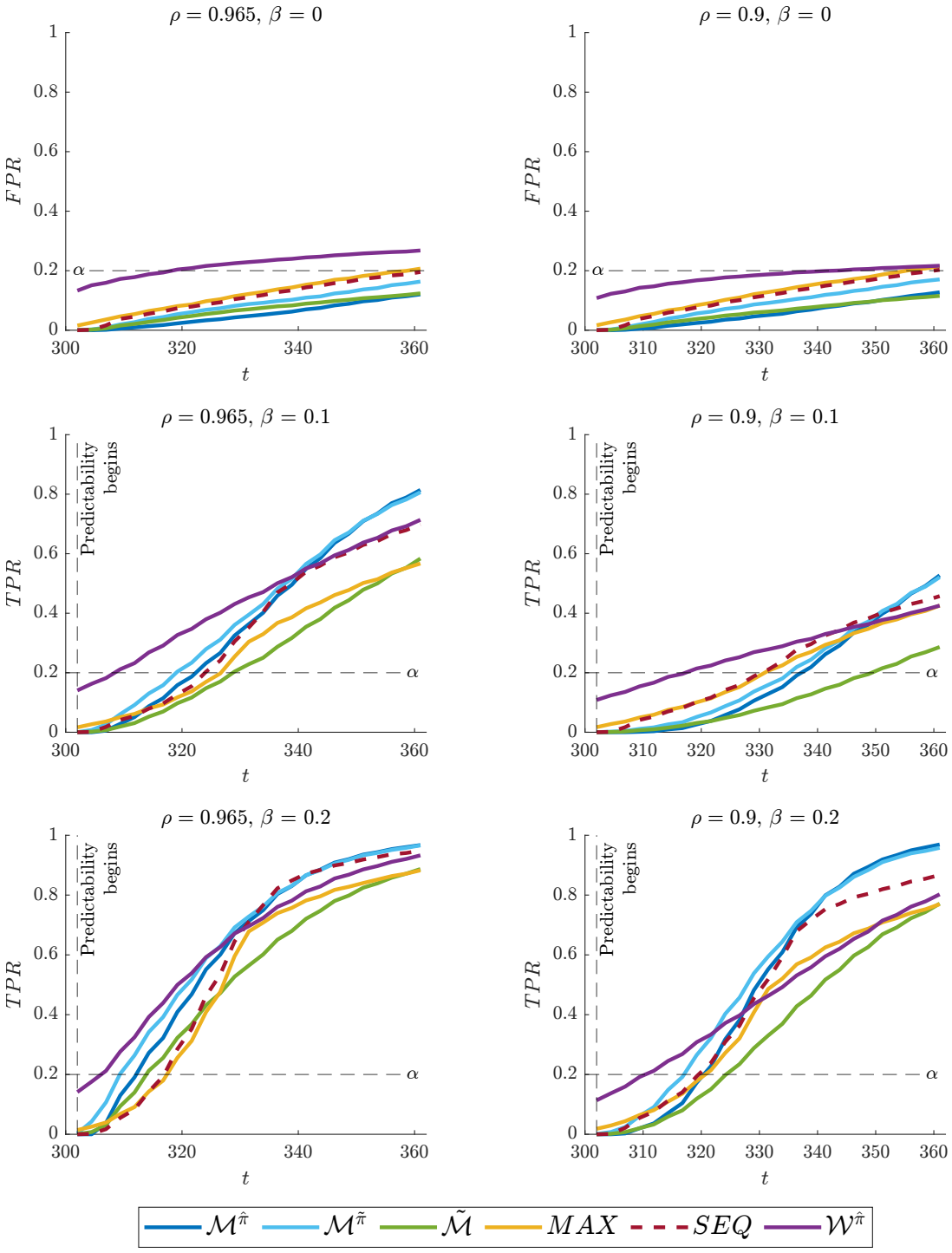

Figure 2: FPR/TPR curves of $\mathcal{M}^{\hat{\pi}}, \mathcal{M}^{\tilde{\pi}}, \widetilde{\mathcal{M}}, MAX, SEQ$ and $\mathcal{W}^{\hat{\pi}}$ for $\rho \in \{0.965, 0.9\}$ and $\beta \in \{0, 0.1, 0.2\}$ and data generated from DGP1 (baseline).

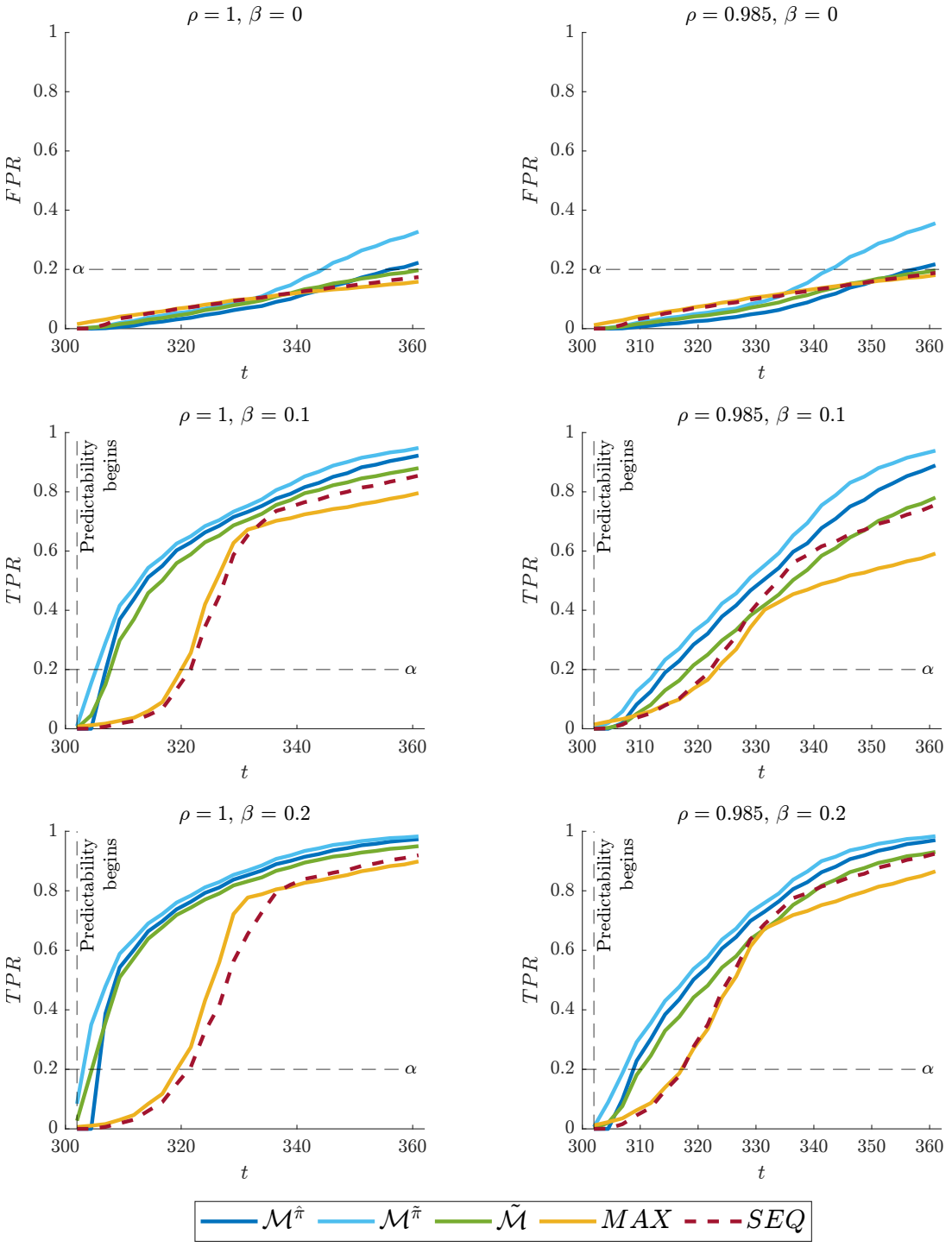

Figure 3: FPR/TPR curves of $\mathcal{M}^{\hat{\pi}}, \mathcal{M}^{\tilde{\pi}}, \widetilde{\mathcal{M}}, MAX$ and SEQ for $\rho \in \{1, 0.985\}$ and $\beta \in \{0, 0.1, 0.2\}$ and data generated from DGP3 (upward variance transition).

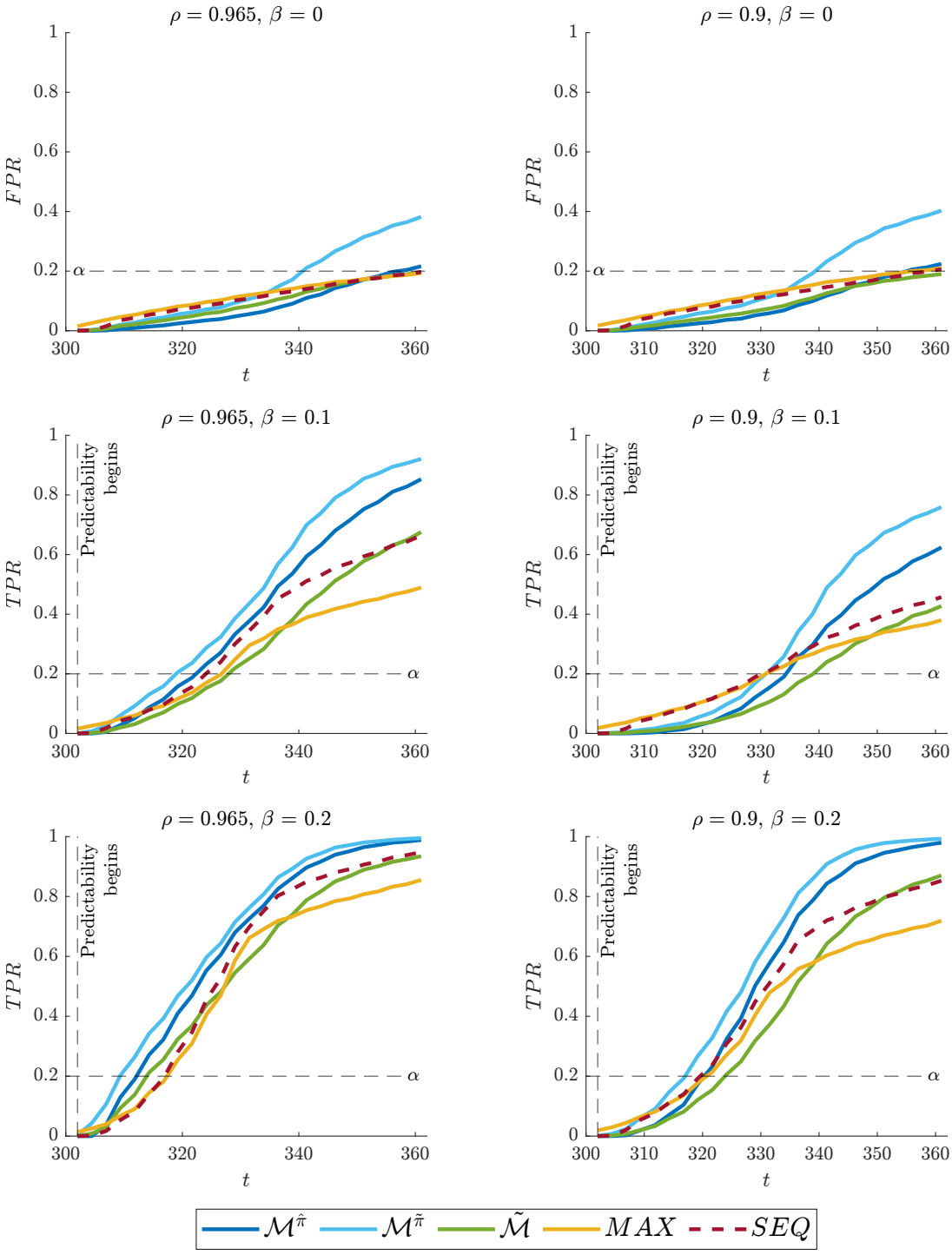

Figure 4: FPR/TPR curves of $\mathcal{M}^{\hat{\pi}}, \mathcal{M}^{\tilde{\pi}}, \widetilde{\mathcal{M}}, MAX$ and SEQ for $\rho \in \{0.965, 0.9\}$ and $\beta \in \{0, 0.1, 0.2\}$ and data generated from DGP3 (upward variance transition).

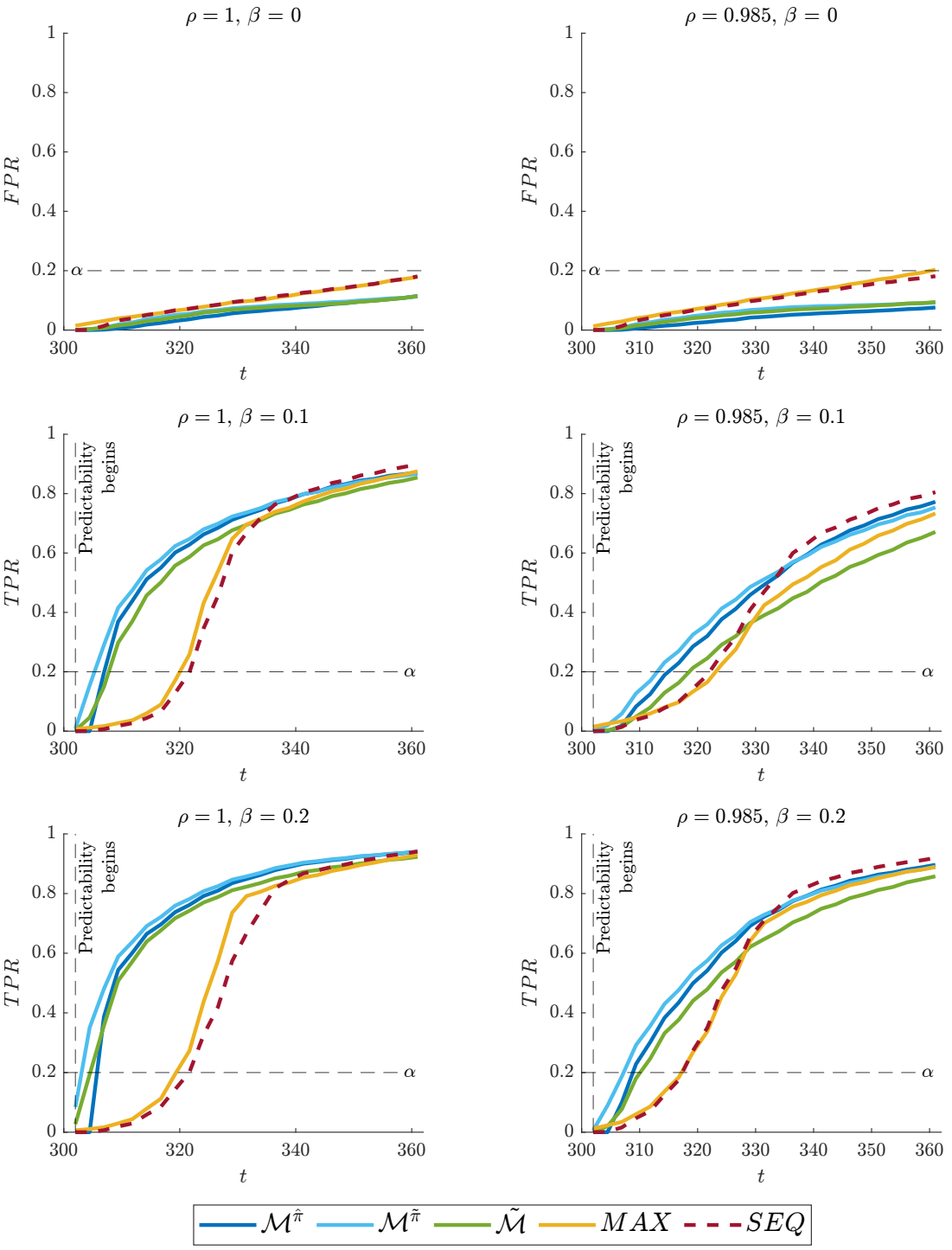

Figure 5: FPR/TPR curves of $\mathcal{M}^{\hat{\pi}}, \mathcal{M}^{\tilde{\pi}}, \widetilde{\mathcal{M}}, MAX$ and SEQ for $\rho \in \{1, 0.985\}$ and $\beta \in \{0, 0.1, 0.2\}$ and data generated from DGP4 (downward variance transition).

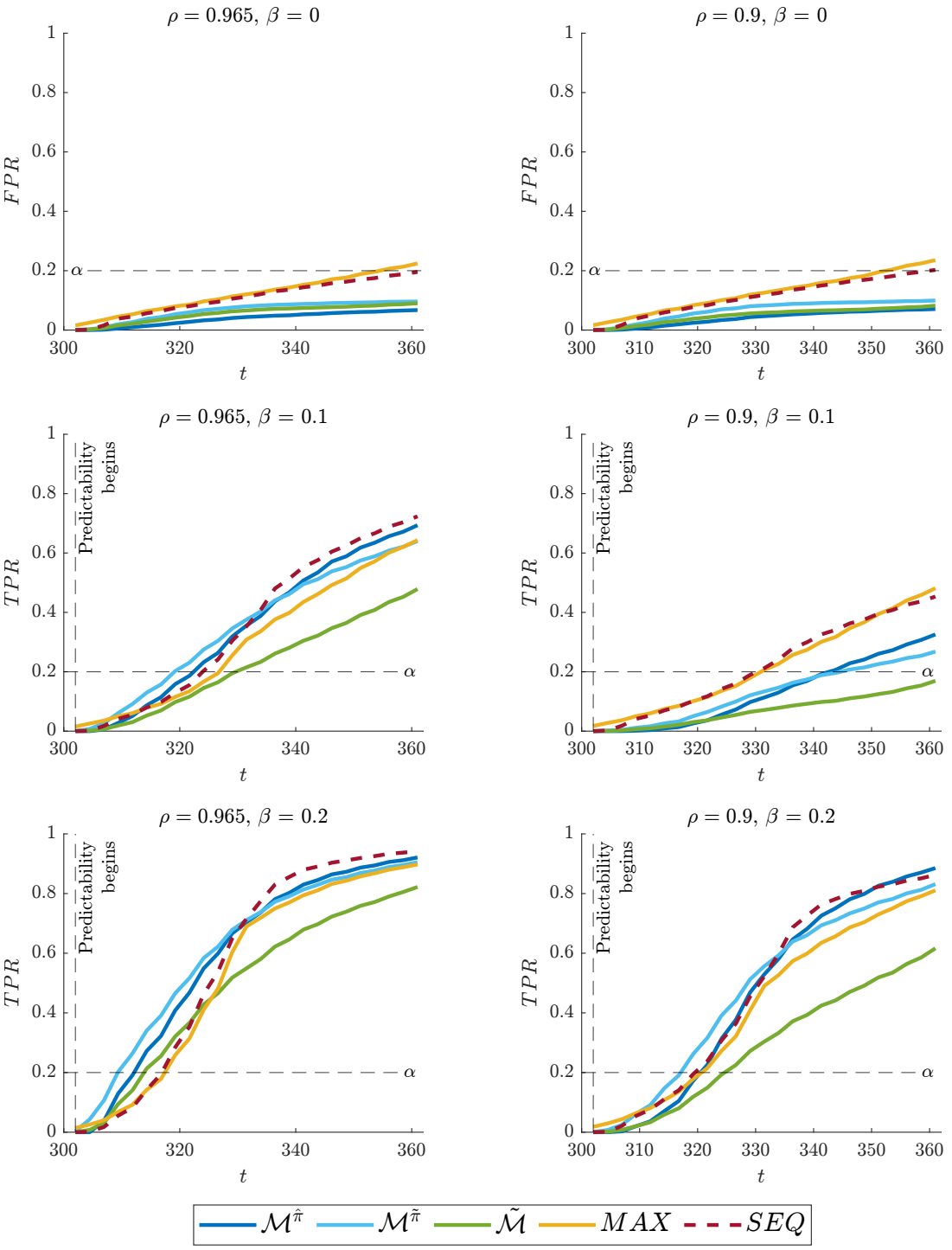

Figure 6: FPR/TPR curves of $\mathcal{M}^{\hat{\pi}}, \mathcal{M}^{\tilde{\pi}}, \widetilde{\mathcal{M}}, MAX$ and SEQ for $\rho \in \{0.965, 0.9\}$ and $\beta \in \{0, 0.1, 0.2\}$ and data generated from DGP4 (downward variance transition).

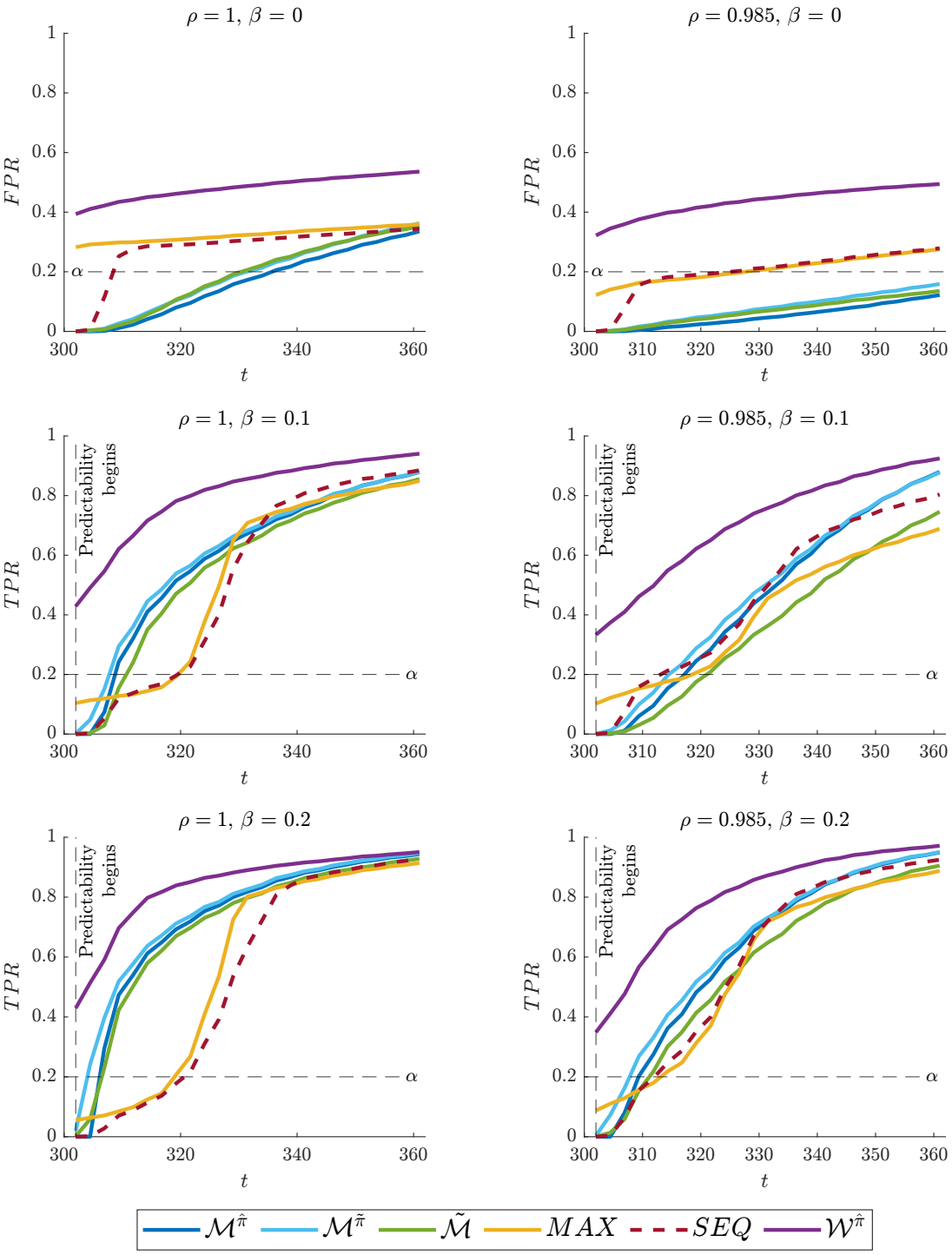

Figure 7: FPR/TPR curves of $\mathcal{M}^{\hat{\pi}}, \mathcal{M}^{\tilde{\pi}}, \widetilde{\mathcal{M}}, MAX, SEQ$ and $\mathcal{W}^{\hat{\pi}}$ for $\rho \in \{1, 0.985\}$ and $\beta \in \{0, 0.1, 0.2\}$ and data generated from DGP7 (predictability in the training period).

