Detecting Regimes of Predictability in the U.S. Equity Premium[*]

David I. Harvey[a], Stephen J. Leybourne[a], Robert Sollis[b] and A.M. Robert Taylor[c]

[a]School of Economics, University of Nottingham

[b]Newcastle University Business School [c]Essex Business School, University of Essex

May 15, 2018

Abstract

We investigate the stability of predictive regression models for the U.S. equity premium. A new approach for detecting regimes of temporary predictability is proposed using sequential implementations of standard (heteroskedasticity-robust) regression t-statistics for predictability applied over relatively short time periods. Critical values for each test in the sequence are provided using subsampling methods. Our primary focus is to develop a real-time monitoring procedure for the emergence of predictive regimes using tests based on end-of-sample data in the sequential procedure, although the procedure could be used for an historical analysis of predictability. Our proposed method is robust to both the degree of persistence and endogeneity of the regressors in the predictive regression and to certain forms of heteroskedasticity in the shocks. We discuss how the detection procedure can be designed such that the false positive rate is pre-set by the practitioner at the start of the monitoring period. We use our approach to investigate for the presence of regime changes in the predictability of the U.S. equity premium at the one-month horizon by traditional macroeconomic and financial variables, and by binary technical analysis indicators. Our results suggest that the one-month ahead equity premium has *temporarily* been predictable (displaying so-called 'pockets of predictability'), and that these episodes of predictability could have been detected in real-time by practitioners using our proposed methodology.

Keywords: Predictive regression; persistence; temporary predictability; subsampling; U.S. equity premium.

JEL Classification: C12, C32.

1 Introduction

Over the last three decades a large body of empirical research has been undertaken investigating stock return predictability. These methods have largely been based on linear predictive regression models and have investigated a wide array of financial and macroeconomic variables as putative

[*]Correspondence to: Robert Taylor, Essex Business School, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, United Kingdom. Email: rtaylor@essex.ac.uk.

predictors for returns. Popular variables investigated for their predictive ability for returns have included valuation ratios such as the dividend-price ratio, earnings-price ratio, book-to-market ratio, various interest rates and interest rate spreads, and macroeconomic variables including inflation and industrial production; see, for example, Fama (1981), Keim and Stambaugh (1986), Campbell (1987), Campbell and Shiller (1988a,1988b), Fama and French (1988,1989) and Fama (1990). Focusing on the in-sample predictability of U.S. stock index returns these studies find that although the statistical evidence on predictability over short horizons is relatively weak, as the forecasting horizon considered increases the evidence on predictability strengthens, and for longer horizons is strongly statistically significant. Finding that stock returns are predictable using financial ratios and macroeconomic variables does not necessarily mean that stock markets are inefficient. It follows from linearising the standard present value model that if the dividend-price ratio for a stock varies over time then it *must* forecast either the dividend growth rate or returns to some extent; see, among others, Campbell and Shiller (1988a,1988b) and Cochrane (2008). More generally, if a stock market is efficient then the expected excess return for the relevant stocks might be predictable using a variety of financial and macroeconomic variables if investors' risk premia are time-varying and correlated with the business cycle.

Although stock return predictability is consistent with orthodox financial theory, several authors have argued there are statistical reasons to suspect that the strong support for predictability obtained in earlier studies could be spurious. Nelson and Kim (1993) and Stambaugh (1999) show that high persistence predictors lead to biased coefficients in predictive regressions if the innovations driving the predictors are correlated with returns, as is known to be the case for many of the popular macroeconomic and financial variables used as predictors. Goyal and Welch (2003) show that the persistence of dividend-based valuation ratios increased significantly over the typical sample periods used in empirical studies of predictability, and argue that as a consequence out-of-sample predictions using these variables are no better than those from a no-change strategy. When estimation and inference techniques are used that take account of the high degree of persistence of the typical financial and macroeconomic variables used as predictors, the statistical evidence of short-horizon and long-horizon predictability is considerably weaker and in some cases disappears completely; see, among others, Ang and Bekaert (2007), Boudoukh, *et al.* (2007), Welch and Goyal (2008) and Breitung and Demetrescu (2015).

The vast majority of empirical studies of stock market predictability 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. A growing body of empirical evidence is also supportive of this view. For example, 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. 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. In particular, if predictability exists as a result of market inefficiency rather than because of time-varying risk premia, then rational investors will attempt to exploit its presence to earn abnormal profits. Assuming that a large-enough proportion of the total number of investors are rational, this behaviour will eventually cause the predictive power of the relevant predictor to be eliminated. 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, but it will eventually disappear; see, in particular, Paye and Timmermann (2006) and Timmermann (2008). It therefore seems reasonable to consider the possibility that the predictive relationship might change over time, so that over a long span of data one may observe some, possibly relatively short, windows of time during which predictability occurs. In such cases, standard predictability tests based on the full sample of available data will have very low power to detect these short-lived predictive episodes.

Several empirical studies have found evidence suggesting that parameter instability is indeed a feature of return prediction models. For example, Lettau and Ludvigsson (2001) find evidence of instability in the predictive ability of the dividend and earnings yield in the second half of the 1990s. Goyal and Welch (2003) and Ang and Bekaert (2007) find instability in prediction models for U.S. stock returns based on the dividend yield in the 1990s. Paye and Timmermann (2006) undertake a comprehensive analysis of prediction model instability for international stock market indices using the structural break tests developed by Bai and Perron (1998, 2003). They find statistically significant evidence of structural breaks for many of the countries considered, arguing that the “Empirical evidence of predictability is not uniform over time and is concentrated in certain periods.” *op.cit.* p.312. They find some evidence of a common break for the U.S. and U.K. in 1974-1975, and for European stock markets linked to the introduction of the European Monetary System in 1979.