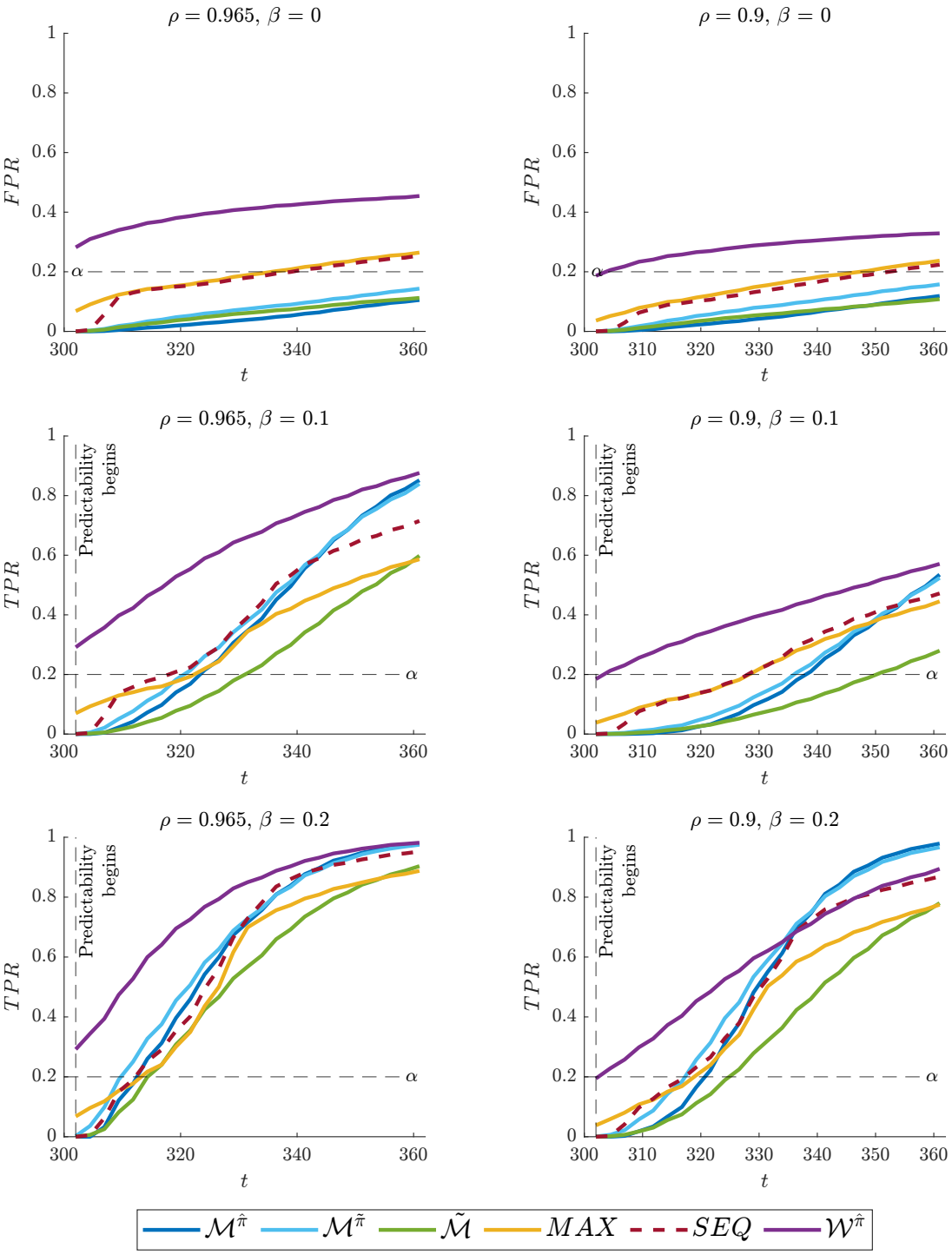

Figure 8: FPR/TPR curves of $\mathcal{M}^{\hat{\pi}}, \mathcal{M}^{\tilde{\pi}}, \widetilde{\mathcal{M}}, MAX, SEQ$ and $\mathcal{W}^{\hat{\pi}}$ for $\rho \in \{0.965, 0.9\}$ and $\beta \in \{0, 0.1, 0.2\}$ and data generated from DGP7 (predictability in the training period).

-	$\mathcal{M}_t^{\hat{\pi}}$		$\mathcal{M}_t^{\tilde{\pi}}$	$\widetilde{\mathcal{M}}_t$	
dp	-	-	-	-	-
dy	-	-	-	-	-
ep	-	-	08/74-01/75	-	-
e10p	-	-	-	-	-
ep10	-	-	08/74-01/75	-	-
de	-	-	-	03/87-09/87	-
svar	-	-	11/87-04/88	11/87-12/24	-
impvar	-	-	-	-	-
rsvix	-	-	-	-	-
vp	-	-	-	-	-
vrp	-	-	11/83-09/84	02/91-07/91	03/95-08/98
csp	-	-	-	01/78-06/78	04/96-12/02
skvw	-	-	-	-	-
tail	-	-	-	-	-
fbmlr	-	-	-	-	-
dtoy	-	-	09/74-02/75	-	-
dtoat	-	-	-	-	-
ygap	-	-	08/74-01/75	-	-
rdsp	-	-	-	-	-
tchi	-	-	-	-	-
avgcor	-	-	-	11/82-04/85	-
shtint	-	-	-	-	-
disag	-	-	-	-	-
lzrt	01/18-09/18	06/22-11/22	11/08-04/09	-	-
bm	-	-	-	-	-
ntis	-	-	-	-	-
tbl	-	-	09/74-02/75	08/74-03/75	10/87-12/88
lty	-	-	08/74-01/75	07/74-10/75	-
ltr	09/74-04/75	-	10/74-03/75	08/73-01/74	-
tms	12/85-09/87	-	06/87-11/87	10/82-04/84	-
AAA	-	-	-	08/74-01/75	-
BAA	-	-	-	08/74-01/75	-
corpr	11/73-06/75	-	04/74-12/74	08/73-12/04	10/08-04/10
dfy	09/74-08/75	-	08/74-08/81	08/74-08/96	-
dfr	-	-	-	-	-
infl	-	-	09/74-02/75	12/87-09/88	-
ogap	05/75-10/75	-	02/75-07/75	06/75-11/75	-
wtexas	-	-	11/73-04/74	11/73-09/75	-
sntmlr	-	-	-	02/74-01/80	-
ndrbl	01/78-06/78	-	08/74-07/75	07/97-12/97	08/14-06/16

Table 1: Detected windows of equity premium predictability. Two-tailed tests conducted at nominal 10% FPR using training data from January 1965 to December 1969 and a pseudo-out-of-sample window from January 1970 to December 2024.

Supplementary Appendix to "Real-Time Monitoring for Stock Return Predictability in Nonstationary Volatility Environments"

Matei Demetrescu[a], Fabian Schmidt[a] and A.M. Robert Taylor[b]

[a] Department of Statistics, TU Dortmund University

[b] Essex Business School, University of Essex

November 2025

This supplementary appendix contains three sections. Section S.1 provides proofs of the technical results stated in the main paper. Section S.2 details how the finite-horizon critical values were generated for our proposed monitoring procedures. Section S.3, provides the additional Monte Carlo results referred to in section 5 of the main paper. Finally, section S.4 provides additional material relating to the empirical application reported in section 6 of the main paper.

S.1 Mathematical Proofs

Throughout the proofs we will use C to denote a generic positive constant. We will also use the standard notation $\|A\|_p := \sqrt[p]{\mathrm{E}\,|A|^p}$ to denote the L_p-norm of a random variable.

S.1.1 Proof of Lemma 1

Definitions of the limiting processes $M_v(\cdot)$ and $M_u(\cdot)$ together with their corresponding quadratic variations, $[M_u]$ and $[M_v]$ respectively, which appear in the statement of Lemma 1 can be found in Lemma 4 and Lemma 3 of the supplementary material of Demetrescu *et al.* (2023). For completeness, we also provide these definitions in section S.1.9 below, along with those for $M_1(\cdot)$ and $M_2(\cdot)$ and their respective quadratic variations.

Notice, from the definition of $(u_t, v_t)' := \mathbf{H}(t/T)\mathbf{a}_t$ given in Remark 4, that $w_t = B(L)v_t$, where v_t is a uniformly L_4-bounded MD array under Assumption 3. Furthermore, since u_t shares these properties, it holds that $\mathrm{Var}(\bar{u}_t) \leq C/t$ for all t.

For later reference, we note that, for any $t > 0$ and positive integer k,

$$\sum_{j=0}^{t} \varrho^{kj} = \frac{1-\varrho^{k(t+1)}}{1-\varrho^{k}} \leq CT^{\eta}$$

while

$$\sum_{j=1}^{t} \frac{1}{j} \leq C \ln t$$

and, for any $s > 0$,

$$\sum_{j=t}^{t+s} \frac{1}{j^2} \leq \sum_{j=t}^{\infty} \frac{1}{j(j-1)} \leq \frac{1}{t-1}.$$

Consider first the proof of the first result given in part (i) of Lemma 1 relating to the strongly persistent case. Writing

$$\frac{1}{T^{1/2+\eta/2}} \sum_{j=1}^{\lfloor \tau T \rfloor} z_{j-1}\left(u_j - \bar{u}_j\right) = \frac{1}{T^{1/2+\eta/2}} \sum_{j=1}^{\lfloor \tau T \rfloor} z_{j-1}u_j - \frac{1}{T^{1/2+\eta/2}} \sum_{j=1}^{\lfloor \tau T \rfloor} z_{j-1}\bar{u}_j, \tag{S.1}$$

we will show that the second summand on the right hand side of (S.1) is asymptotically negligible, uniformly in τ. To do so, we note that

$$z_t = B(1)\sum_{j=0}^{t-1} \varrho^j v_{t-j} + r_t$$

where

$$\sup_{1 \leq t \leq \lfloor \omega T \rfloor - 1} \|r_t\|_4 = O(T^{\eta - 1/2}).$$

This result obtains by checking the steps of the proof of Lemma 3.1 of Demetrescu and Hillmann (2022) pertaining to the moment behaviour of r_t under Assumption 3.[1] Given that v_t is a uniformly L_4-bounded MD array, we may use Lemma 2.5.2 of Giraitis *et al.* (2012) to obtain that

$$\mathrm{E}\left(\left|\sum_{j=0}^{t-1}\varrho^j v_{t-j}\right|^4\right) \leq C\left(\sum_{j=0}^{t-1}\varrho^{2j}\left\|v_s\right\|_4^2\right)^2 \leq CT^{2\eta}$$

which implies that $T^{-\eta/2}\sum_{j=0}^{t-1}\varrho^j v_{t-j}$ is uniformly L_4-bounded and consequently so is $T^{-\eta/2}z_t$. It is then a straightforward exercise to demonstrate that $\max_{1\leq t\leq\lfloor\omega T\rfloor}\left|T^{-\eta/2}z_{t-1}\right| = O_p(T^{1/2})$.

Recalling that $z_0 = 0$, we can write

$$\sum_{j=1}^{\lfloor\tau T\rfloor} z_{j-1}\bar{u}_j = B(1)\sum_{t=2}^{\lfloor\tau T\rfloor}\left(\sum_{j=0}^{t-2}\varrho^j v_{t-1-j}\right)\bar{u}_t + \sum_{j=2}^{\lfloor\tau T\rfloor} r_{j-1}\bar{u}_j. \tag{S.2}$$

The triangle and Cauchy-Schwarz inequalities imply that the second term on the right hand side of (S.2) is such that

$$\sup_{0\leq\tau\leq\omega}\left|\sum_{j=2}^{\lfloor\tau T\rfloor} r_{j-1}\bar{u}_j\right| \leq \sup_{0\leq\tau\leq\omega}\sum_{j=2}^{\lfloor\tau T\rfloor}\left|r_{j-1}\bar{u}_j\right| \leq \sum_{j=2}^{\lfloor\omega T\rfloor}\left|r_{j-1}\bar{u}_j\right| \leq \sqrt{\sum_{j=2}^{\lfloor\omega T\rfloor} r_{j-1}^2\sum_{j=2}^{\lfloor\omega T\rfloor}\bar{u}_j^2}.$$

Markov's and Jensen's inequalities together with the moment behaviour of r_t and $\bar{u}_t$ then further imply that

$$\sum_{j=2}^{\lfloor\omega T\rfloor} r_{j-1}^2 = O_p\left(\sum_{j=2}^{\lfloor\omega T\rfloor}\mathrm{E}\left(r_{j-1}^2\right)\right) = O_p\left(\sum_{j=2}^{\lfloor\omega T\rfloor}\left\|r_{j-1}\right\|_4^2\right) = O_p\left(T^{2\eta}\right)$$

and

$$\sum_{j=2}^{\lfloor\omega T\rfloor}\bar{u}_j^2 = O_p\left(\sum_{j=2}^{\lfloor\omega T\rfloor}\mathrm{E}\left(\bar{u}_j^2\right)\right) = O_p\left(\ln T\right)$$

such that

$$\sup_{0\leq\tau\leq\omega}\left|\sum_{j=2}^{\lfloor\tau T\rfloor} r_{t-1}\bar{u}_j\right| = O_p\left(T^{\eta}\sqrt{\ln T}\right) = o_p\left(T^{1/2+\eta/2}\right)$$

as $\eta < 1$. Examining the first term on the right hand side of (S.2), we obtain upon

[1] For the benefit of the reader, we note that Lemma 3.1 of Demetrescu and Hillmann (2022) only states the result that $\|r_t\|_4 = o(T^{1/2+\eta/2})$. However, the proof they provide derives the stricter $T^{\eta-1/2}$ rate given above.

re-arranging the summation terms that

$$\begin{aligned}\sum_{t=2}^{\lfloor \tau T\rfloor}\left(\sum_{j=0}^{t-2}\varrho^{j}v_{t-j-1}\right)\left(\frac{1}{t}\sum_{k=1}^{t}u_k\right) &= \sum_{t=1}^{\lfloor \tau T\rfloor-1}v_tu_t\left(\sum_{j=t+1}^{\lfloor \tau T\rfloor}\frac{1}{j}\varrho^{j-t-1}\right)+\sum_{t=2}^{\lfloor \tau T\rfloor}u_t\sum_{s=1}^{t-1}v_s\left(\sum_{j=t}^{\lfloor \tau T\rfloor}\frac{1}{j}\varrho^{j-s-1}\right)\\ &\quad+\sum_{t=2}^{\lfloor \tau T\rfloor-1}v_t\sum_{s=1}^{t-1}u_s\left(\sum_{j=t+1}^{\lfloor \tau T\rfloor}\frac{1}{j}\varrho^{j-t-1}\right)\\ &=: A_{\tau,T}+B_{\tau,T}+C_{\tau,T}.\end{aligned}$$

Let us consider the three terms, $A_{\tau,T}$, $B_{\tau,T}$ and $C_{\tau,T}$ in turn. For $A_{\tau,T}$, because $\mathrm{E}\left(|v_tu_t|\right)\leq\sqrt{\mathrm{E}\left(e_t^2\right)\mathrm{E}\left(u_t^2\right)}$ is uniformly bounded, Markov's inequality implies that

$$\begin{aligned}\sup_{0\leq\tau\leq\omega}|A_{\tau,T}| &\leq \sup_{0\leq\tau\leq\omega}\sum_{t=1}^{\lfloor \tau T\rfloor-1}|v_tu_t|\left(\sum_{j=t+1}^{\lfloor \tau T\rfloor}\frac{1}{j}\varrho^{j-t-1}\right)\leq\sum_{t=1}^{\lfloor \omega T\rfloor-1}|v_tu_t|\frac{1}{t}\left(\sum_{j=t+1}^{\lfloor \omega T\rfloor}\varrho^{j-t-1}\right)\\ &= O_p\left(T^{\eta}\ln T\right)=o_p\left(T^{1/2+\eta/2}\right).\end{aligned}$$

Turning to $B_{\tau,T}$, write

$$B_{\tau,T}=\sum_{t=2}^{\lfloor \tau T\rfloor}u_t\sum_{s=1}^{t-1}v_s\left(\sum_{j=t}^{\lfloor \tau T\rfloor}\frac{1}{j}\varrho^{j-s-1}\right)=\sum_{t=2}^{\lfloor \tau T\rfloor}b_{t,T}$$

where $\{b_{t,T}\}$ is seen to possess the MD property. Consequently,

$$\mathrm{Var}\left(B_{\tau,T}\right)=\sum_{t=2}^{\lfloor \tau T\rfloor}\mathrm{Var}\left(b_{t,T}\right)$$

where, using the Cauchy-Schwarz inequality,

$$\mathrm{Var}\left(b_{t,T}\right)=\mathrm{E}\left(u_t^2\left(\sum_{s=1}^{t-1}v_s\left(\sum_{j=t}^{\lfloor \tau T\rfloor}\frac{1}{j}\varrho^{j-s-1}\right)\right)^2\right)\leq\sqrt{\mathrm{E}\left(u_t^4\right)\mathrm{E}\left(\left|\sum_{s=1}^{t-1}v_s\left(\sum_{j=t}^{\lfloor \tau T\rfloor}\frac{1}{j}\varrho^{j-s-1}\right)\right|^4\right)}.$$

Recalling that v_t is a uniformly L_4-bounded MD array, Lemma 2.5.2 of Giraitis *et al.* (2012) implies that

$$\begin{aligned}\mathrm{E}\left(\left|\sum_{s=1}^{t-1}v_s\left(\sum_{j=t}^{\lfloor \tau T\rfloor}\frac{1}{j}\varrho^{j-s-1}\right)\right|^4\right) &\leq C\left(\sum_{s=1}^{t-1}\frac{1}{t^2}\left(\sum_{j=t}^{\lfloor \tau T\rfloor}\varrho^{j-s-1}\right)^2\|v_s\|_4^2\right)^2\leq C\left(T^{2\eta}\sum_{s=1}^{t-1}\frac{1}{t^2}\right)^2\\ &\leq CT^{4\eta}\frac{(t-1)^2}{t^4}\end{aligned}$$

for all τ and t. Consequently,

$$\mathrm{Var}\left(b_{t,T}\right)\leq CT^{2\eta}\frac{1}{t}\qquad\text{and thus}\qquad\mathrm{Var}\left(B_{\tau,T}\right)\leq CT^{2\eta}\ln T$$

uniformly in τ. Doob's inequality then implies that

$$\sup_{0\leq\tau\leq\omega}|B_{\tau,T}| = O_p\left(T^{\eta}\sqrt{\ln T}\right) = o_p\left(T^{1/2+\eta/2}\right)$$

as $\eta < 1$. Using the same arguments, it is then not difficult to show that $C_{\tau,T}$ is also asymptotically negligible, uniformly in τ.

Summing up, $\frac{1}{T^{1/2+\eta/2}}\sum_{j=2}^{\lfloor\tau T\rfloor} z_{j-1}\bar{u}_j$ is asymptotically negligible, uniformly in τ, and so the large sample behaviour of $\frac{1}{T^{1/2+\eta/2}}\sum_{j=1}^{\lfloor\tau T\rfloor} z_{j-1}\left(u_j-\bar{u}_j\right)$ in (S.1) is determined by $\frac{1}{T^{1/2+\eta/2}}\sum_{j=1}^{\lfloor\tau T\rfloor} z_{j-1}u_j$. Using Lemma 5(d) of Demetrescu *et al.* (2023) (which trivially extends to the interval $[0,\omega]$), we therefore have that

$$\frac{1}{T^{1/2+\eta/2}}\sum_{j=2}^{\lfloor\tau T\rfloor} z_{j-1}\left(u_j-\bar{u}_j\right) \Rightarrow M_1(\tau) := \frac{B(1)}{\sqrt{2a}}\int_0^{\tau}\sigma_u(s)\sigma_v(s)\mathrm{d}W(s)$$

where $W(\cdot)$ is a standard Brownian motion process which is independent of $M_v(\cdot)$ and, consequently, also of $J_{c,\mathrm{H}}(\cdot)$ defined in (11). We note that the result in Lemma 5(d) of Demetrescu *et al.* (2023) although established for finite-order stable AR filters can straightforwardly be shown to also be valid for 1-summable filters of the form considered here.

We next turn to establishing the first result given in part (ii) of Lemma 1 relating to the weakly persistent case. This case requires different normalisation of the partial sums from the strongly persistent case, but the proof steps are the essentially same. For example, we have for the numerator that

$$\frac{1}{\sqrt{T}}\sum_{j=2}^{\lfloor\tau T\rfloor} z_{j-1}\left(u_j-\bar{u}_j\right) = \frac{1}{\sqrt{T}}\sum_{j=2}^{\lfloor\tau T\rfloor} z_{j-1}u_j - \frac{1}{\sqrt{T}}\sum_{j=2}^{\lfloor\tau T\rfloor} z_{j-1}\bar{u}_j, \tag{S.3}$$

with the first summand on the right hand side of (S.3) weakly converging to a limiting Gaussian random variable, $M_2(\tau)$. Recalling that, in the weakly persistent case, z_t is approximately ξ_t plus a rest term whose magnitude can be controlled for in the relevant sums, the form of $M_2(\tau)$ can be found obtained from Lemma 4 and its associated proof in Section C.2 of the supplementary appendix of Demetrescu *et al.* (2023). However, note that the coefficients b_j given there differ from those which apply here because they result from inverting a stable $AR(p)$ lag polynomial, whereas ours originate directly from the lag polynomial $B(L)$ in Assumption 3.1. Nevertheless, the assumption of 1-summability placed on those coefficients entails that the result from Demetrescu *et al.* (2023) also applies here, and we therefore omit its full derivation. Regarding the second summand on the right hand side of (S.3), it is not difficult to show that

$$\frac{1}{\sqrt{T}}\sum_{j=2}^{\lfloor\tau T\rfloor} z_{j-1}\bar{u}_j = \frac{1}{\sqrt{T}}\sum_{j=2}^{\lfloor\tau T\rfloor} \xi_{j-1}\bar{u}_j + o_p(1),$$

where the o_p term is uniform. Then, noting that $\xi_t = (1-\rho L)^{-1}B(L)v_t = \sum_{j\geq 0} c_j v_{t-j}$, up to an asymptotically dominated term arising from the initialisation at $\xi_0 = O_p(1)$, we

have that

$$\begin{aligned}\sum_{j=2}^{\lfloor \tau T\rfloor} \xi_{j-1}\bar{u}_j &= \sum_{t=2}^{\lfloor \tau T\rfloor}\left(\sum_{j=0}^{t-2} c_j v_{t-j-1}\right)\left(\frac{1}{t}\sum_{k=1}^{t} u_k\right)\\ &= \sum_{t=1}^{\lfloor \tau T\rfloor -1} v_t u_t \left(\sum_{j=t+1}^{\lfloor \tau T\rfloor}\frac{1}{j}c_{j-t-1}\right) + \sum_{t=2}^{\lfloor \tau T\rfloor} u_t \sum_{s=1}^{t-1} v_s \left(\sum_{j=t}^{\lfloor \tau T\rfloor}\frac{1}{j}c_{j-s-1}\right)\\ &\quad + \sum_{t=2}^{\lfloor \tau T\rfloor -1} v_t \sum_{s=1}^{t-1} u_s \left(\sum_{j=t+1}^{\lfloor \tau T\rfloor}\frac{1}{j}c_{j-t-1}\right)\\ &=: A^*_{\tau,T} + B^*_{\tau,T} + C^*_{\tau,T}.\end{aligned}$$

We now show these three terms all vanish uniformly in τ in large samples. To that end, we first observe that the coefficients c_j are 1-summable given that $|\rho| < 1$ is fixed and the b_j coefficients are 1-summable. We then have, similarly to the analysis of $A_{\tau,T}$ in the strongly persistent case, that

$$\begin{aligned}\sup_{0\le\tau\le\omega}\left|A^*_{\tau,T}\right| &\le \sup_{0\le\tau\le\omega}\sum_{t=1}^{\lfloor \tau T\rfloor -1}|v_t u_t|\left(\sum_{j=t+1}^{\lfloor \tau T\rfloor}\frac{1}{j}|c_{j-t-1}|\right) \le \sum_{t=1}^{\lfloor \omega T\rfloor -1}|v_t u_t|\,\frac{1}{t}\left(\sum_{j=t+1}^{\lfloor \omega T\rfloor}|c_{j-t-1}|\right)\\ &= O_p\left(\ln T\right) = o_p\left(T^{1/2}\right)\end{aligned}$$

since $\sum_{j=t+1}^{\lfloor \omega T\rfloor}|c_{j-t-1}|$ is uniformly bounded by $\sum_{j\ge 0}|c_j|$. Exploiting in the same manner the 1-summability of c_j, the terms $B^*_{\tau,T}$ and $C^*_{\tau,T}$ can also be shown to be asymptotically negligible (uniformly in τ) using arguments similar to those used to deal with $B_{\tau,T}$ and $C_{\tau,T}$ in the strongly persistent case, and we omit the details.

Moving on to the second result in part (i) of Lemma 1, relating to the strongly persistent case, we have that

$$\begin{aligned}\frac{1}{T^{1+\eta}}\sum_{j=2}^{\lfloor \tau T\rfloor} z_{j-1}^2\left(u_j-\bar{u}_j\right)^2 &= \frac{1}{T^{1+\eta}}\sum_{j=2}^{\lfloor \tau T\rfloor} z_{j-1}^2 u_j^2 - \frac{2}{T^{1+\eta}}\sum_{j=2}^{\lfloor \tau T\rfloor} z_{j-1}^2 u_j\bar{u}_j\\ &\quad + \frac{1}{T^{1+\eta}}\sum_{j=2}^{\lfloor \tau T\rfloor} z_{j-1}^2\bar{u}_j^2. \qquad \text{(S.4)}\end{aligned}$$

Consider first the third term on the right hand side of (S.4). For any $0 \le \tau \le \omega$,

$$\frac{1}{T^{1+\eta}}\sum_{j=2}^{\lfloor \tau T\rfloor} z_{j-1}^2\bar{u}_j^2 \le \frac{1}{T^{1+\eta}}\sum_{j=2}^{\lfloor \omega T\rfloor} z_{j-1}^2\bar{u}_j^2 \le \frac{1}{T}\max_{j=1,\ldots,\lfloor \omega T\rfloor}\left(\frac{z_{j-1}}{T^{\eta/2}}\right)^2\sum_{j=2}^{\lfloor \omega T\rfloor}\bar{u}_j^2$$

where $\max_{j=1,\ldots,\lfloor \omega T\rfloor}\frac{z_{j-1}^2}{T^\eta} = O_p(\sqrt{T})$. Also,

$$\mathrm{E}\left(\sum_{j=2}^{\lfloor \omega T\rfloor}\bar{u}_j^2\right) \le C\sum_{j=2}^{\lfloor \omega T\rfloor}\frac{1}{j} = O\left(\ln T\right),$$

such that, uniformly in τ,

$$\frac{1}{T^{1+\eta}}\sum_{j=2}^{\lfloor \tau T\rfloor} z_{j-1}^2\bar{u}_j^2 = o_p\left(1\right).$$

Turning to the second term on the right hand side of (S.4), using the Cauchy-Schwarz inequality, we have that

$$\sup_{0\leq\tau\leq\omega}\left|\frac{1}{T^{1+\eta}}\sum_{j=2}^{\lfloor \tau T\rfloor} z_{j-1}^2 u_j\bar{u}_j\right| \leq \frac{1}{T^{1+\eta}}\sum_{j=2}^{\lfloor \omega T\rfloor}\left|z_{j-1}^2 u_j\bar{u}_j\right| \leq \sqrt{\frac{1}{T^{2+2\eta}}\sum_{j=2}^{\lfloor \omega T\rfloor} z_{j-1}^2 u_j^2 \sum_{j=2}^{\lfloor \omega T\rfloor} z_{j-1}^2\bar{u}_j^2}.$$

Consequently, we have that, uniformly in τ,

$$\frac{1}{T^{1+\eta}}\sum_{j=2}^{\lfloor \tau T\rfloor} z_{j-1}^2\left(u_j-\bar{u}_j\right)^2 = \frac{1}{T^{1+\eta}}\sum_{j=2}^{\lfloor \tau T\rfloor} z_{j-1}^2 u_j^2 + o_p(1).$$

The first term on the right hand side converges to the quadratic variation process, $[M_1](\tau)$, as required; this result is a by-product of the proof of Lemma 5(d) of Demetrescu *et al.* (2023), where the weak convergence of the normalised partial sums of $z_{j-1}u_j$ to $M_1(\tau)$ is established for the strongly persistent case.

The weak convergence of the partial sums of $(u_t, v_t)'$ is established in Demetrescu *et al.* (2023); notice that the partial sums of $(u_t, v_t)'$ are the same under strong and weak persistence and so the part of Lemma 4 of Demetrescu *et al.* (2023) referring to the partial sums of $(u_t, v_t)'$ holds under either type of persistence. Notice further that convergence to $M_v(\tau)$ occurs jointly with the previous weak convergence results; again see Demetrescu *et al.* (2023). Consequently, convergence to $J_{c,\mathrm{H}}$ also occurs jointly in the strongly persistent case.

The convergence of the partial sums of squares in the weakly persistent case follows along the same lines and we omit the details.

S.1.2 Proof of Lemma 2

We build on the following probability statement for the exceedances of a generic standard Brownian motion process, denoted $W(s)$, from Chu *et al.* (1996, eq. (7))

$$\mathrm{P}\left(|W(s)| > \sqrt{s\left(a^2+\ln s\right)} \text{ for some } s\geq 1\right) = 2\left(1-\Phi(a)+a\phi(a)\right).$$

The event $|W(s)| > \sqrt{s\left(a^2+\ln s\right)}$ is invariant to one-to-one transformations of the time, s. Noting that the two variance profiles, $\pi_i(\cdot)$, $i=1,2$, (in each case extended to the real positive half-line such that it possesses a uniformly bounded derivative at all but a countable number of points), are both continuous and strictly increasing, and therefore constitute one-to-one mappings. Consequently, using the substitution $s=\pi_i(\tau)$, and

noting that $\pi_i(1) = 1$, by construction, it follows that

$$\mathrm{P}\left(|W(s)| > \sqrt{s\left(a^2 + \ln s\right)} \text{ for some } s \geq 1\right) = \mathrm{P}\left(W^2(\pi_i(\tau)) > \pi_i(\tau)\left(a^2 + \ln \pi_i(\tau)\right) \text{ for some } \pi_i(\tau) \geq 1\right)$$

Given the distributional equivalence, $M_i(\tau) \stackrel{d}{=} \sqrt{[M_i](1)}\, W(\pi_i(\tau))$, $i = 1, 2$, and the monotonicity of $\pi_i(\cdot)$, we therefore have that, for both $i = 1$ and $i = 2$,

$$\mathrm{P}\left(\frac{M_i^2(\tau)}{[M_i](1)} > \pi_i(\tau)\left(a^2 + \ln \pi_i(\tau)\right) \text{ for some } \tau \geq 1\right) = 2\left(1 - \Phi(a) + a\phi(a)\right)$$

and the desired result follows noting that $\pi_i(\tau) = [M_i](\tau)/[M_i](1)$, $i = 1, 2$, and that $[M_i](1)$ is almost surely positive, $i = 1, 2$.

S.1.3 Proof of Proposition 1

We first examine the case of strong persistence. Here it follows from item (i) of Lemma 1 that, for $\tau \in [0, \omega]$, $\hat{\pi}(\tau) \Rightarrow \pi_1(\tau)$. Then, let

$$\begin{aligned} Q_t^2 \;:=\; & \frac{1}{T^{1+\eta}}\left(\sum_{j=1}^{t} z_{j-1}\left(u_j - \bar{u}_j\right)\right)^2 \\ & -\frac{1}{T^{1+\eta}}\sum_{j=1}^{t} z_{j-1}^2\left(y_j - \bar{y}_j\right)^2\left(a^2 + \ln \hat{\pi}(t/T)\right) + [M_1](t/T)\left(a^2 + \ln \pi_1(t/T)\right). \end{aligned}$$

Notice that the event $\mathcal{W}_t > b_{t,\hat{\pi}}$ (for some $T + 1 \leq t \leq E$) is equivalent to the event

$$Q_t^2 \geq [M_1](t/T)\left(a^2 + \ln \pi_1(t/T)\right) \quad \text{for some } T + 1 \leq t \leq E$$

such that

$$\mathrm{P}\left(\mathcal{W}_t \geq b_{t,\hat{\pi}} \text{ for some } T + 1 \leq t \leq E\right) = \mathrm{P}\left(Q_t^2 \geq [M_1](t/T)\left(a^2 + \ln \pi_1(t/T)\right) \text{ for some } T + 1 \leq t \leq E\right)$$

Lemma 1 also implies that, under the null hypothesis, $Q_{[\tau T]}^2 \Rightarrow M_1^2(\tau)$, such that, following the arguments of Chu *et al.* (1996), we may then conclude that the probability

$$\mathrm{P}\left(Q_t^2 \geq [M_1](t/T)\left(a^2 + \ln \pi_1(t/T)\right) \text{ for some } T + 1 \leq t \leq E\right)$$

converges as $T \to \infty$ to

$$\mathrm{P}\left(M_1^2(\tau) \geq [M_1](\tau)\left(a^2 + \ln \pi_1(\tau)\right) \text{ for some } 1 \leq \tau \leq \omega\right).$$

Finally, Lemma 2 implies

$$\begin{aligned} & \mathrm{P}\left(M_1^2(\tau) \geq [M_1](\tau)\left(a^2 + \ln \pi_1(\tau)\right) \text{ for some } 1 \leq \tau \leq \omega\right) \\ & \quad \leq \mathrm{P}\left(M_1^2(\tau) \geq [M_1](\tau)\left(a^2 + \ln \pi_1(\tau)\right) \text{ for some } 1 \leq \tau\right) \\ & \quad = 2\left(1 - \Phi(a) + a\phi(a)\right), \end{aligned}$$

as required.

Turning to the weakly persistent case, define

$$\begin{aligned} Q_t^2 \;:=\; & \frac{1}{T}\left(\sum_{j=1}^{t} z_{j-1}\left(u_j-\bar{u}_j\right)\right)^2 \\ & -\frac{1}{T}\sum_{j=1}^{t} z_{j-1}^2\left(y_j-\bar{y}_j\right)^2\left(a^2+\ln\hat{\pi}(t/T)\right)+[M_1](t/T)\left(a^2+\ln\pi_2(t/T)\right). \end{aligned}$$

The stated result then follows along the lines of the proof for the strongly persistent case above, now using item (ii) of Lemma 1 such that $\hat{\pi}(\tau) \Rightarrow \pi_2(\tau)$.

S.1.4 Proof of Lemma 3

First, from Chu *et al.* (1996, eq. (8)) we have the following probability statement for the exceedances of a standard Brownian motion processes,

$$\mathrm{P}\left(|W(\tau)| > \sqrt{(\tau+1)\left(\tilde{a}_\alpha+\ln(\tau+1)\right)} \text{ for some } \tau \geq 0\right) = e^{-\tilde{a}_\alpha/2}.$$

Using the change of variable, $s = \tau + 1$, and exploiting the distributional equivalence $W(s-1) \stackrel{d}{=} W(s) - W(1)$, $s \geq 1$, we obtain that

$$\mathrm{P}\left(|W(s)-W(1)| > \sqrt{s\left(\tilde{a}_\alpha+\ln s\right)} \text{ for some } s \geq 1\right) = e^{-\tilde{a}_\alpha/2}.$$

Again noting that the variance profiles, $\pi_i(\cdot)$, $i = 1, 2$, are both monotonic transformations, we have, as in the proof of Lemma 2, that

$$\mathrm{P}\left(|W(\pi_i(\tau))-W(\pi_i(1))| > \sqrt{\pi_i(\tau)\left(\tilde{a}_\alpha+\ln\pi_i(\tau)\right)} \text{ for some } \pi_i(\tau) \geq 1\right) = e^{-\tilde{a}_\alpha/2},$$

where we again note that $\pi_i(1) = 1$, by construction. Again using the distributional equivalence, $M_i(\tau)/\sqrt{[M_i](1)} \stackrel{d}{=} W(\pi(\tau))$, and noting that $\pi_i(\tau) \geq 1$ if and only if $\tau \geq 1$, the stated result follows.

S.1.5 Proof of Proposition 2

The proof steps are analogous to those for Proposition 1. Let $\pi(\cdot)$ denote the relevant variance profile, *i.e.* $\pi(\cdot) = \pi_1(\cdot)$ under strong persistence and $\pi(\cdot) = \pi_2(\cdot)$ under weak persistence, and analogously let $M(\cdot)$ denote the relevant limit process, *i.e.* $M(\cdot) = M_1(\cdot)$ under strong persistence and $M(\cdot) = M_2(\cdot)$ under weak persistence.

It follows from Lemma 1 that, irrespective of whether the predictor is weakly or strongly

persistent,

$$\hat{\pi}(\tau) \Rightarrow \pi(\tau) \qquad \text{and} \qquad \mathcal{M}_{\lfloor \tau T \rfloor} \Rightarrow \frac{(M(\tau) - M(1))^2}{[M](1)}. \tag{S.5}$$

Define

$$Q_t^2 := \mathcal{M}_t - \hat{\pi}\left(\frac{t}{T}\right)\left(\tilde{a}_\alpha + \ln \hat{\pi}\left(\frac{t}{T}\right)\right) + \pi\left(\frac{t}{T}\right)\left(\tilde{a}_\alpha + \ln \pi\left(\frac{t}{T}\right)\right)$$

such that

$$\mathrm{P}\left(\mathcal{M}_t \geq c^2_{t,\hat{\pi}} \text{ for some } T < t \leq E\right) = \mathrm{P}\left(Q_t^2 \geq \pi\left(\frac{t}{T}\right)\left(\tilde{a}_\alpha + \ln \pi\left(\frac{t}{T}\right)\right) \text{ for some } T < t \leq E\right).$$

From (S.5), it follows immediately that, for $\tau \in [1, \omega]$,

$$Q^2_{\lfloor \tau T \rfloor} \Rightarrow \frac{(M(\tau) - M(1))^2}{[M](1)},$$

so that we may conclude, analogously to the proof of Proposition 1, that the probability

$$\mathrm{P}\left(Q_t^2 \geq \pi\left(\frac{t}{T}\right)\left(\tilde{a}_\alpha + \ln \pi\left(\frac{t}{T}\right)\right) \text{ for some } T < t \leq E\right)$$

converges as $T \to \infty$ to

$$\mathrm{P}\left(\frac{(M(\tau) - M(1))^2}{[M](1)} \geq \pi(\tau)\left(\tilde{a}_\alpha + \ln \pi(\tau)\right) \text{ for some } 1 \leq \tau \leq \omega\right).$$

Finally, using the fact that $[M](1)$ is almost surely positive, Lemma 3 implies that

$$\begin{aligned}
&\mathrm{P}\left(\frac{(M(\tau) - M(1))^2}{[M](1)} \geq \pi(\tau)\left(\tilde{a}_\alpha + \ln \pi(\tau)\right) \text{ for some } 1 \leq \tau \leq \omega\right) \\
&\quad \leq \mathrm{P}\left((M(\tau) - M(1))^2 \geq [M](1)\pi(\tau)\left(\tilde{a}_\alpha + \ln \pi(\tau)\right) \text{ for some } 1 \leq \tau\right) = e^{-\tilde{a}_\alpha/2}
\end{aligned}$$

as required.

S.1.6 Proof of Lemma 4

Consider first the result for the sequence of $\mathcal{W}_{\lfloor \tau T \rfloor}$ statistics in the weakly persistent case. To that end, recalling that $\bar{u}_j = \frac{1}{j}\sum_{k=1}^{j} u_k$ and that $z_0 = 0$, we have for the (scaled) numerator of $\mathcal{W}_{\lfloor \tau T \rfloor}$ that

$$\frac{1}{\sqrt{T}}\sum_{j=1}^{\lfloor \tau T \rfloor} z_{j-1}(y_j - \bar{y}_j) = \frac{1}{\sqrt{T}}\sum_{j=2}^{\lfloor \tau T \rfloor} z_{j-1}(u_j - \bar{u}_j) + \frac{1}{T}\sum_{j=1}^{\lfloor \tau T \rfloor} z_{j-1}\left(b(j/T)x_{j-1} - \frac{1}{j}\sum_{k=1}^{j} b(k/T)x_{k-1}\right). \tag{S.6}$$

The first summand on the right hand side of (S.6) converges to $M_2(\tau)$, as demonstrated in Lemma 1. For the second summand, it is not difficult to show that

$$\frac{1}{T}\sum_{j=1}^{\lfloor \tau T \rfloor} z_{j-1} b(j/T) x_{j-1} = \frac{\mu_x}{T}\sum_{j=1}^{\lfloor \tau T \rfloor} z_{j-1} b(j/T) + \frac{1}{T}\sum_{j=1}^{\lfloor \tau T \rfloor} z_{j-1}\xi_j b(j/T) = \frac{1}{T}\sum_{j=1}^{\lfloor \tau T \rfloor} b(j/T)\xi^2_{j-1} + o_p(1)$$

noting that $z_{j-1} - \xi_{j-1}$ is negligible in the relevant summations, analogously to in the proof of Proposition 1 of Demetrescu *et al.* (2023), and that

$$\sum_{j=1}^{\lfloor \tau T \rfloor} \xi_{j-1} b(j/T) = O_p\left(\sqrt{T}\right) \tag{S.7}$$

uniformly in τ, as ξ_t is weakly persistent. Then, using the result established in the proof of Proposition 1 of Demetrescu *et al.* (2023), we have that the second summand on the right hand side of (S.6) weakly converges to $\kappa^2 \int_0^\tau b(s)\mathrm{d}[M_v](s)$. For the last summand on the right hand side of (S.6), write

$$\frac{1}{T}\sum_{j=1}^{\lfloor \tau T \rfloor} z_{j-1} \frac{1}{j}\sum_{k=1}^{j} b(k/T) x_{k-1} = \frac{\mu_x}{T}\sum_{j=1}^{\lfloor \tau T \rfloor} z_{j-1} \frac{1}{j}\sum_{k=1}^{j} b(k/T) + \frac{1}{T}\sum_{j=1}^{\lfloor \tau T \rfloor} z_{j-1} \frac{1}{j}\sum_{k=1}^{j} b(k/T)\xi_{k-1}$$

where $\frac{1}{j}\sum_{k=1}^{j} b(k/T)$ is uniformly bounded such that, analogously to (S.7), we have that

$$\sum_{j=1}^{\lfloor \tau T \rfloor} z_{j-1} \frac{1}{j}\sum_{k=1}^{j} b(k/T) = O_p\left(\sqrt{T}\right)$$

uniformly in τ, and

$$\mathrm{E}\left(\left|\frac{1}{T}\sum_{j=1}^{\lfloor \tau T \rfloor} \xi_{j-1} \frac{1}{j}\sum_{k=1}^{j} b(k/T)\xi_{k-1}\right|\right) \leq \frac{1}{T}\sum_{j=1}^{\lfloor \tau T \rfloor} \frac{1}{j}\sum_{k=1}^{j} |b(k/T)|\,\mathrm{E}\left(|\xi_{j-1}\xi_{k-1}|\right)$$

which, given the uniform boundedness of b, vanishes whenever the covariances of $|\xi_j|$ are absolutely summable. This is indeed the case here, given short memory of ξ_j under weak persistence. Then, Markov's inequality gives us that

$$\frac{1}{T}\sum_{j=1}^{\lfloor \tau T \rfloor} z_{j-1} \frac{1}{j}\sum_{k=1}^{j} b(k/T)\xi_{k-1} = o_p(1).$$

An application of the Continuous Mapping Theorem [CMT] then establishes that the (scaled) numerator of $\mathcal{W}_{\lfloor \tau T \rfloor}$ weakly converges to $\left(M_2(\tau) + \kappa^2 \int_0^\tau b(s)\mathrm{d}[M_v](s)\right)^2$.

Turning to the denominator of $\mathcal{W}_{\lfloor \tau T \rfloor}$, we have that

$$\frac{1}{T}\sum_{j=1}^{\lfloor \tau T \rfloor} z_{j-1}^2 (y_j - \bar{y}_j)^2 = \frac{1}{T}\sum_{j=1}^{\lfloor \tau T \rfloor} z_{j-1}^2 (u_j - \bar{u}_j)^2 + o_p(1)$$

uniformly in τ, as $y_j = u_j + b(j/T)x_{j-1}$ with $b(j/T)$ uniformly bounded and $\sup_{1\leq t\leq \omega T} |x_j| = O_p(T^{1/4})$ given the uniform L_4-boundedness of x_j. Consequently, under the local alternative, $H_{1,\eta_T,b(\cdot)}$ of (21), the suitably normalised cumulative sum of squares weakly converges to the quadratic variation process, $[M_2](\cdot)$, just as it does under the null hypothesis, H_0 of (3). An immediate and useful consequence of this is that $\hat{\pi}(\cdot)$ is consistent under weak persistence for $\pi_2(\cdot)$ under both the null hypothesis and local alternatives of the form in (21). A further application of the CMT then completes the proof. The corresponding result for the $\mathcal{M}_{\lfloor \tau T \rfloor}$ sequence of statistics follows immediately from the foregoing

analysis.

The results for the strongly persistent case builds analogously on the proof of Proposition 1 of Demetrescu *et al.* (2023) and we therefore omit the details here.

S.1.7 Proof of Lemma 5

Recall our choice $\tilde{\sigma}_j^2 = \sum_{s=0}^{\min\{h,j\}} \psi_s z_{j-1-s}^2 (y_{j-s} - \bar{y}_{j-s})^2$ and $\psi_s = \frac{K(s/h)}{\sum_{i=0}^{\min\{h,j\}} K(i/h)}$. Since $\tilde{\sigma}_1^2 = 0$ by construction, this results in

$$\begin{aligned}
\sum_{j=1}^{\lfloor \tau T \rfloor} \tilde{\sigma}_j^2 \sum_{j=2}^{\lfloor \tau T \rfloor} \tilde{\sigma}_j^2 &= \sum_{j=2}^{\lfloor \tau T \rfloor} \left(\sum_{s=0}^{h} \psi_s z_{j-1-s}^2 (y_{j-s} - \bar{y}_{j-s})^2 \right) \\
&= \psi_0 z_1^2 (y_2 - \bar{y}_2)^2 \\
&\quad + \psi_0 z_2^2 (y_3 - \bar{y}_3)^2 + \psi_1 z_1^2 (y_2 - \bar{y}_2)^2 \\
&\quad + \psi_0 z_3^2 (y_4 - \bar{y}_4)^2 + \psi_1 z_2^2 (y_3 - \bar{y}_3)^2 + \psi_2 z_1^2 (y_2 - \bar{y}_2)^2 \\
&\quad \dots \\
&\quad + \psi_0 z_{\lfloor \tau T \rfloor - 1}^2 \left(y_{\lfloor \tau T \rfloor} - \bar{y}_{\lfloor \tau T \rfloor} \right)^2 + \psi_1 z_{\lfloor \tau T \rfloor - 2}^2 \left(y_{\lfloor \tau T \rfloor - 1} - \bar{y}_{\lfloor \tau T \rfloor - 1} \right)^2 \\
&\qquad + \psi_2 z_{\lfloor \tau T \rfloor - 3}^2 \left(y_{\lfloor \tau T \rfloor - 2} - \bar{y}_{\lfloor \tau T \rfloor - 2} \right)^2 + \psi_3 z_{\lfloor \tau T \rfloor - 4}^2 \left(y_{\lfloor \tau T \rfloor - 3} - \bar{y}_{\lfloor \tau T \rfloor - 3} \right)^2 + \dots \\
&\qquad + \psi_h z_{\lfloor \tau T \rfloor - h - 1}^2 \left(y_{\lfloor \tau T \rfloor - h} - \bar{y}_{\lfloor \tau T \rfloor - h} \right)^2,
\end{aligned}$$

i.e. we implicitly use $\tilde{\sigma}_2^2 = 0$ due to our assumption that $K(0) = K(1) = 0$, where re-ordering the summands yields

$$\sum_{j=2}^{\lfloor \tau T \rfloor} \tilde{\sigma}_j^2 = \sum_{j=2}^{\lfloor \tau T \rfloor - h} z_{j-1}^2 (y_j - \bar{y}_j)^2 \left(\sum_{s=0}^{h} \psi_s \right) + \sum_{j=\lfloor \tau T \rfloor - h + 1}^{\lfloor \tau T \rfloor} z_{j-1}^2 (y_j - \bar{y}_j)^2 \left(\sum_{s=0}^{\lfloor \tau T \rfloor - j} \psi_s \right)$$

with $\sum_{s=0}^{h} \psi_s = 1$, by definition. Then, a comparison of $\sum_{j=2}^{\lfloor \tau T \rfloor} \tilde{\sigma}_j^2$ with its non-smooth analogue $\sum_{j=2}^{\lfloor \tau T \rfloor} z_{j-1}^2 (y_j - \bar{y}_j)^2$ indicates that

$$\begin{aligned}
\left| \sum_{j=2}^{\lfloor \tau T \rfloor} \tilde{\sigma}_j^2 - \sum_{j=2}^{\lfloor \tau T \rfloor} z_{j-1}^2 (y_j - \bar{y}_j)^2 \right| &= \left| \sum_{j=\lfloor \tau T \rfloor - h + 1}^{\lfloor \tau T \rfloor} z_{j-1}^2 (y_j - \bar{y}_j)^2 \left(\sum_{s=0}^{\lfloor \tau T \rfloor - j} \psi_s - 1 \right) \right| \\
&= \sum_{j=\lfloor \tau T \rfloor - h + 1}^{\lfloor \tau T \rfloor} z_{j-1}^2 (y_j - \bar{y}_j)^2 \left(1 - \sum_{s=0}^{\lfloor \tau T \rfloor - j} \psi_s \right) \\
&\leq \sup_{\tau \in [1,\omega]} \sum_{j=\lfloor \tau T \rfloor - h + 1}^{\lfloor \tau T \rfloor} z_{j-1}^2 (y_j - \bar{y}_j)^2
\end{aligned}$$

as $\left(\sum_{s=0}^{\lfloor \tau T \rfloor - j} \psi_s \right) \in [0, 1]$ for $j = \lfloor \tau T \rfloor - h + 1, \dots, \lfloor \tau T \rfloor$.

Under both the null hypothesis, H_0 of (3), and the local alternative, $H_{1,\eta_T,b(\cdot)}$ of (21), it

is straightforward to show that the difference between $\sup_{\tau\in[1,\omega]}\sum_{j=\lfloor\tau T\rfloor-h+1}^{\lfloor\tau T\rfloor} z_{j-1}^2\left(y_j-\bar{y}_j\right)^2$ and $\sup_{\tau\in[1,\omega]}\sum_{j=\lfloor\tau T\rfloor-h+1}^{\lfloor\tau T\rfloor} z_{j-1}^2\left(u_j-\bar{u}_j\right)^2$ is asymptotically negligible. The Cauchy-Schwarz inequality implies that $z_{j-1}^2\left(u_j-\bar{u}_j\right)^2$ is uniformly L_2 bounded in the weakly persistent case, and the same can be shown to hold for $T^{-\eta}z_{j-1}^2\left(u_j-\bar{u}_j\right)^2$ in the strongly persistent case. Noting that the supremum of a sequence of $\lfloor\omega T\rfloor$ uniformly bounded L_2 variables is of $O_p(\sqrt{\lfloor\omega T\rfloor})$, it follows that $\sup_{\tau\in[1,\omega]}\sum_{j=\lfloor\tau T\rfloor-h+1}^{\lfloor\tau T\rfloor} z_{j-1}^2\left(u_j-\bar{u}_j\right)^2$ is of $O_p\left(h\sqrt{T}\right)$ under weak persistence and of $O_p\left(hT^{\eta}\sqrt{T}\right)$ under strong persistence, and hence in each case is asymptotically negligible, provided the bandwidth satisfies the rate condition $h=o(\sqrt{T})$.

S.1.8 Proof of Lemma 6

Consider first the result in part (i) of Lemma 6 for the case of weakly persistent predictors. Here we have that

$$\frac{1}{\sqrt{T}}\sum_{j=T+1}^{\lfloor\tau T\rfloor}\frac{1}{\hat{\sigma}_j}z_{j-1}(u_j-\bar{u}_j)=\frac{1}{\sqrt{T}}\sum_{j=T+1}^{\lfloor\tau T\rfloor}\frac{1}{\sigma_2(j/T)}z_{j-1}(u_j-\bar{u}_j)+\frac{1}{\sqrt{T}}\sum_{j=T+1}^{\lfloor\tau T\rfloor}\left(\frac{1}{\hat{\sigma}_j}-\frac{1}{\sigma_2(j/T)}\right)z_{j-1}(u_j-\bar{u}_j)$$

where, as a consequence of Lemma 1, the first summand on the right hand side weakly converges:

$$\frac{1}{\sqrt{T}}\sum_{j=T+1}^{\lfloor\tau T\rfloor}\frac{1}{\sigma_2(j/T)}z_{j-1}(u_j-\bar{u}_j)\Rightarrow\int_1^{\tau}\frac{1}{\sigma_2(s)}\mathrm{d}M_2(s)\overset{d}{=}W(\tau)-W(1).$$

Turning to the second summand, this can be expressed as

$$\begin{aligned}&\frac{1}{\sqrt{T}}\sum_{j=T+1}^{\lfloor\tau T\rfloor}\left(\frac{1}{\hat{\sigma}_j}-\frac{1}{\sigma_2(j/T)}\right)z_{j-1}(u_j-\bar{u}_j)\\&\quad=\frac{1}{\sqrt{T}}\sum_{j=T+1}^{\lfloor\tau T\rfloor}\left(\frac{1}{\hat{\sigma}_j}-\frac{1}{\sigma_2(j/T)}\right)z_{j-1}u_j-\frac{1}{\sqrt{T}}\sum_{j=T+1}^{\lfloor\tau T\rfloor}\left(\frac{1}{\hat{\sigma}_j}-\frac{1}{\sigma_2(j/T)}\right)z_{j-1}\bar{u}_j.\end{aligned}\tag{S.8}$$

Noting that u_j possesses the MD property and that $\hat{\sigma}_j$ is adapted, the first summand on the right hand side of (S.8) cumulates MDs and, hence, Doob's martingale inequality implies that this first summand will be asymptotically negligible provided the following condition holds,

$$\frac{1}{T}\sum_{j=T+1}^{\lfloor\omega T\rfloor}\left(\frac{1}{\hat{\sigma}_j}-\frac{1}{\sigma_2(j/T)}\right)^2z_{j-1}^2u_j^2\xrightarrow{p}0.$$

This holds true because

$$\frac{1}{T}\sum_{j=T+1}^{\lfloor\omega T\rfloor}\left(\frac{1}{\hat{\sigma}_j}-\frac{1}{\sigma_2(j/T)}\right)^2z_{j-1}^2u_j^2\leq\sup_{T<j\leq\lfloor\omega T\rfloor}\left(\frac{1}{\hat{\sigma}_j}-\frac{1}{\sigma_2(j/T)}\right)^2\frac{1}{T}\sum_{j=T+1}^{\lfloor\omega T\rfloor}z_{j-1}^2u_j^2=o_p(1)$$

where we have used the fact that $\sum_{j=T+1}^{\lfloor \omega T \rfloor} z_{j-1}^2 u_j^2 = O_p(T)$, by virtue of Lemma 1, and that

$$\left(\frac{1}{\hat{\sigma}_j} - \frac{1}{\sigma_2(j/T)}\right)^2 = \left(\frac{\sigma_2^2(j/T) - \hat{\sigma}_j^2}{\hat{\sigma}_j \sigma_2(j/T)(\sigma_2(j/T) + \hat{\sigma}_j)}\right)^2 = \frac{\left(\sigma_2^2(j/T) - \hat{\sigma}_j^2\right)^2}{4\sigma_2^6(j/T) + o_p(1)} \tag{S.9}$$

where the $o_p(1)$ term is uniform in $T < j \leq \lfloor \omega T \rfloor$ by virtue of the assumption that $\hat{\sigma}_j^2$ is uniformly consistent for σ_2^2 with σ_2^2 bounded away from zero (and thus $\hat{\sigma}_j$ being uniformly consistent for σ_2). These assumptions in turn then ensure the ratio in (S.9) is of $o_p(1)$.

The second summand in (S.8) is also seen to be of $o_p(1)$ as

$$\left| \frac{1}{\sqrt{T}} \sum_{j=T+1}^{\lfloor \tau T \rfloor} \left(\frac{1}{\hat{\sigma}_j} - \frac{1}{\sigma_2(j/T)}\right) z_{j-1} \bar{u}_j \right| \leq \sup_{T<j\leq\lfloor \omega T \rfloor} \left| \frac{1}{\hat{\sigma}_j} - \frac{1}{\sigma_2(j/T)} \right| \sup_{T<j\leq\lfloor \omega T \rfloor} |\bar{u}_j| \frac{1}{\sqrt{T}} \sum_{j=T+1}^{\lfloor \tau T \rfloor} |z_{j-1}|$$

where, for $T < j \leq \lfloor \omega T \rfloor$,

$$\sup_j |\bar{u}_j| \leq \frac{1}{\sqrt{T}} \sup_j \left| \frac{1}{\sqrt{T}} \sum_{k=1}^{j} \right| = O_p(T^{-1/2})$$

due to the weak convergence of $\frac{1}{\sqrt{T}} \sum_{k=1}^{\lfloor \tau T \rfloor}$ to $M_u(\tau)$ established in Lemma 1. Markov's inequality and the Cauchy-Schwarz inequality imply that

$$\sum_{j=T+1}^{\lfloor \tau T \rfloor} |z_{j-1}| \leq \sum_{j=T+1}^{\lfloor \omega T \rfloor} |z_{j-1}| = O_p(T)$$

uniformly in τ. Consequently, the uniform consistency of $\hat{\sigma}_j^2$ coupled with the assumption that σ_2^2 is bounded and bounded away from zero, suffice to complete the result.

For the proof of part (i) of Lemma 6 in the strongly persistent case, the required normalisation is $T^{1/2+\eta/2}$, and the stated result follows using the same arguments as above for the weakly persistent case as, again from Lemma 1,

$$\frac{1}{T^{1+\eta}} \sum_{j=T+1}^{\lfloor \omega T \rfloor} z_{j-1}^2 u_j^2 = O_p(1)$$

and

$$\sum_{j=T+1}^{\lfloor \tau T \rfloor} |z_{j-1}| \leq \sqrt{\omega T \sum_{j=T+1}^{\lfloor \omega T \rfloor} z_{j-1}^2} = O_p(T^{1+\eta/2})$$

as required for the result in part (i) of Lemma 6.

Establishing the result in part (ii) of Lemma 6 is straightforward for both the weakly and strongly persistent cases, in each case following using similar arguments to those given above, and so we omit the details.

S.1.9 The processes $M_u(\tau)$, $M_v(\tau)$, $M_1(\tau)$ and $M_2(\tau)$

For the strongly persistent case, the joint properties of $M_u(\tau)$, $M_v(\tau)$ and $M_1(\tau)$ are as follows:

$$\begin{pmatrix} M_u(\tau) \\ M_v(\tau) \\ M_1(\tau) \end{pmatrix} := \begin{pmatrix} \int_0^\tau h_{11}(s)\mathrm{d}W_1(s) + \int_0^\tau h_{12}(s)\mathrm{d}W_2(s) \\ \int_0^\tau h_{21}(s)\mathrm{d}W_1(s) + \int_0^\tau h_{22}(s)\mathrm{d}W_2(s) \\ \frac{B(1)}{\sqrt{2a}} \int_0^\tau \sigma_u(s)\sigma_v(s)\mathrm{d}W_3(s) \end{pmatrix}$$

where $h_{ij}(\tau)$ is the (i,j)th element of $\mathbf{H}(\tau)$ from Assumption 3, $W_1(\tau)$, $W_2(\tau)$ and $W_3(\tau)$ are independent standard Brownian motion processes, and $\sigma_u^2(\tau)$ and $\sigma_v^2(\tau)$ are the derivatives of the quadratic variation processes of $M_u(\tau)$ and $M_v(\tau)$ respectively, the latter given as

$$\begin{pmatrix} [M_u](\tau) \\ [M_v](\tau) \end{pmatrix} = \int_0^\tau \begin{pmatrix} h_{11}^2(s) + h_{12}^2(s) \\ h_{21}^2(s) + h_{22}^2(s) \end{pmatrix} \mathrm{d}s.$$

For the weakly persistent case, the joint properties of $M_u(\tau)$, $M_v(\tau)$ and $M_2(\tau)$ are as follows:

$$\begin{pmatrix} M_u(\tau) \\ M_v(\tau) \\ M_2(\tau) \end{pmatrix} := \int_0^\tau \mathbf{G}(s)\mathrm{d}\boldsymbol{B}(s),$$

where

$$\mathbf{G}(\tau) = \begin{pmatrix} h_{11}(\tau) & h_{12}(\tau) & 0 & 0 & 0 & 0 \\ h_{21}(\tau) & h_{22}(\tau) & 0 & 0 & 0 & 0 \\ 0 & 0 & h_{11}(\tau)h_{21}(\tau) & h_{11}(\tau)h_{22}(\tau) & h_{12}(\tau)h_{21}(\tau) & h_{12}(\tau)h_{22}(\tau) \end{pmatrix}$$

with $\boldsymbol{B}(\tau)$ a 6-variate Brownian motion with covariance matrix equal to the covariance matrix of the vector

$$\boldsymbol{\psi}_t := \begin{pmatrix} a_{1,t} \\ a_{2,t} \\ a_{1,t}\sum_{j\geq 0} c_j a_{1,t-1-j} \\ a_{1,t}\sum_{j\geq 0} c_j a_{2,t-1-j} \\ a_{2,t}\sum_{j\geq 0} c_j a_{1,t-1-j} \\ a_{2,t}\sum_{j\geq 0} c_j a_{2,t-1-j} \end{pmatrix},$$

with $(a_{1,t}, a_{2,t})' =: \boldsymbol{a}_t$ and $\sum_{j\geq 0} c_j L^j = (1-\rho L)^{-1}B(L)$, $B(L) = \sum_{j\geq 0} b_j L^j$ being from Assumption 3.

S.2 Finite-Horizon Critical Value Generation

In practical implementations where the monitoring horizon is finite, we can adjust $\tilde{a}_\alpha$ such that we obtain an empirical FPR close to a chosen nominal level, α, at the end of the monitoring period. Given that we do not recommend using the FLUC-type monitoring procedures outlined in section 4.1, we will only discuss finite-horizon critical value generation in the context of the CUSUM-based monitoring procedures. Finite-horizon critical values for the FLUC-type procedures can be obtained using a similar approach, but as mentioned in section 5.1 the procedures based on these were more liberal than the procedures based on the asymptotic boundary function.

For a finite monitoring horizon, the relevant parameter that determines the empirical FPR of the monitoring procedure is the length of the monitoring period, relative to the length of the training period. We thus specify, *ex ante*, an end time ωT, where $\omega = E/T$ captures the relative length of the monitoring period; for example, $\omega = 1.5$ denotes a monitoring period that is half as long as the training period, while $\omega = 3$ refers to a monitoring period that is twice as long as the training period. The critical bound parameter, $\tilde{a}_\alpha$, is then adjusted using preliminary simulation methods such that the desired FPR level, α, approximately holds for the given value of ω.

To operationalise this approach, we use direct simulation of the limiting null distributions of the sequence of $\mathcal{M}_t$ statistics in (17). Because we do not know the true variance profile, we perform our simulations under the assumption of homoskedasticity. This does matter asymptotically because the term $\tilde{a}_\alpha$ is common to both the homoskedastic limiting boundary function in (19) and the corresponding heteroskedastic limiting boundary function in (20) (though, of course, the FPR approximation our procedure will provide would be expected to be better in the homoskedastic case *vis-à-vis* homoskedastic environments). We therefore directly simulate the limiting null distribution using the discretisation

$$\mathcal{M}_t^\infty = T^{-1}\left(\sum_{j=T+1}^{t} u_j\right)^2, \quad t = T+1, \ldots, \lfloor \omega T \rfloor,$$

setting $T = 1000$ and $u_t \sim NIID(0,1)$, the latter generated using the Threefry 4x64 random number generator on MATLAB 2024b. We then compare $\mathcal{M}_t^\infty$ with the critical bound

$$c_t^2 = \frac{t}{T}\left(\tilde{a}_{\alpha,\omega} + \ln\left(\frac{t}{T}\right)\right). \tag{S.10}$$

This is repeated over $n = 10000$ Monte Carlo replications to yield an approximation of the empirical FPR for this choice of ω. Next, $\tilde{a}_{\alpha,\omega}$ is raised (reduced) if the resulting empirical FPR is larger (smaller) than the desired theoretical FPR, α, yielding some $\tilde{a}_{\alpha,\omega}^{\text{new}}$. These steps are then repeated until the empirical FPR resulting from $\tilde{a}_{\alpha,\omega}^*$ is within 0.001 of the theoretical FPR, α, for the given α-ω combination involved, or until 250 iterations have been completed (in which case the value at the last iteration is used).[2]

[2]The adjustment from one iteration to the next proceeds as follows. Given the previous critical bound based on some $\tilde{a}_{\alpha,\omega} = -2\ln(\hat{\alpha}_{\alpha,\omega})$, the adjustment depends on the resulting empirical FPR, $\hat{\alpha}$, and the target FPR, α, via $\Delta = \alpha/\hat{\alpha}$ where Δ is constrained to $[0.9, 1.1]$ to prevent overshooting the target value (recall the nonlinear relationship between $\tilde{a}_{\alpha,\omega}$ and $\hat{\alpha}_{\alpha,\omega}$). Then, the new critical bound

This exercise was repeated for each possible combination of $\alpha \in \{0.005, 0.01, 0.02, \ldots, 0.21, 0.22\}$ and $\omega \in \{1.2, 1.4, 1.6, \ldots, 5, 5.5, 6, 6.5, \ldots, 21\}$,[3] to which we fitted the following piecewise polynomial response surface:

$$\hat{\alpha}^*_{\alpha,\omega} = \begin{cases} 1.979 - 2.014\,\omega + 5.994\,\alpha + 0.7614\,\omega^2 + 1.264\,\omega\alpha - 62.03\,\alpha^2 \\ \quad - 0.1252\,\omega^3 - 0.6822\,\omega^2\alpha + 9.155\,\omega\alpha^2 + 213.2\,\alpha^3 \\ \quad + 0.007549\,\omega^4 + 0.06527\,\omega^3\alpha - 0.1997\,\omega^2\alpha^2 \\ \quad - 18.85\,\omega\alpha^3 - 257.5\,\alpha^4 & \text{for } \omega < 5 \\ 0.09331 - 0.01168\,\omega + 2.197\,\alpha + 0.0003751\,\omega^2 \\ \quad - 0.02321\,\omega\alpha - 1.866\,\alpha^2 & \text{for } 5 \leq \omega \leq 21 \end{cases} \tag{S.11}$$

where $\tilde{a}^*_{\alpha,\omega} = -2\ln(\hat{\alpha}^*_{\alpha,\omega})$. The response surface in (S.11) is formulated in terms of $\hat{\alpha}^*_{\alpha,\omega}$, rather than $\tilde{a}^*_{\alpha,\omega}$, because this yielded a superior response surface fit. The R^2 for the fitted response surface for $\omega < 5$ was 0.9929, while that for the fitted response surface for $5 \leq \omega \leq 21$ was 0.9969.

The value of $\tilde{a}^*_{\alpha,\omega}$ to be used is based on evaluating (S.11) for the given values of α and ω, to obtain the value $\hat{\alpha}^*_{\alpha,\omega}$, and we then calculate $\tilde{a}^*_{\alpha,\omega} = -2\ln(\hat{\alpha}^*_{\alpha,\omega})$. The finite-horizon critical value boundary function, c^{*2}_t, say, for $t = T+1, \ldots, \lfloor \omega T \rfloor$ is then given by (20) evaluated at $\tilde{a}_\alpha = \tilde{a}^*_{\alpha,\omega}$, for the given choice of variance profile estimator adopted (that is, either $\hat{\pi}$ or $\tilde{\pi}$).

Remark S.1 It is important to emphasise that continuing to monitor beyond time $t = \lfloor \omega T \rfloor$ would be expected to lead to an exceedance of the nominal FPR, α, when working with the finite-horizon critical bounds generated, as outlined above, for the given value of ω. □

Remark S.2 In cases where $\omega > 21$, we recommend using the asymptotic boundary function in (20), for the given choice of variance profile estimate, setting $\tilde{a}_\alpha = -2\ln\alpha$ for the chosen nominal FPR level, α. □

parameter for the next iteration is given by $\tilde{a}^{\text{new}}_{\alpha,\omega} = -2\ln(\hat{\alpha}^{\text{new}}_{\alpha,\omega})$ where $\hat{\alpha}^{\text{new}}_{\alpha,\omega} = \Delta \cdot \hat{\alpha}_{\alpha,\omega}$.

[3]The simulation proceeded in two nested loops: first along ω, where each matched $\hat{\alpha}^*_{\alpha,\omega}$ value served as the starting point for the next step; and then along α, incrementing after completing all steps for a given ω range. For example, the result from the search procedure for $(\omega = 1.2, \alpha = 0.005)$ initialised the search procedure for $(\omega = 1.4, \alpha = 0.005)$, and similarly for higher values of ω. After iterating over ω, the search result for $(\omega = 1.2, \alpha = 0.005)$ initialises the next ω search loop, i.e. the search result for $(\omega = 1.2, \alpha = 0.005)$ initialises the ω search loop for $\alpha = 0.01$, the result for the search procedure for $(\omega = 1.2, \alpha = 0.01)$ initialises the search loop starting at $(\omega = 1.2, \alpha = 0.02)$, etc. The entire search procedure is initialised using $\tilde{a}_\alpha = -2\ln(0.6)$ as a starting value for the search procedure for $(\omega = 1.2, \alpha = 0.005)$.

S.3 Additional Monte Carlo Results

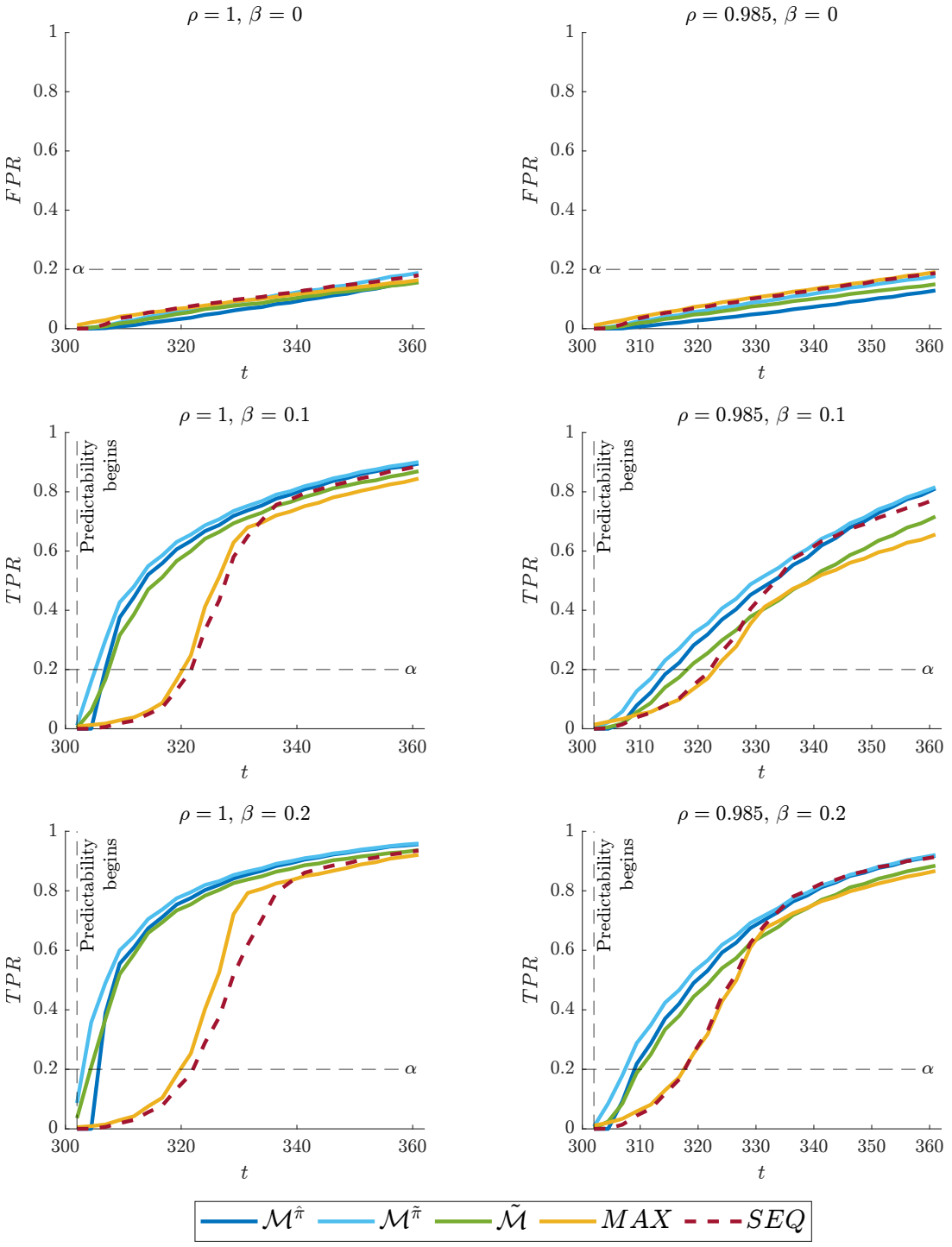

Figure S.1: FPR/TPR curves of $\mathcal{M}^{\hat{\pi}}, \mathcal{M}^{\tilde{\pi}}, \widetilde{\mathcal{M}}, MAX$ and SEQ for $\rho \in \{1, 0.985\}$ and $\beta \in \{0, 0.1, 0.2\}$ and data generated from DGP2 (GARCH(1,1)).

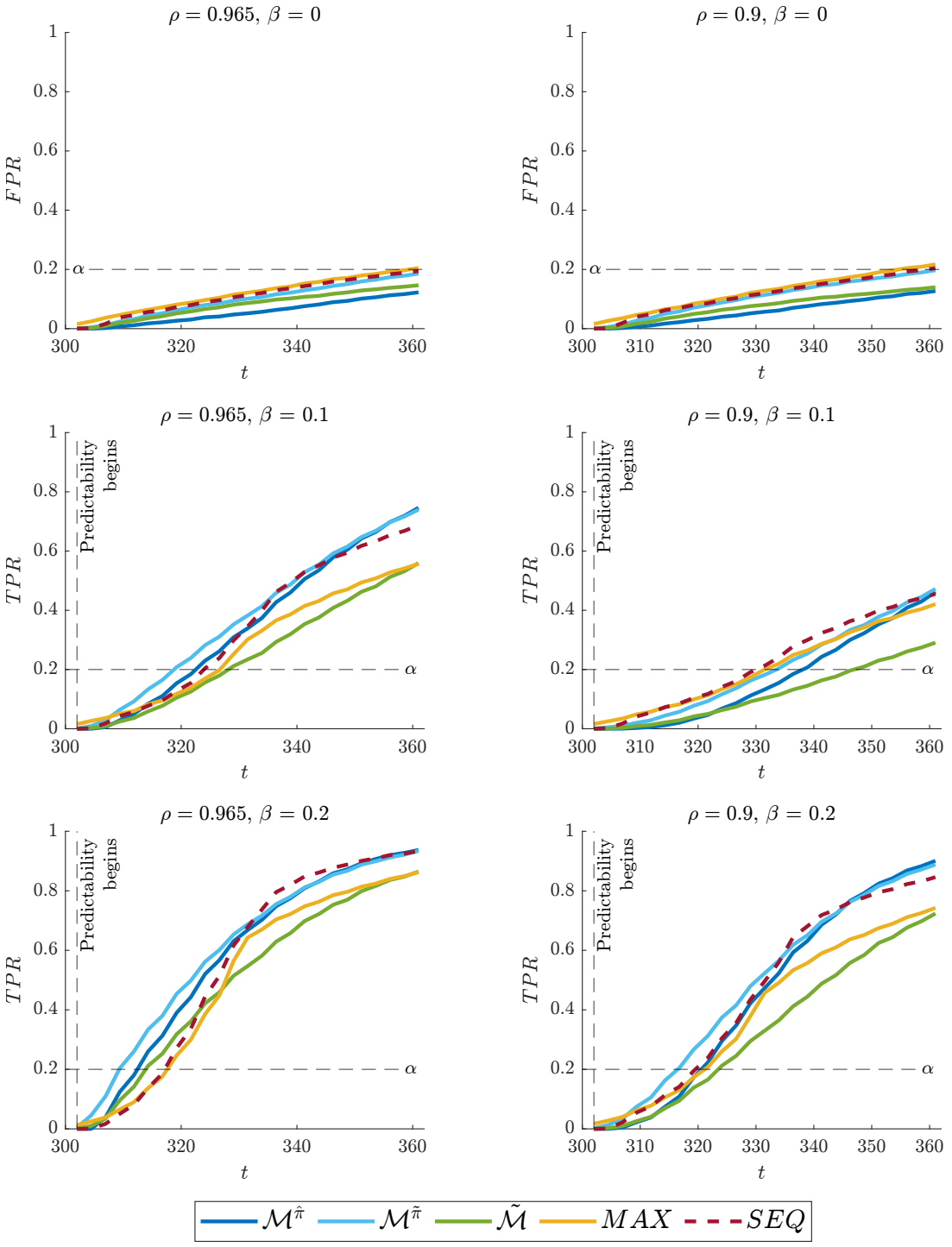

Figure S.2: FPR/TPR curves of $\mathcal{M}^{\hat{\pi}}, \mathcal{M}^{\tilde{\pi}}, \widetilde{\mathcal{M}}, MAX$ and SEQ for $\rho \in \{0.965, 0.9\}$ and $\beta \in \{0, 0.1, 0.2\}$ and data generated from DGP2 (GARCH(1,1)).

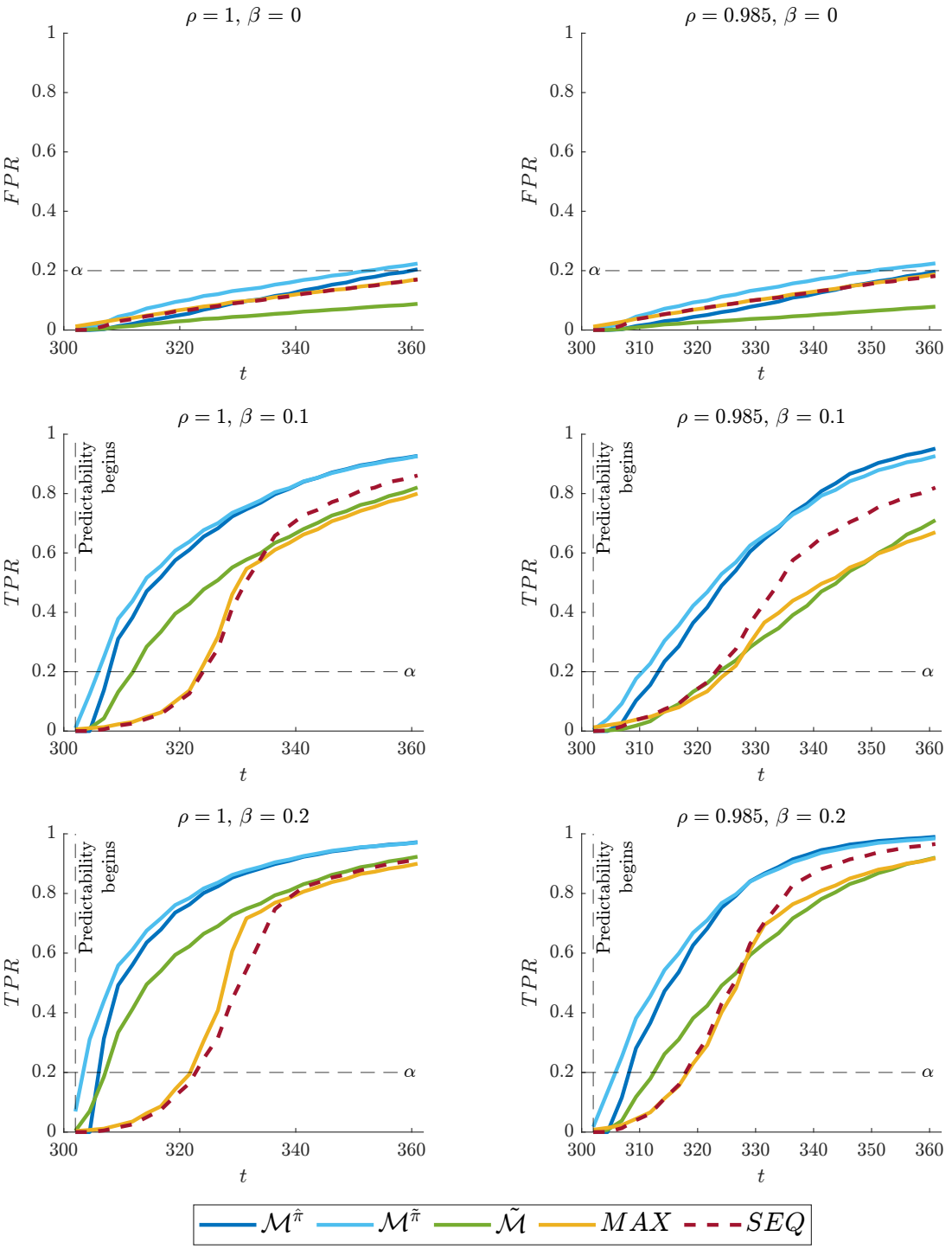

Figure S.3: FPR/TPR curves of $\mathcal{M}^{\hat{\pi}}, \mathcal{M}^{\tilde{\pi}}, \widetilde{\mathcal{M}}, MAX$ and SEQ for $\rho \in \{1, 0.985\}$ and $\beta \in \{0, 0.1, 0.2\}$ and data generated from DGP5 (training period upward volatility transition).

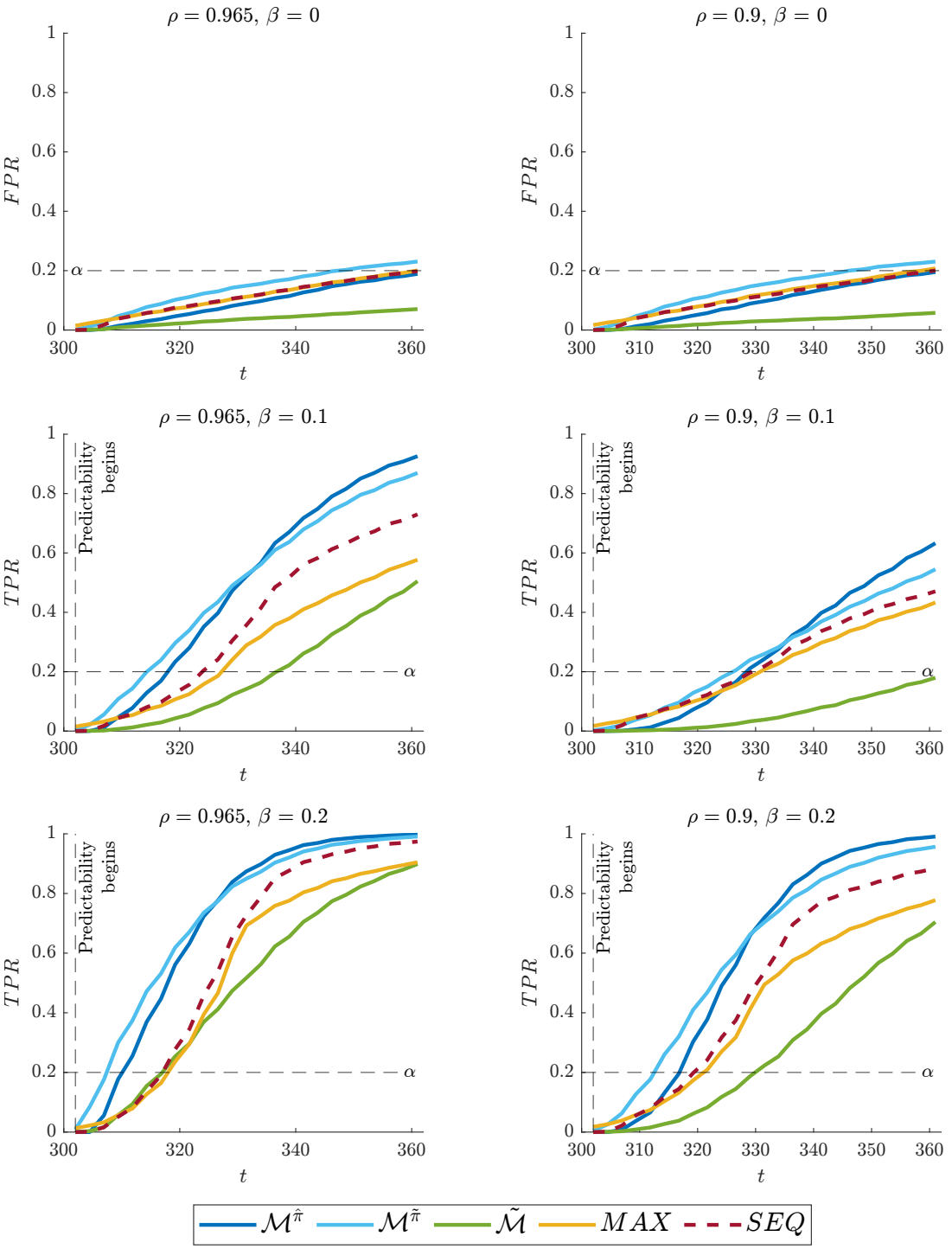

Figure S.4: FPR/TPR curves of $\mathcal{M}^{\hat{\pi}}, \mathcal{M}^{\tilde{\pi}}, \widetilde{\mathcal{M}}, MAX$ and SEQ for $\rho \in \{0.965, 0.9\}$ and $\beta \in \{0, 0.1, 0.2\}$ and data generated from DGP5 (training period upward volatility transition).

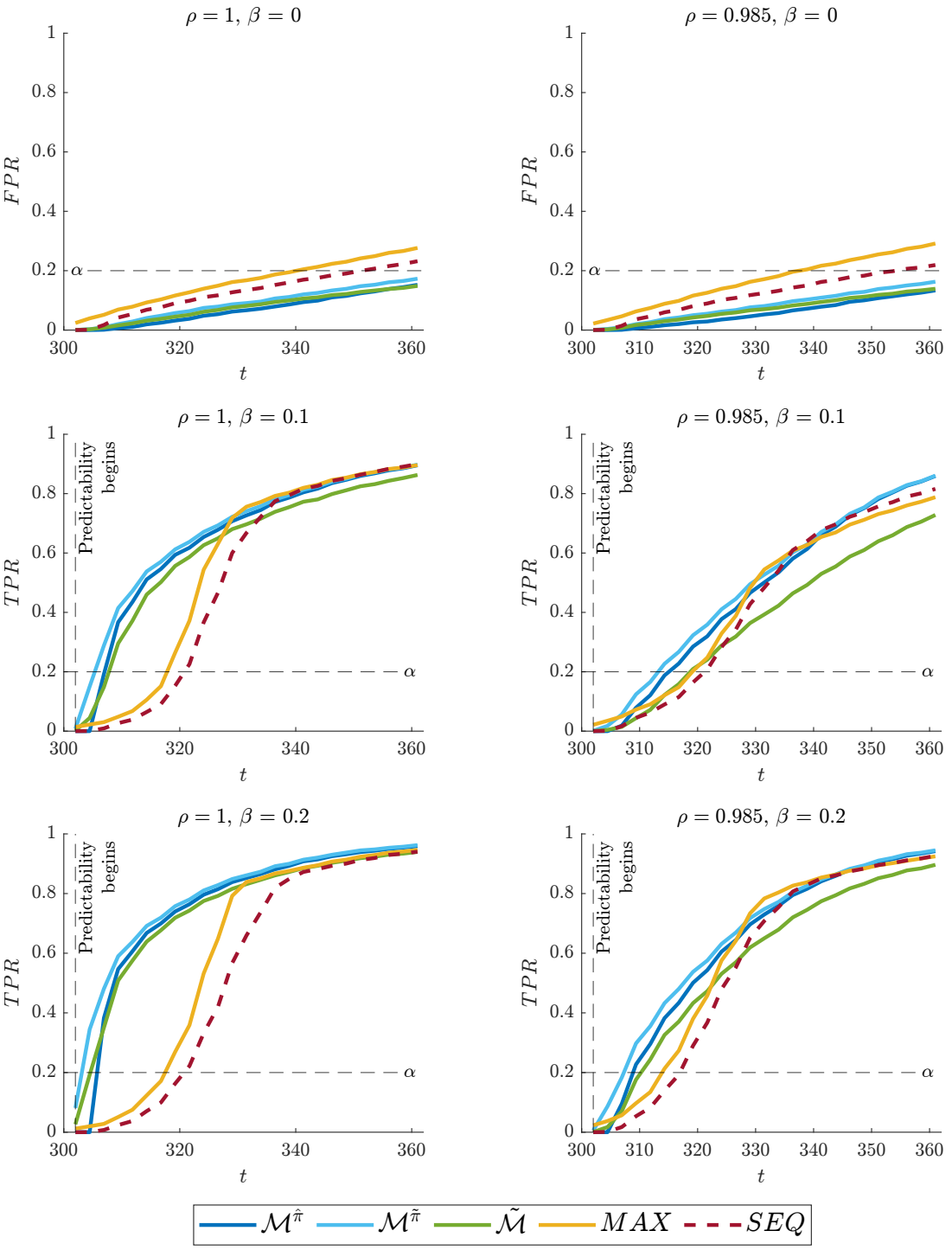

Figure S.5: FPR/TPR curves of $\mathcal{M}^{\hat{\pi}}, \mathcal{M}^{\tilde{\pi}}, \widetilde{\mathcal{M}}, MAX$ and SEQ for $\rho \in \{1, 0.985\}$ and $\beta \in \{0, 0.1, 0.2\}$ and data generated from DGP6 (training period endogeneity transition).

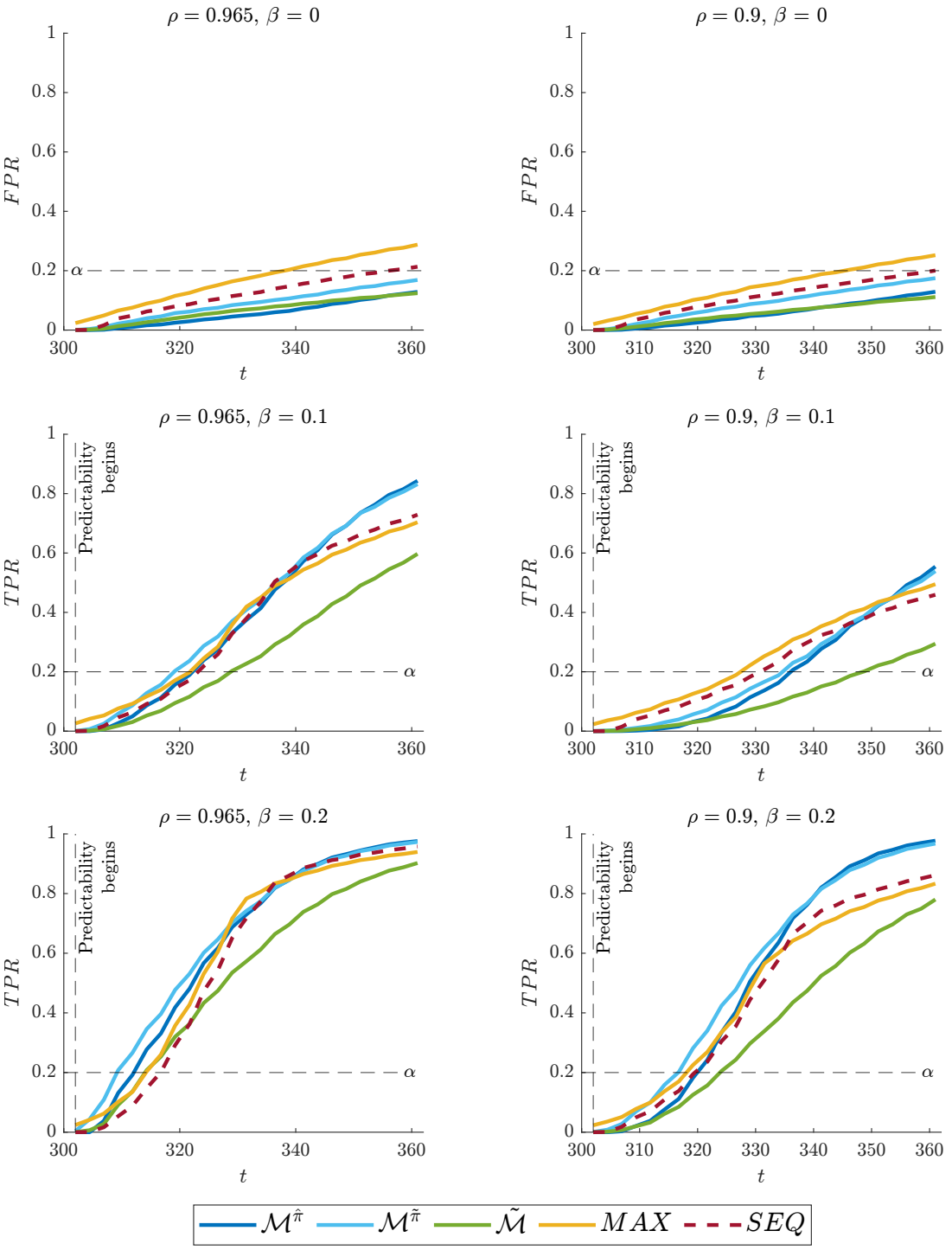

Figure S.6: FPR/TPR curves of $\mathcal{M}^{\hat{\pi}}, \mathcal{M}^{\tilde{\pi}}, \widetilde{\mathcal{M}}, MAX$ and SEQ for $\rho \in \{0.965, 0.9\}$ and $\beta \in \{0, 0.1, 0.2\}$ and data generated from DGP6 (training period endogeneity transition).

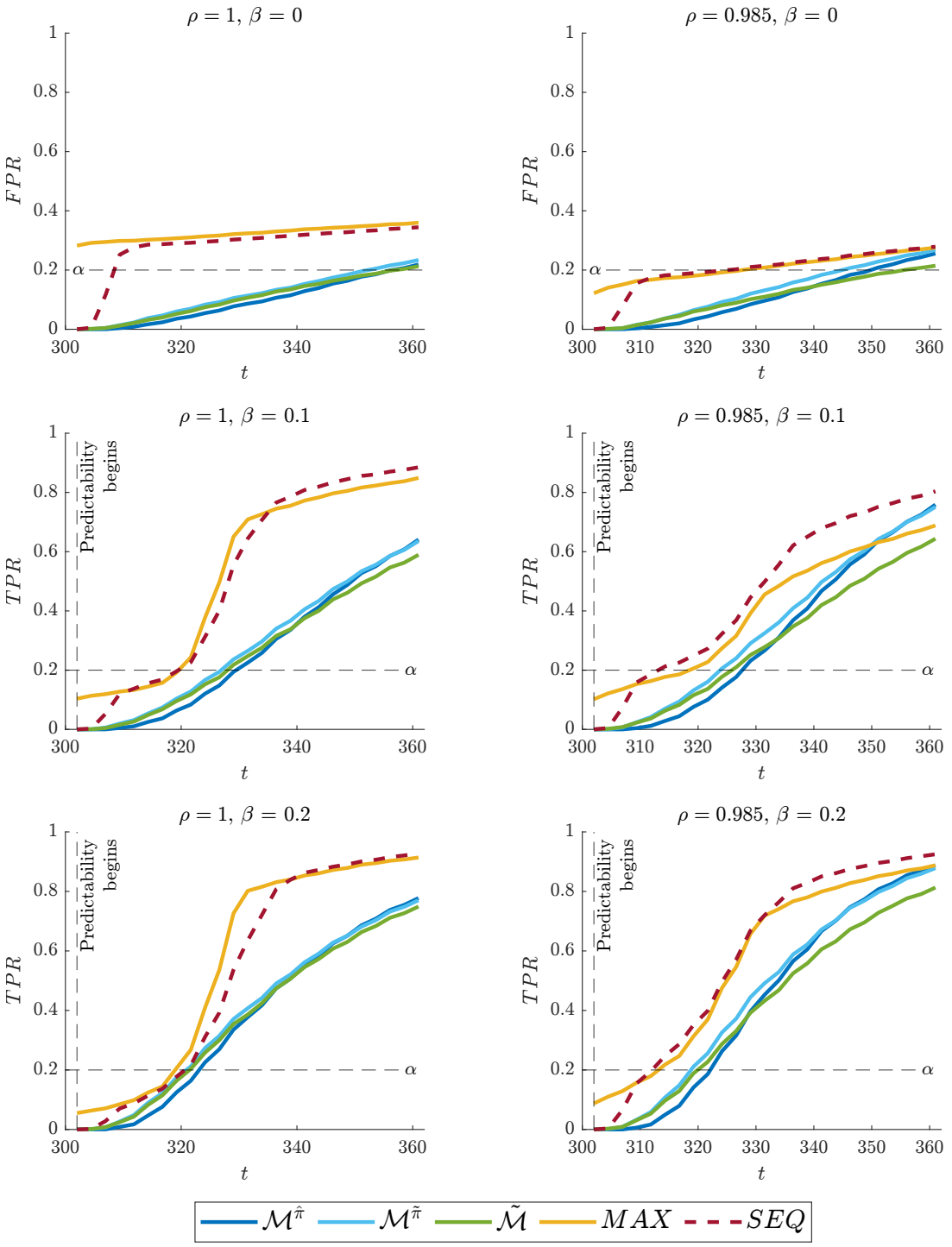

Figure S.7: FPR/TPR curves of $\mathcal{M}^{\hat{\pi}}, \mathcal{M}^{\tilde{\pi}}, \widetilde{\mathcal{M}}, MAX$ and SEQ for $\rho \in \{1, 0.985\}$ and $\beta \in \{0, 0.1, 0.2\}$ and data generated from DGP7 (predictability in the training period). Calculation of $\bar{y}_j$ re-initialized at $t = T + 1$.

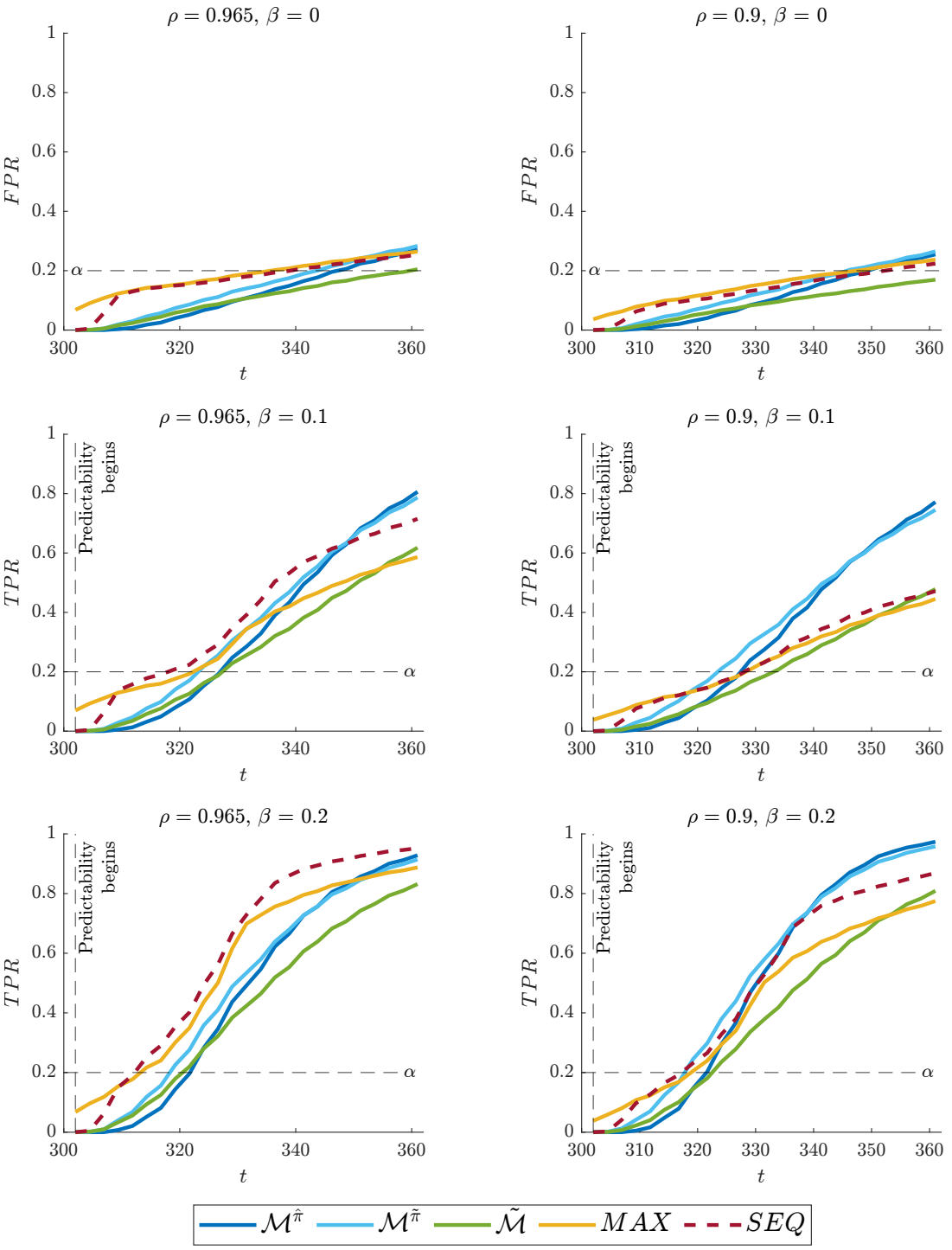

Figure S.8: FPR/TPR curves of $\mathcal{M}^{\hat{\pi}}, \mathcal{M}^{\tilde{\pi}}, \widetilde{\mathcal{M}}, MAX$ and SEQ for $\rho \in \{0.965, 0.9\}$ and $\beta \in \{0, 0.1, 0.2\}$ and data generated from DGP7 (predictability in the training period). Calculation of $\bar{y}_j$ re-initialized at $t = T + 1$.

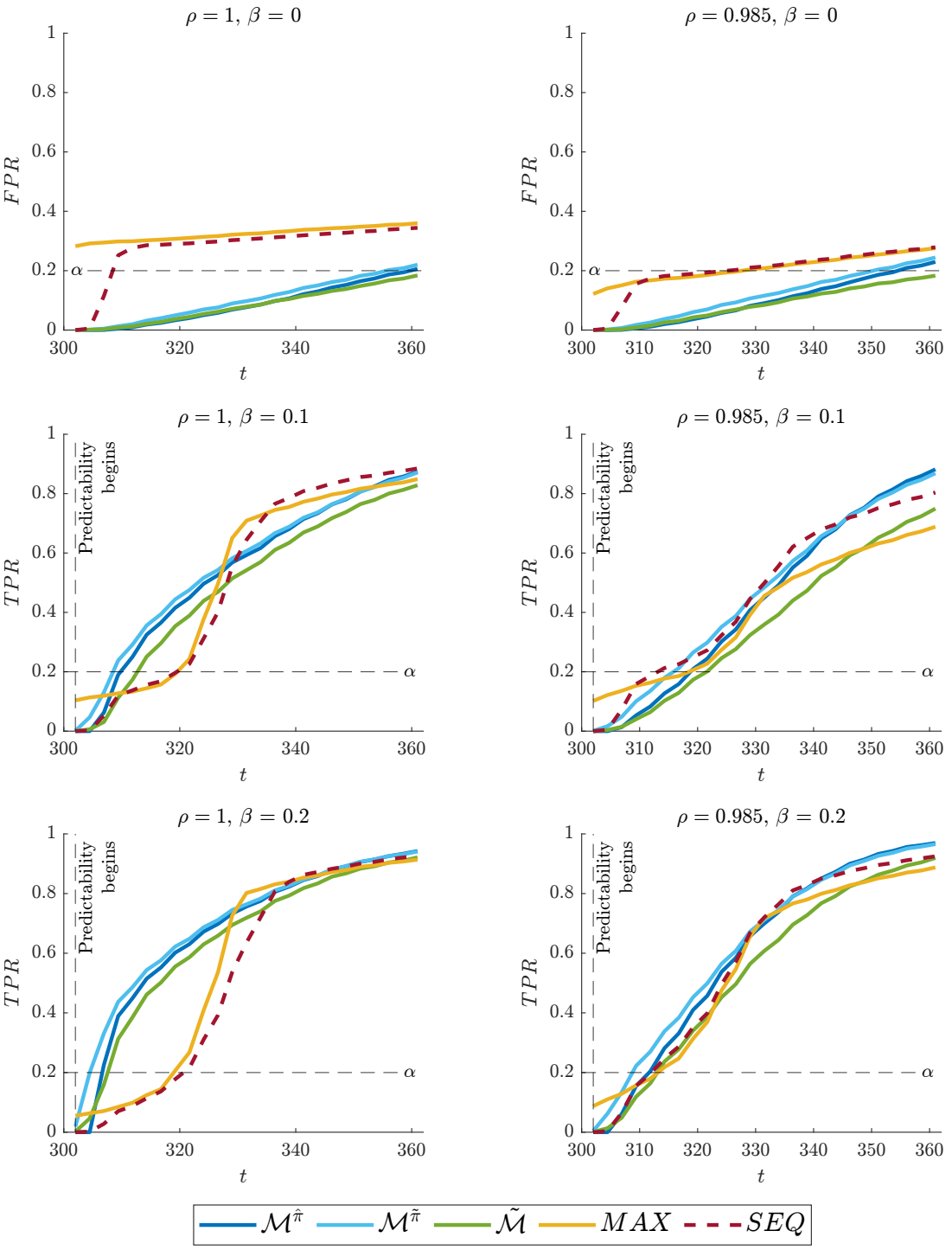

Figure S.9: FPR/TPR curves of $\mathcal{M}^{\hat{\pi}}, \mathcal{M}^{\tilde{\pi}}, \widetilde{\mathcal{M}}, MAX$ and SEQ for $\rho \in \{1, 0.985\}$ and $\beta \in \{0, 0.1, 0.2\}$ and data generated from DGP7 (predictability in the training period). Calculation of $\bar{y}_j$ re-initialized at $t = T + 1 - 6$.

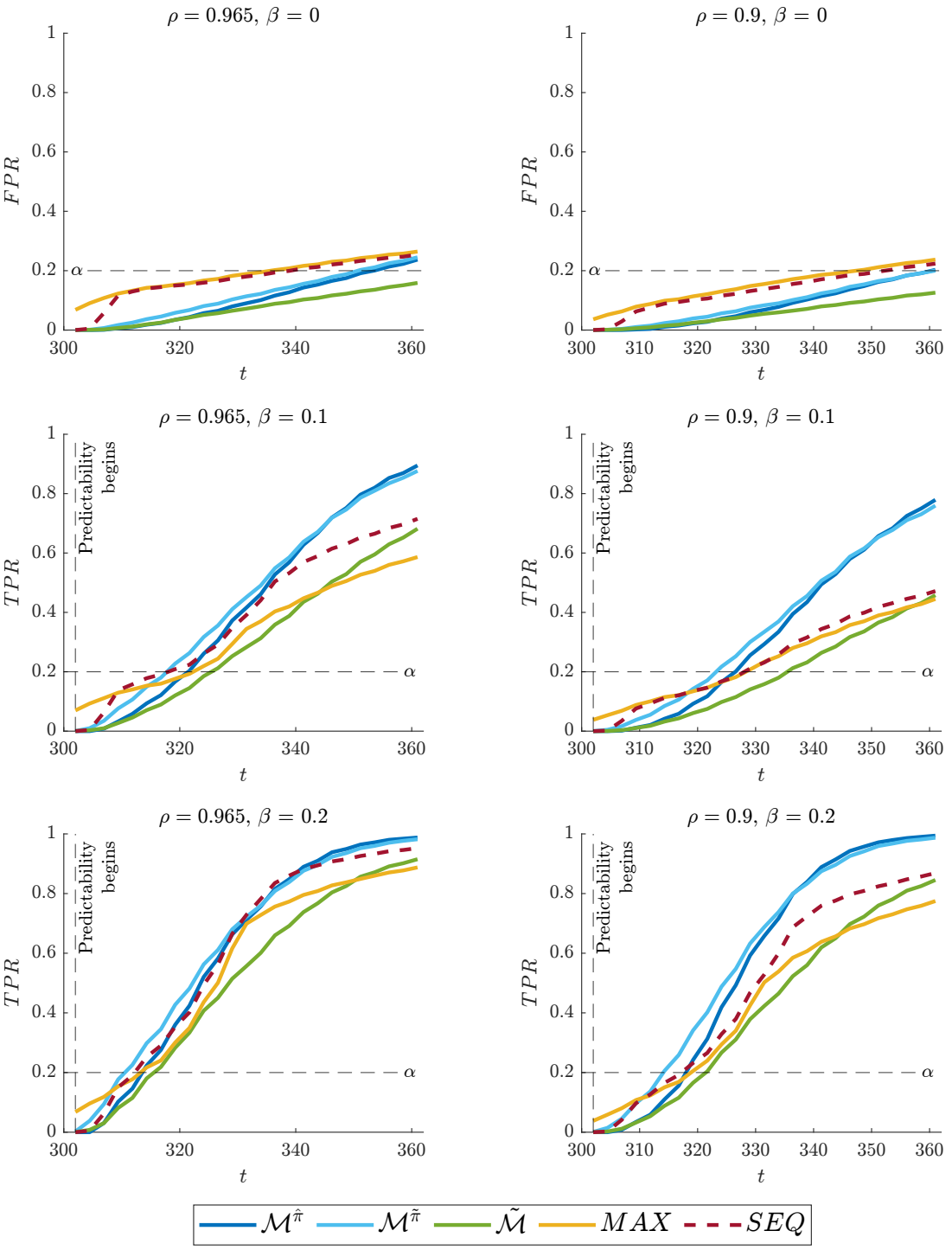

Figure S.10: FPR/TPR curves of $\mathcal{M}^{\hat{\pi}}, \mathcal{M}^{\tilde{\pi}}, \widetilde{\mathcal{M}}, MAX$ and SEQ for $\rho \in \{0.965, 0.9\}$ and $\beta \in \{0, 0.1, 0.2\}$ and data generated from DGP7 (predictability in the training period). Calculation of $\bar{y}_j$ re-initialized at $t = T + 1 - 6$.

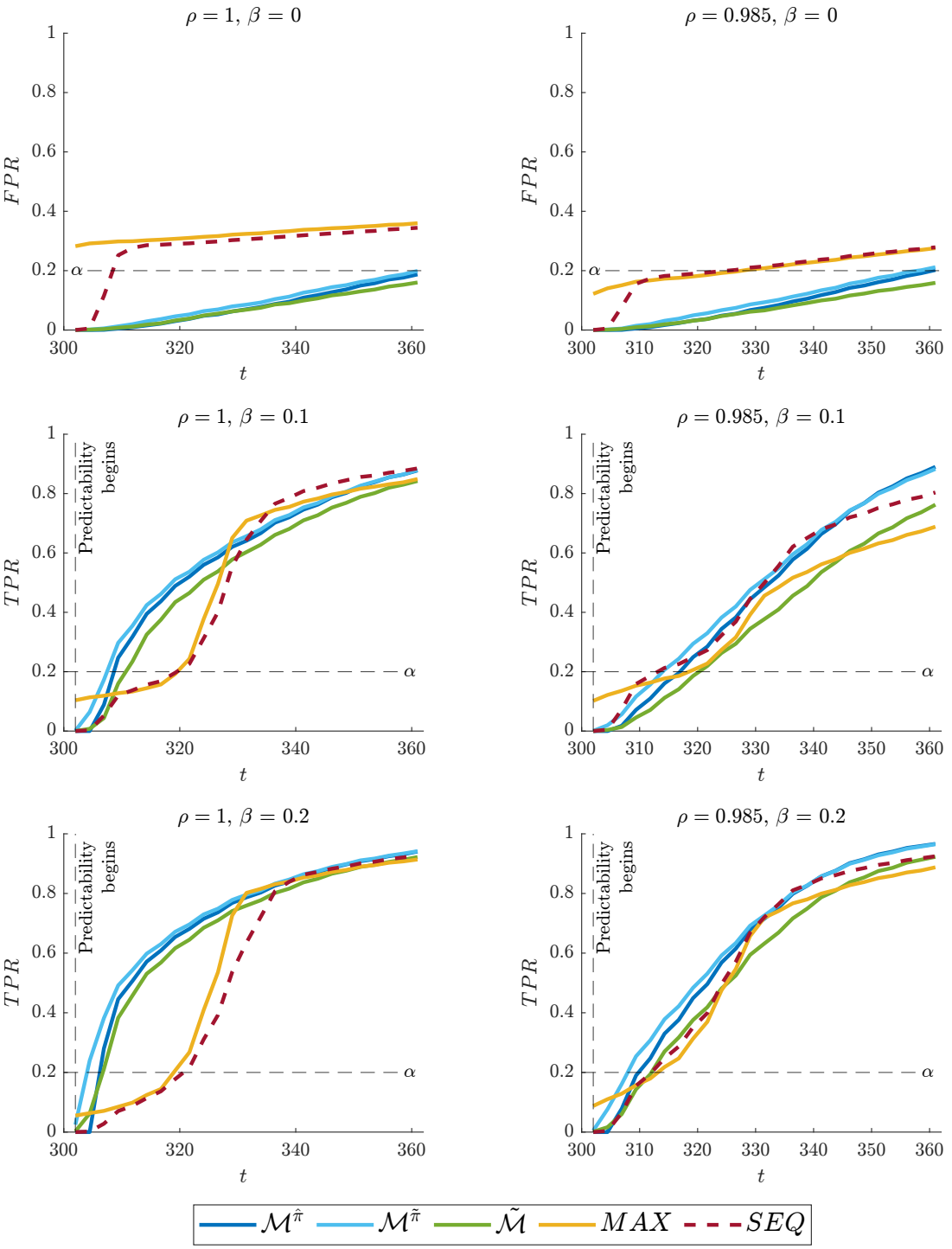

Figure S.11: FPR/TPR curves of $\mathcal{M}^{\hat{\pi}}, \mathcal{M}^{\tilde{\pi}}, \widetilde{\mathcal{M}}, MAX$ and SEQ for $\rho \in \{1, 0.985\}$ and $\beta \in \{0, 0.1, 0.2\}$ and data generated from DGP7 (predictability in the training period). Calculation of $\bar{y}_j$ re-initialized at $t = T + 1 - 12$.

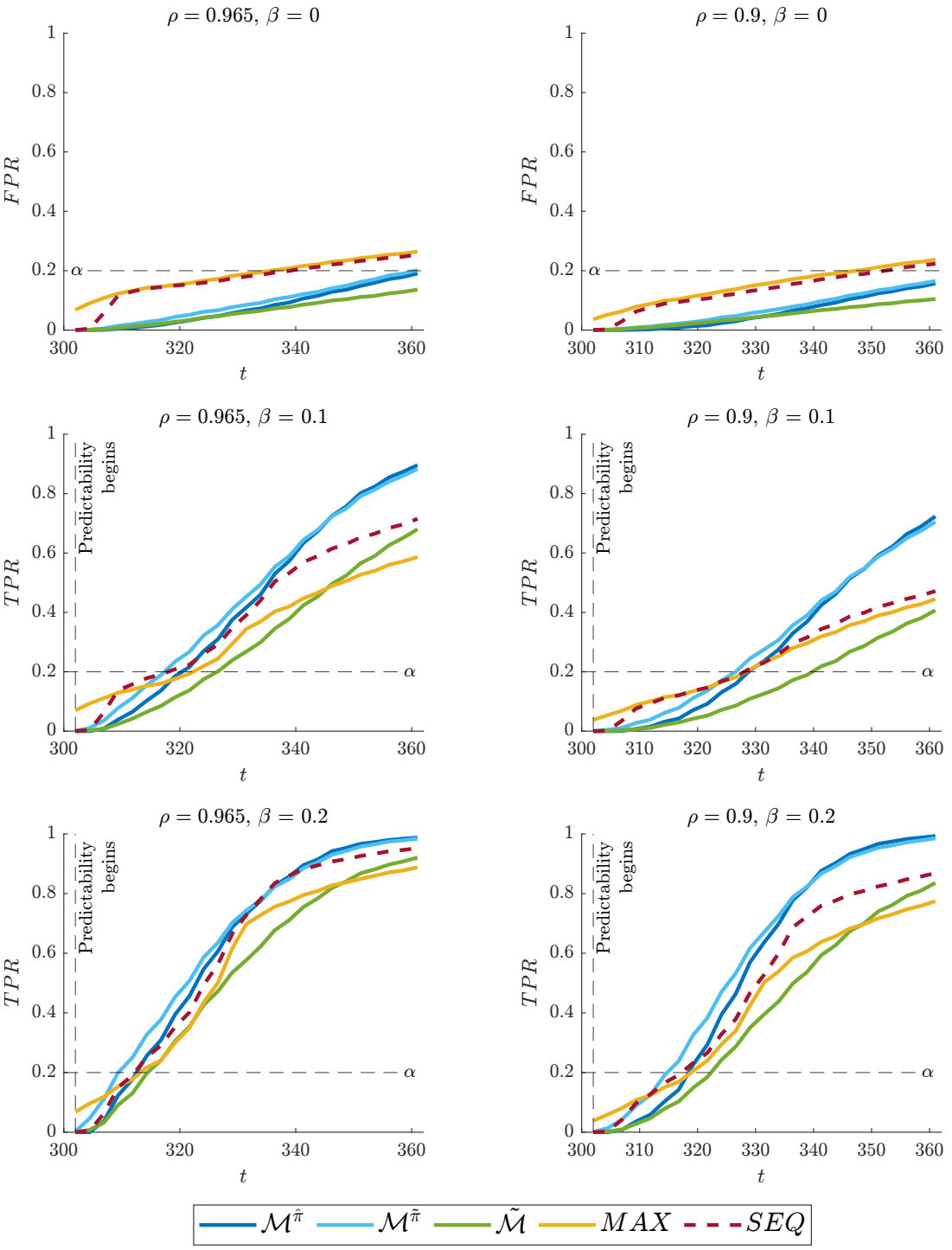

Figure S.12: FPR/TPR curves of $\mathcal{M}^{\hat{\pi}}, \mathcal{M}^{\tilde{\pi}}, \widetilde{\mathcal{M}}, MAX$ and SEQ for $\rho \in \{0.965, 0.9\}$ and $\beta \in \{0, 0.1, 0.2\}$ and data generated from DGP7 (predictability in the training period). Calculation of $\bar{y}_j$ re-initialized at $t = T + 1 - 12$.

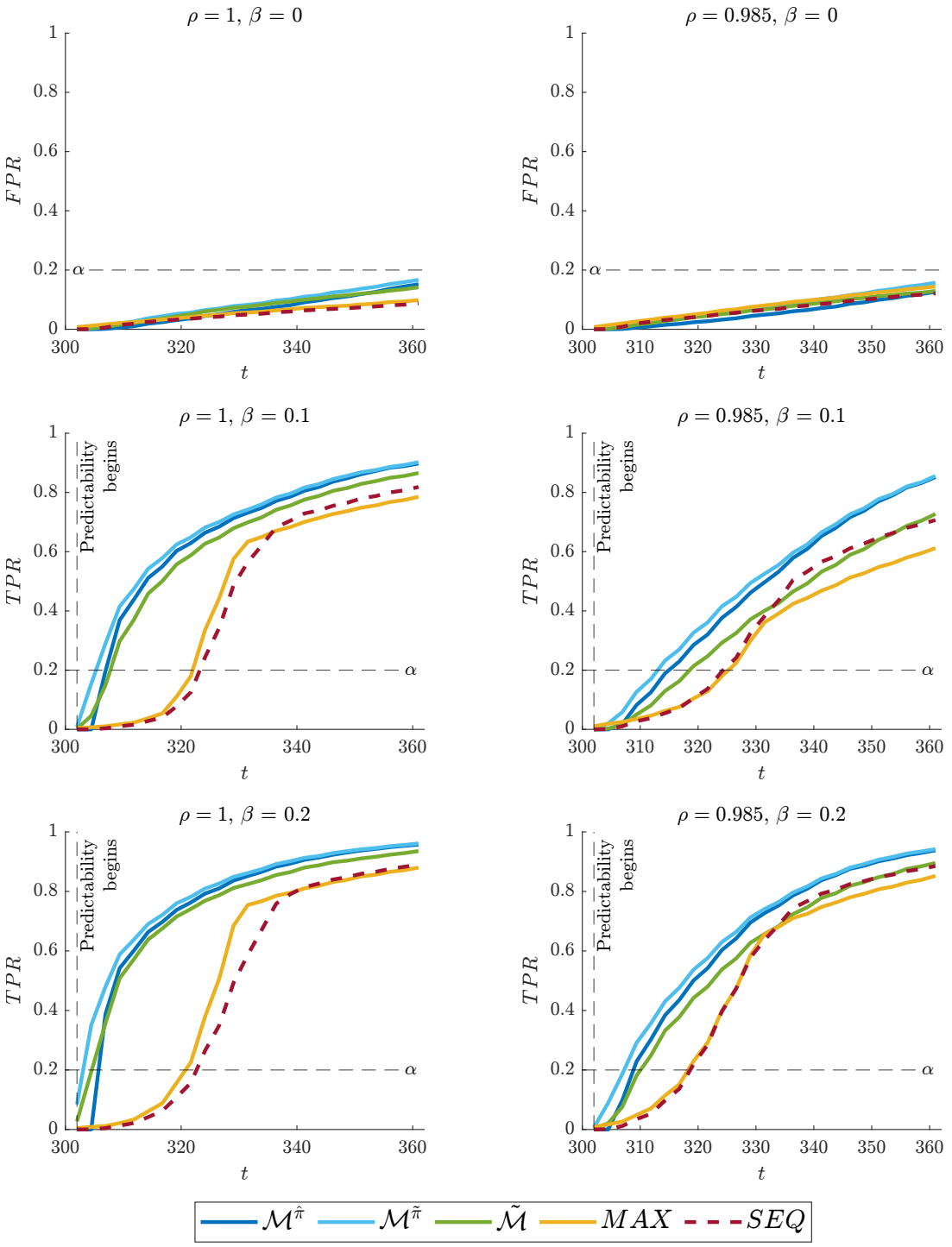

Figure S.13: FPR/TPR curves of $\mathcal{M}^{\hat{\pi}}, \mathcal{M}^{\tilde{\pi}}, \widetilde{\mathcal{M}}, MAX$ and SEQ for $\rho \in \{1, 0.985\}$ and $\beta \in \{0, 0.1, 0.2\}$ and data generated from DGP8.1 (predictability in the training period).

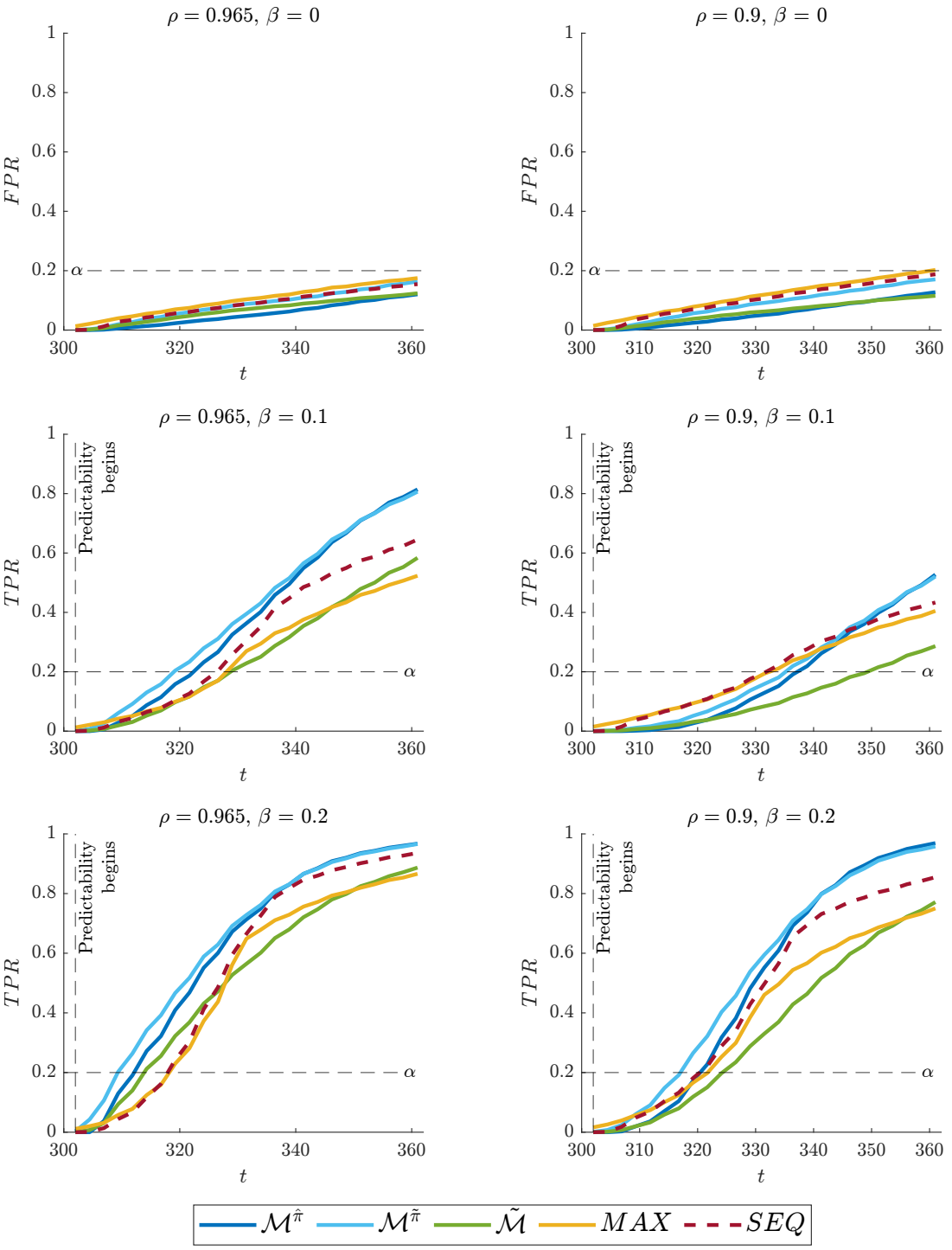

Figure S.14: FPR/TPR curves of $\mathcal{M}^{\hat{\pi}}, \mathcal{M}^{\tilde{\pi}}, \widetilde{\mathcal{M}}, MAX$ and SEQ for $\rho \in \{0.965, 0.9\}$ and $\beta \in \{0, 0.1, 0.2\}$ and data generated from DGP8.1 (predictability in the training period).

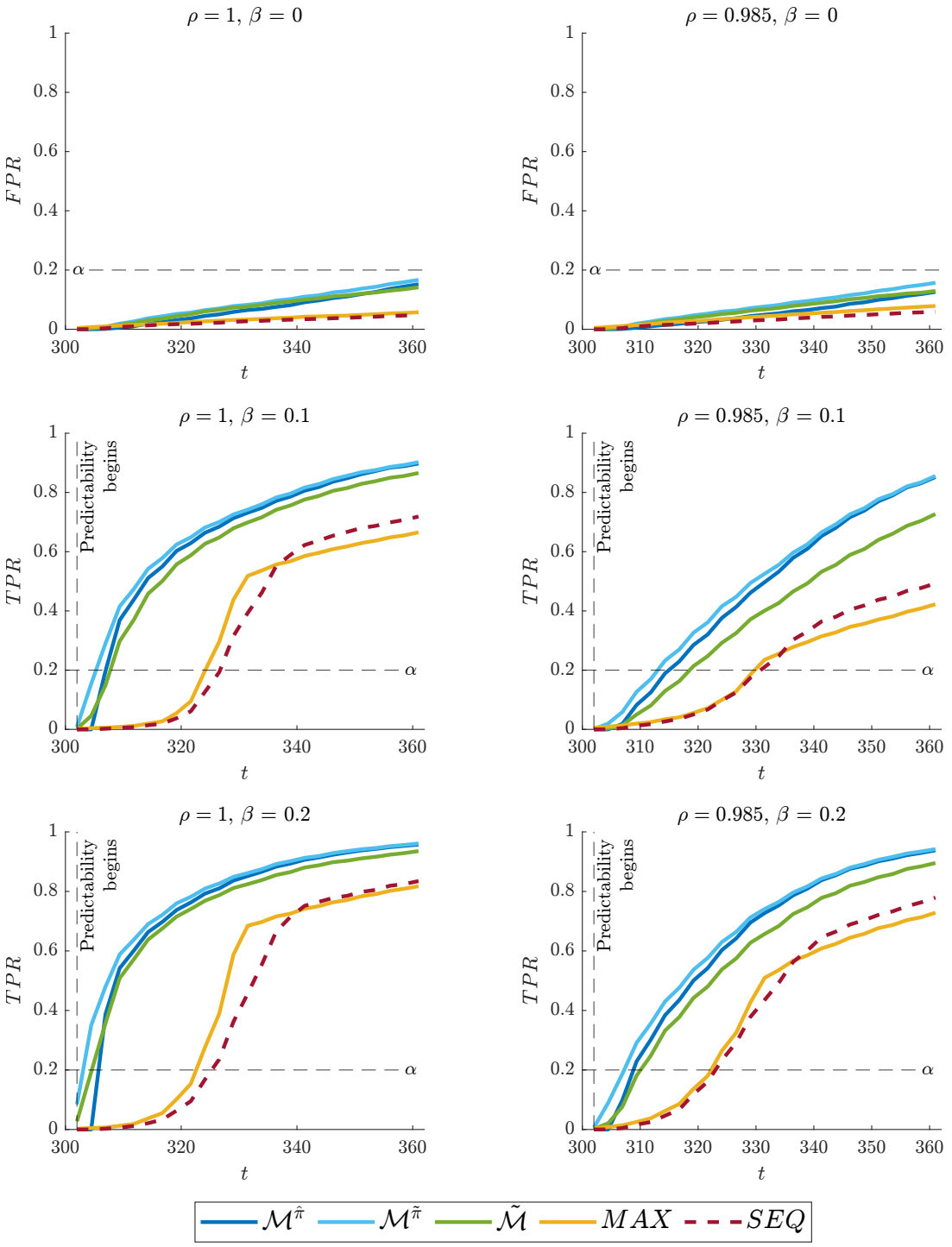

Figure S.15: FPR/TPR curves of $\mathcal{M}^{\hat{\pi}}, \mathcal{M}^{\tilde{\pi}}, \widetilde{\mathcal{M}}, MAX$ and SEQ for $\rho \in \{1, 0.985\}$ and $\beta \in \{0, 0.1, 0.2\}$ and data generated from DGP8.2 (predictability in the training period with $\beta^* = 0.3$).

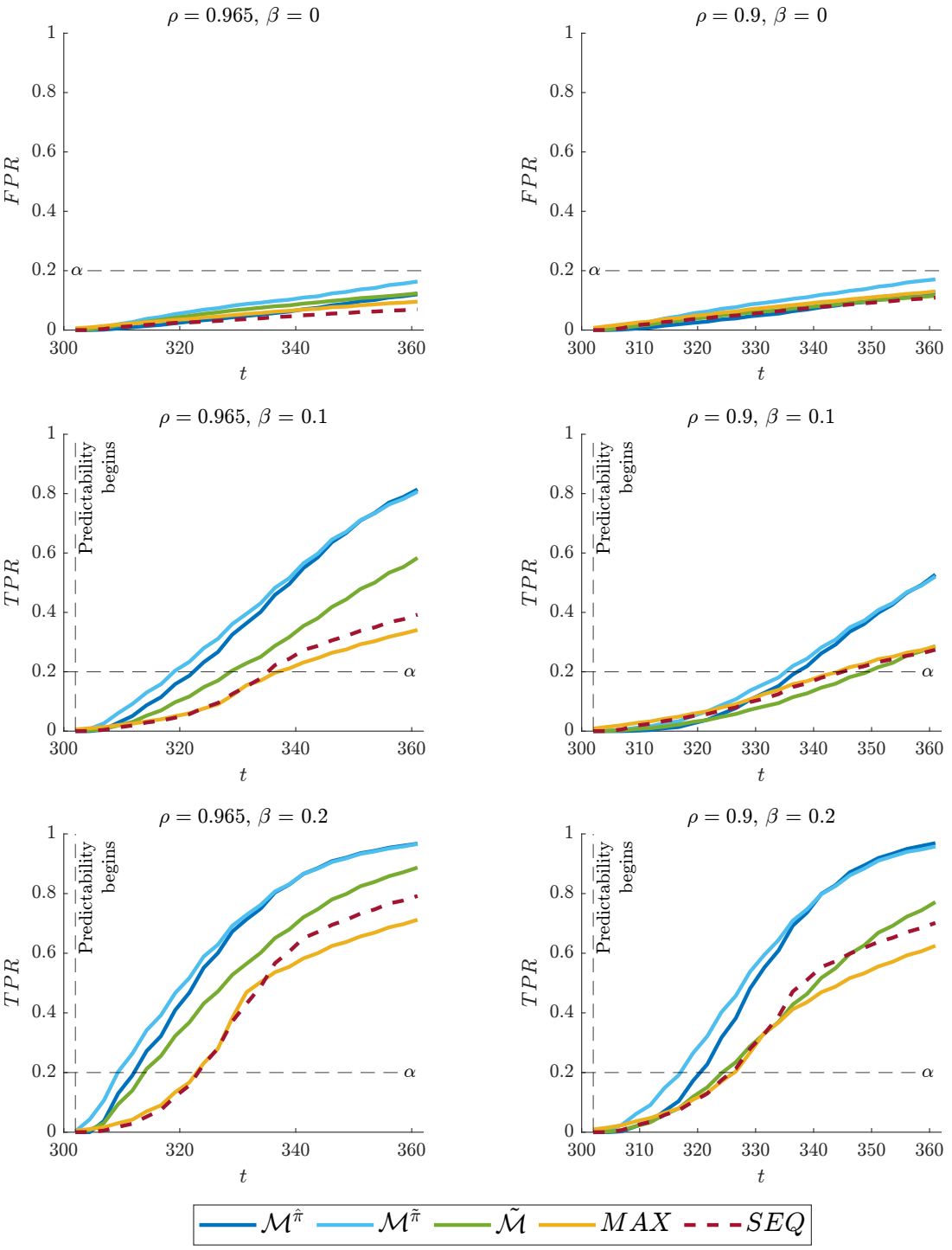

Figure S.16: FPR/TPR curves of $\mathcal{M}^{\hat{\pi}}, \mathcal{M}^{\tilde{\pi}}, \widetilde{\mathcal{M}}, MAX$ and SEQ for $\rho \in \{0.965, 0.9\}$ and $\beta \in \{0, 0.1, 0.2\}$ and data generated from DGP8.2 (predictability in the training period with $\beta^* = 0.3$).

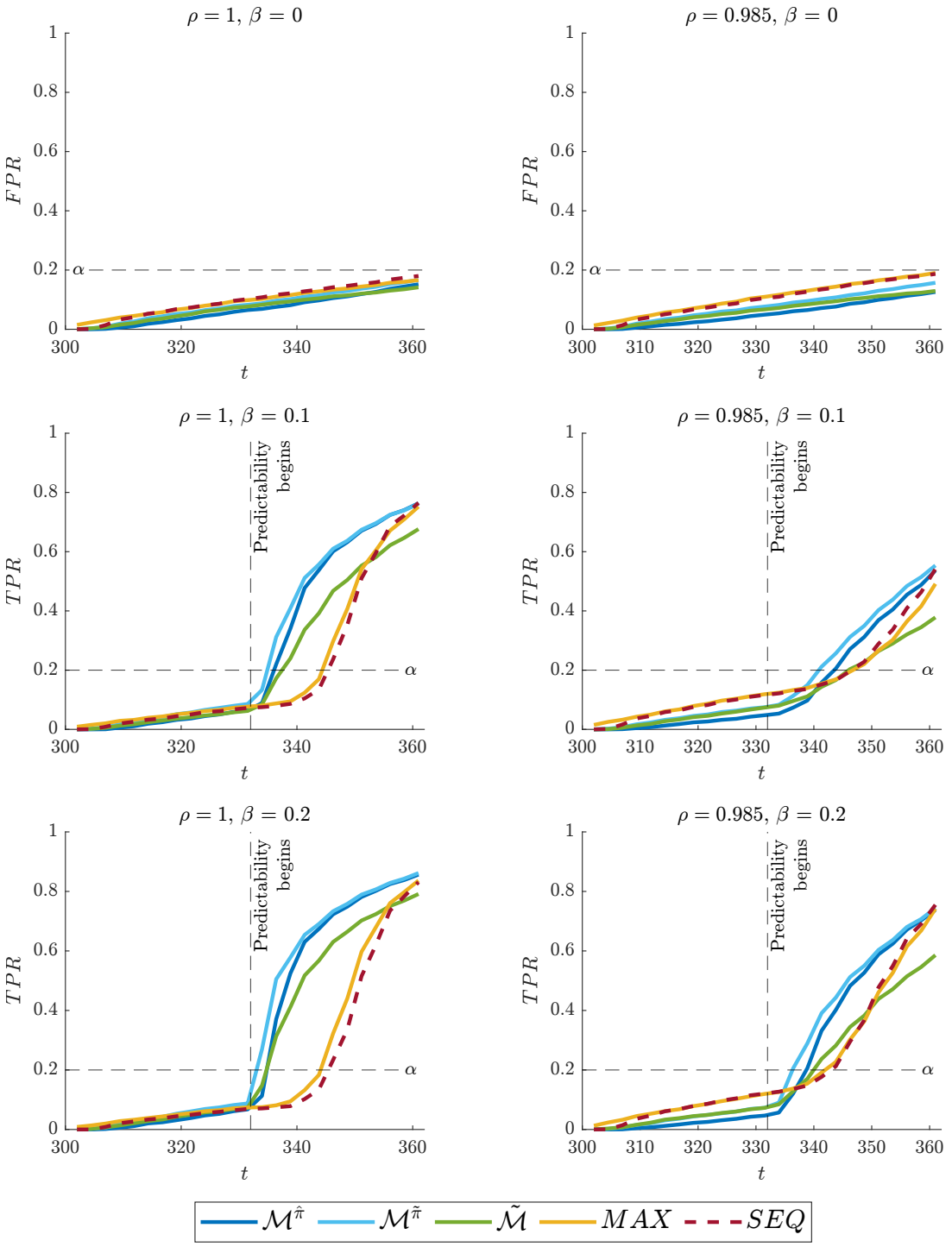

Figure S.17: FPR/TPR curves of $\mathcal{M}^{\hat{\pi}}, \mathcal{M}^{\tilde{\pi}}, \widetilde{\mathcal{M}}, MAX$ and SEQ for $\rho \in \{1, 0.985\}$ and $\beta \in \{0, 0.1, 0.2\}$ and data generated from DGP9 (start of predictability delayed).

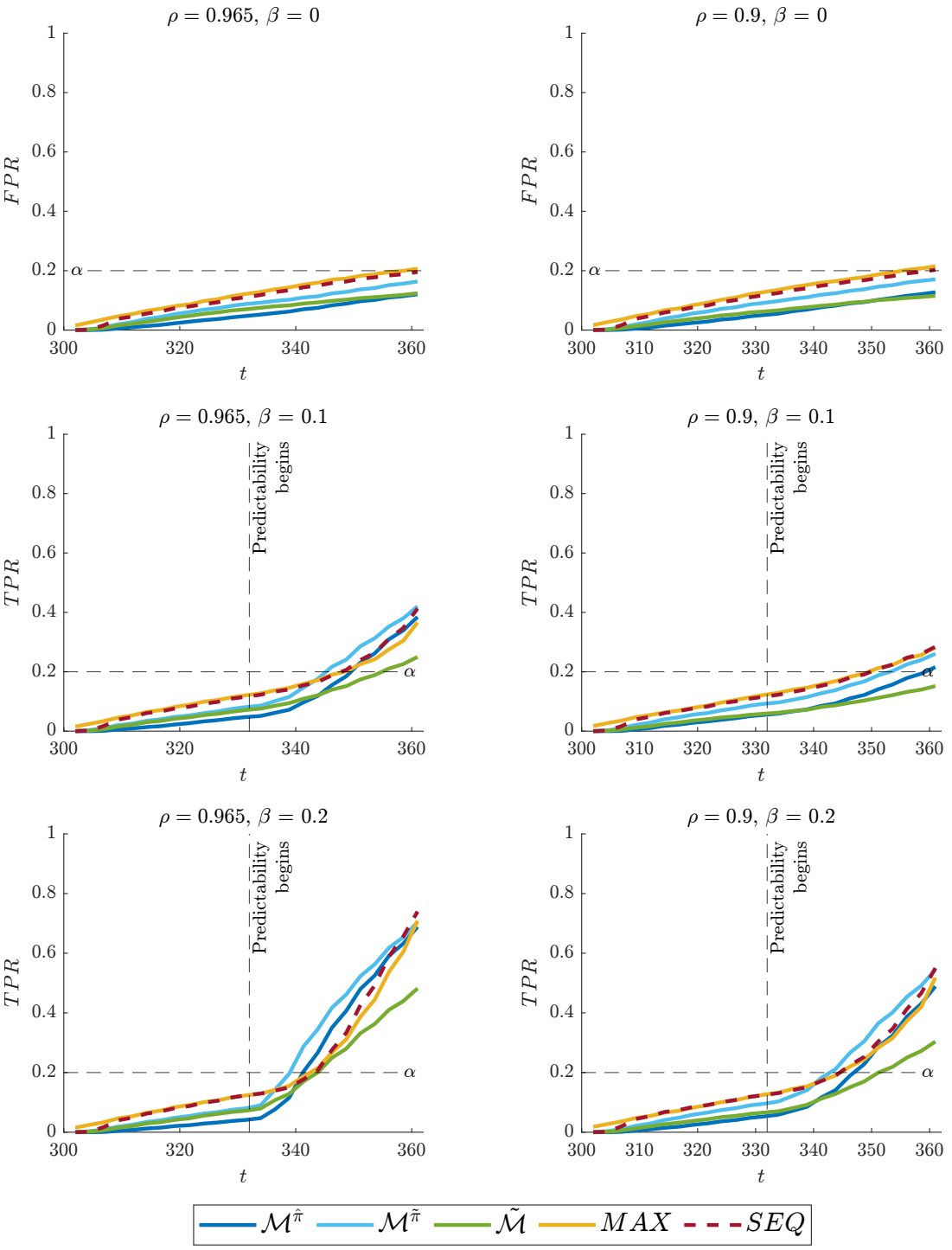

Figure S.18: FPR/TPR curves of $\mathcal{M}^{\hat{\pi}}, \mathcal{M}^{\tilde{\pi}}, \widetilde{\mathcal{M}}, MAX$ and SEQ for $\rho \in \{0.965, 0.9\}$ and $\beta \in \{0, 0.1, 0.2\}$ and data generated from DGP9 (start of predictability delayed).

S.4 Empirical Application - Supporting Material

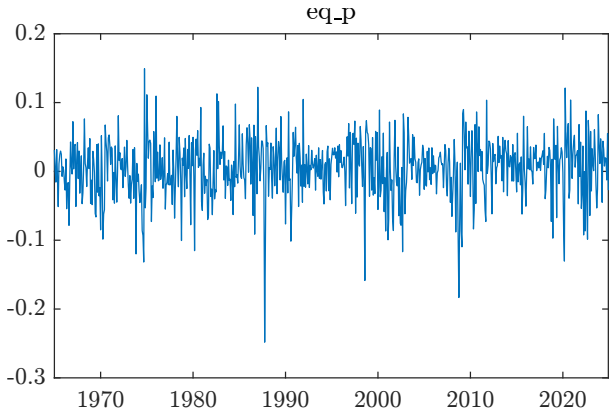

Figure S.19: Estimated equity premium time series from 01/1965 to 12/2024.

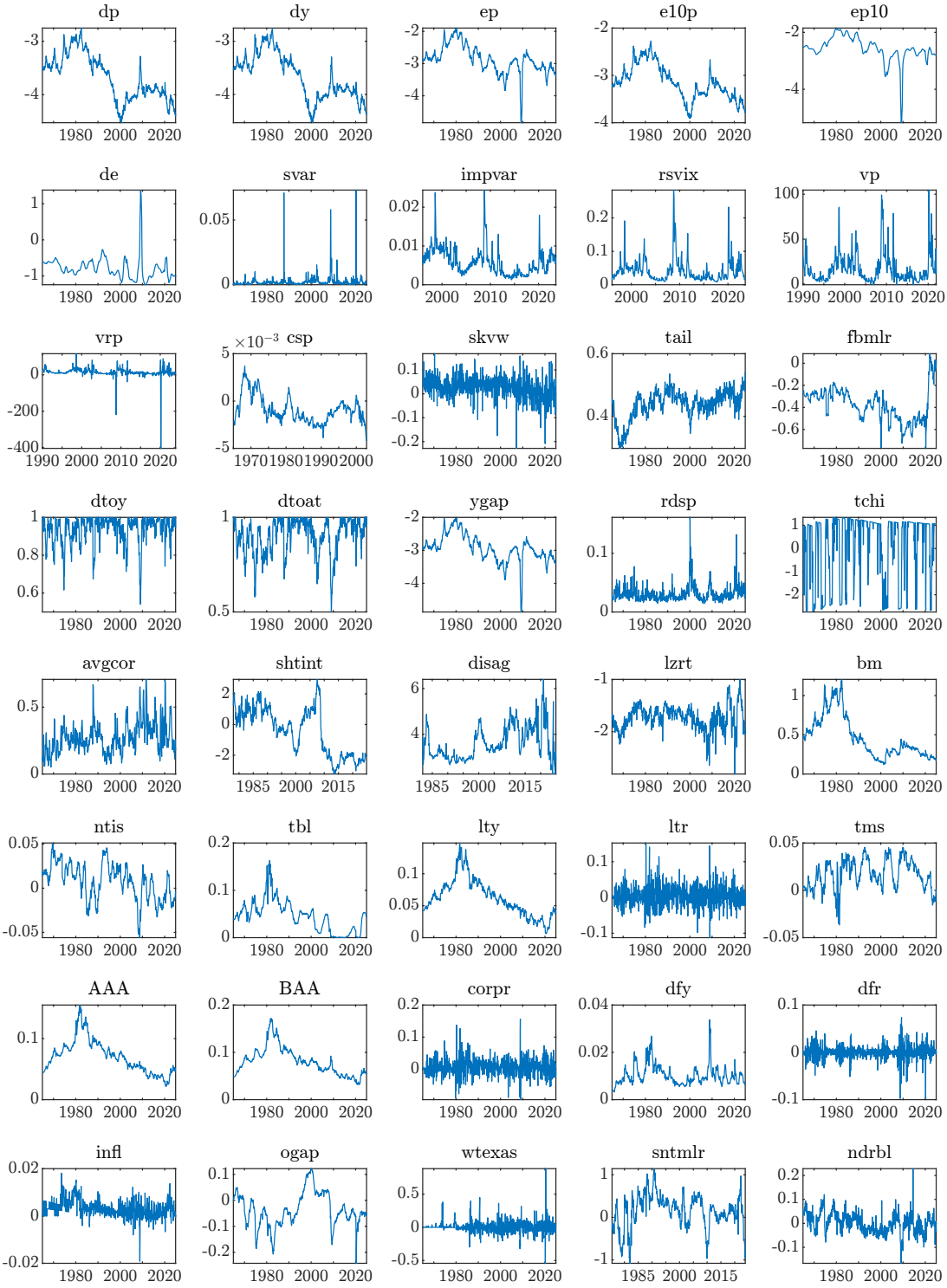

Figure S.20: Time Series Plots of the Candidate Predictors.

	Description	Source	Sample begin	Sample end
dp	Dividend-price ratio	GW2	01/1965	12/2024
dy	Dividend yield	GW2	01/1965	12/2024
ep	Earnings-price ratio	GW2	01/1965	12/2024
e10p	Earnings' ten-year moving average divided by price	GW	01/1965	12/2024
ep10	Earnings divided by prices' ten-year moving average	GW	01/1965	12/2024
de	Dividend-earnings ratio	GW2	01/1965	12/2024
svar	Stock market variance	GW	01/1965	12/2024
impvar	Implied volatility	GW2	01/1996	08/2023
rsvix	Scaled risk-neutral VIX	GW2	01/1996	08/2023
vp	VIX squared minus implied volatiltiy	GW2	01/1990	12/2021
vrp	Variance risk premium	GW2	01/1990	12/2023
csp	Cross-sectional premium	GW	01/1965	12/2002
skvw	Average stock skewness	GW2	01/1965	12/2024
tail	Tail risk from cross section	GW2	01/1965	12/2024
fbmlr	Single factor from book-to-market cross section (log returns)	GW2	01/1965	11/2024
dtoy	Nearness to Dow 52-week high	GW2	01/1965	12/2024
dtoat	Nearness to Dow all-time high	GW2	01/1965	12/2024
ygap	Stock-bond yield gap	GW2	01/1965	12/2024
rdsp	Stock return dispersion	GW2	01/1965	12/2024
tchi	14 technical indicators	GW2	01/1965	12/2024
avgcor	Average correlation of daily stock returns	GW2	01/1965	12/2024
shtint	Short stock interest	GW2	01/1978	12/2024
disag	Analyst forecast disagreement	GW2	12/1981	12/2024
lzrt	9 illiquidity measures	GW2	01/1965	12/2024
bm	Book-to-market ratio	GW	01/1965	12/2024
ntis	Net equity expansion	GW	01/1965	12/2024
tbl	Treasury bill rate	GW	01/1965	12/2024
lty	Long-term government bond yield	GW	01/1965	12/2024
ltr	Long-term government bond return	GW	01/1965	12/2024
tms	Term spread (lty - tbl)	GW2	01/1965	12/2024
AAA	Corporate bond yield; AAA-rated firms	GW	01/1965	12/2024
BAA	Corporate bond yield; BAA-rated firms	GW	01/1965	12/2024
corpr	Long-term corporate bond returns	GW	01/1965	12/2024
dfy	Default yield spread (BAA - AAA)	GW	01/1965	12/2024
dfr	Default return spread (corpr - ltr)	GW	01/1965	12/2024
infl	Inflation	GW	01/1965	12/2024
ogap	Industrial production output gap	GW2	01/1965	12/2024
wtexas	Oil price changes	GW2	01/1965	12/2024
sntmlr	Optimized investor sentiment index (log returns)	GW2	07/1975	11/2023
ndrbl	New orders/shipments to durable goods	GW2	01/1965	12/2024

Table S.1: List of candidate predictors, their sample beginning and ending dates, and their source; see Welch and Goyal (2008) (GW) and Goyal *et al.* (2024) (GW2) for more details.

	ρ	$\text{corr}(u_t, v_t)$	p_{IVX}	R^2	Adj. R^2	p^T_{IVX}
dp	0.9963	−0.9892	0.7204	0.0002	−0.0011	0.1267
dy	0.9964	−0.0238	0.6807	0.0003	−0.0011	0.1076
ep	0.9902	−0.6108	0.9733	0.0000	−0.0014	0.2213
e10p	0.9946	−0.9959	0.6367	0.0005	−0.0009	0.1172
ep10	0.9935	0.0897	0.9014	0.0000	−0.0013	0.9084
de	0.9861	−0.1019	0.7159	0.0003	−0.0011	0.7210
svar	0.3949	−0.3362	0.6540	0.0011	−0.0003	0.2663
impvar	0.7886	−0.6445	0.5182	0.0016	−0.0014	0.0024
rsvix	0.7484	−0.7413	0.6932	0.0017	−0.0013	0.0018
vp	0.7427	−0.5586	0.1220	0.0149	0.0123	0.6023
vrp	0.0757	0.0585	0.7671	0.0011	−0.0013	0.6198
csp	0.9735	−0.0053	0.0471	0.0075	0.0053	0.4460
skvw	0.1133	0.2401	0.0978	0.0063	0.0050	0.3519
tail	0.9221	−0.0234	0.0184	0.0073	0.0059	0.4724
fbmlr	0.9809	0.1467	0.2441	0.0027	0.0013	0.6259
dtoy	0.8788	0.8802	0.5516	0.0011	−0.0003	0.3107
dtoat	0.9300	0.9156	0.1611	0.0048	0.0034	0.1051
ygap	0.9895	−0.6086	0.9108	0.0000	−0.0014	0.1890
rdsp	0.5921	−0.0103	0.6865	0.0005	−0.0009	0.5457
tchi	0.9069	0.5284	0.0889	0.0057	0.0043	0.9606
avgcor	0.8906	−0.3317	0.0644	0.0082	0.0068	0.0993
shtint	0.9841	0.1232	0.0577	0.0062	0.0044	0.9583
disag	0.9193	0.1903	0.8270	0.0000	−0.0019	0.4566
lzrt	0.8538	0.1946	0.4771	0.0010	−0.0004	0.7848
bm	0.9961	−0.7502	0.8167	0.0001	−0.0013	0.4155
ntis	0.9821	−0.0414	0.3493	0.0018	0.0004	0.2122
tbl	0.9913	−0.1236	0.0760	0.0044	0.0030	0.3493
lty	0.9945	−0.1536	0.1701	0.0023	0.0009	0.2956
ltr	0.0478	0.0973	0.0554	0.0069	0.0055	0.5582
tms	0.9612	0.0193	0.2462	0.0025	0.0011	0.7786
AAA	0.9961	−0.2220	0.2189	0.0017	0.0003	0.5276
BAA	0.9966	−0.2577	0.3206	0.0010	−0.0004	0.6572
corpr	0.0975	0.2691	0.0209	0.0134	0.0120	0.6916
dfy	0.9632	−0.0549	0.4019	0.0021	0.0007	0.5360
dfr	−0.0550	0.2967	0.4826	0.0018	0.0004	0.1550
infl	0.5821	−0.0738	0.0743	0.0061	0.0047	0.1636
ogap	0.9894	−0.0131	0.0456	0.0069	0.0056	0.5347
wtexas	0.1703	0.1138	0.1093	0.0058	0.0044	0.4437
sntmlr	0.9335	0.1062	0.8044	0.0001	−0.0017	0.4568
ndrbl	0.7087	−0.0181	0.0803	0.0059	0.0045	0.9651

Table S.2: In-sample results of the following preliminary analysis steps applied to the full sample: AR(1) coefficient estimates ρ; $\text{corr}(u_t, v_t)$ estimated as the sample correlation between the residuals from the fitted bivariate predictive regression model and the residuals from the $AR(1)$ model for each candidate predictor; p-values, p_{IVX}, for the predictive regression IVX tests (using $a = 1, \eta = 0.95$ to obtain z_t, and the finite sample correction of Kostakis *et al.* (2015) with lag truncation parameter $\lfloor T^{1/3} \rfloor$); (Adj.) R^2 values based on OLS predictive regressions, *i.e.* the best linear prediction. Additionally, the final column headed p^T_{IVX} reports p-values of the IVX predictability tests of Kostakis *et al.* (2015) applied to each candidate predictor's first 60 observations (the initial training period).

	$\mathcal{KS}$	$\mathcal{R}$	$\mathcal{CVM}$	$\mathcal{AD}$
dp	-	-	-	-
dy	-	-	-	-
ep	-	-	-	-
e10p	-	-	-	-
ep10	-	-	*	-
de	-	-	*	-
svar	-	-	-	-
impvar	-	-	-	-
rsvix	-	-	-	-
vp	-	-	-	-
vrp	-	-	-	-
csp	-	-	-	-
skvw	**	-	**	**
tail	**	*	***	***
fbmlr	**	-	**	**
dtoy	-	-	-	-
dtoat	-	-	-	-
ygap	-	-	-	-
rdsp	-	-	-	-
tchi	-	-	-	-
avgcor	*	-	**	**
shtint	***	**	***	***
disag	***	**	***	***
lzrt	***	***	***	***
bm	***	***	***	***
ntis	-	-	-	-
tbl	***	**	***	**
lty	*	***	**	**
ltr	-	-	-	-
tms	**	**	**	**
AAA	-	***	*	*
BAA	-	-	-	-
corpr	-	-	-	-
dfy	-	-	-	-
dfr	**	*	*	-
infl	-	-	-	-
ogap	-	-	-	-
wtexas	-	-	**	**
sntmlr	-	**	*	*
ndrbl	-	-	*	*
eq_p	-	-	-	-

Table S.3: In-sample stationary volatility test results. The table summarizes the significance results of the four tests $\mathcal{KS}$ (Kolmogorov-Smirnov), $\mathcal{R}$ (Kuiper), $\mathcal{CVM}$ (Cramer-von Mises), and $\mathcal{AD}$ (Anderson-Darling) proposed by Cavaliere and Taylor (2008). The tests (using a lag truncation parameter of 4 in the Bartlett long run variance estimator) have been applied to AR(1) residuals for all predictors as well as the dependent variable eq_p. A dash (-) indicates no significance, an asterisk ($*$) indicates significance at the 10% level, two asterisks ($**$) at the 5% level, and three asterisks ($***$) at the 1% level.