A limitation of the statistical techniques used in previous research on the instability of return prediction models, such as orthodox regression t-tests, Chow tests, and the Bai-Perron tests for structural breaks, is that they are not designed for use with highly persistent, endogenous predictors. Paye and Timmermann (2006) investigate the effects of persistence and endogeneity of the regressors on the Bai-Perron tests for structural breaks using Monte Carlo simulations. Their simulations reveal that size distortions, whereby parameter change is falsely signalled when none is present, can be substantial. They also show that some of the tests lack power in this context because of the large amount of noise typically present in predictive regression models. It should also be noted that traditional regression t-tests for predictability and structural break tests are an *ex post* tool for detecting the statistical significance of regressors and structural breaks in a historical sample of data. They are less useful in monitoring for change in real-time because their repeated application in prediction models can lead to size distortions (with the probability of at least one of the tests rejecting tending to unity as the number of tests in the sequence increases) and, as a consequence, spurious evidence of in-sample predictive ability; see Inoue and Rossi (2005) for a detailed discussion of this problem in relation to t-tests.

Motivated by this, in this paper we develop new statistical techniques which we will use to investigate the stability of predictive regression models for the U.S. equity premium. As putative predictors we will consider various commonly used traditional macroeconomic and financial variables. We will also consider a range of technical analysis rules where only price or volume data is used to predict returns. In an early paper in this direction, Brock *et al.* (1992) study the ability of moving average and trading range break trading rules to predict the Dow Jones Industrial Average (DJIA) index using daily data from 1897 through to 1986, finding strong statistically significant evidence that the trading strategies generated abnormal returns that cannot be explained by serial correlation or conditional heteroskedasticity in the returns. Sullivan *et al.* (1999) analyse a longer sample of data on the DJIA, and find that the rules employed by Brock *et al.* (1992) were unable to identify profitable trading strategies for the period 1987-1996, although there was some evidence that they managed to do so prior to this period. Hudson *et al.* (1996) undertake a similar analysis to Brock *et al.* (1992) for UK stock index returns and find that although the rules examined do have predictive power, their use would not enable investors to make abnormal returns once trading transaction costs are taken into account. More recently Neely *et al.* (2014) have investigated the in-sample and out-of-sample predictive power of binary technical analysis indicators in a predictive regression-based context using monthly data. Indicators are constructed from moving-average rules, momentum rules, and on-balance volume rules. They find that the indicators have predictive power that matches or exceeds that of the traditional financial and macroeconomic variables used as predictors. They also show that combining information from technical analysis indicators and macroeconomic variables significantly improves equity risk premium forecasts versus using either type of predictor in isolation.

The testing methods we propose in this paper are designed with the aim of detecting relatively short windows of predictability arising from shifts in the parameter on the predictor variable in the predictive regression. These predictive regimes could occur at the end of the sample, the beginning of the sample, or somewhere in between. Our detection procedures are based around the sequential application of simple heteroskedasticity-robust regression t-statistics for the significance of the predictor variable calculated over a subsample of fixed length m. These statistics are then compared to critical values obtained using the subsampling-like method of Andrews (2003) and Andrews and Kim (2006). Specifically, to take the end-of-sample case to illustrate, suppose we have a sample of size $T^* + m$ and we form a predictability test statistic based on the last m observations. To obtain a critical value, one uses the *training period* $t = 1, ..., T^*$, to compute the $T^* - m + 1$ test statistics that are analogous to this statistic but calculated over the m observations that start at the jth observation (rather than the $(T^* + 1)$th observation, as for our end-of-sample statistic) for $j = 1, ..., T^* - m + 1$. The $(1 - \alpha)$ sample quantile of these statistics is the estimated significance level-α critical value for the end-of-sample predictability test. Computation of the critical value is relatively easy and p-values can also be readily obtained using this method. This methodology has distinct advantages when compared with the application of traditional regression-based tests for predictability and structural change. In particular, it is robust to the degree of persistence and endogeneity of the predictor, making it ideal for this type of application. We use t-statistics constructed using heteroskedasticity-robust standard errors and, hence, our approach is also robust to certain forms of heteroskedasticity in the model errors.

It is important to notice that an implication of the subsample critical value method being used is that these resulting one-shot tests will be able to detect general structural change in the slope parameter on the predictor variable (in that particular subsample, relative to the rest of the sample) not just a change to predictability within the given subsample. This is because a rejection will occur where the estimated slope coefficient on the predictor differs significantly between the subsample over which the one-shot test is based and the subsamples used in the critical value generation. This could occur either because the slope coefficient in the given subsample was non-zero while it was zero in the rest of the data, or vice versa, or because it took a different (non-zero) value in the given subsample from the rest of the sample. Conversely, if the slope on the predictor was fixed throughout the sample then the one-shot test would not reject (beyond its nominal significance level) regardless of what that fixed value was. However, and based on the arguments above and the work of, among others, Paye and Timmermann (2006) and Timmermann (2008), it seems reasonable to focus attention on the null model of no predictive relationship, such that structural change where it should occur is between no predictability and a short window of predictability. It is this interpretation that we will focus on in motivating and outlining our procedure. In our application to U.S. equity data we first apply standard predictability tests to the full data sets (and indeed the training periods used to obtain the estimated critical value) to check for any evidence of sustained predictability in those samples.

Our proposed approach is based on the sequential application of these one-shot subsample tests, commencing from a given start date, with a predictability regime being deemed to have occurred if a certain number of consecutive rejections (at a given marginal significance level) by these tests is observed. Where this occurs the run of rejections can be used to form estimates of the locations of the predictive regimes. When applied using end-of-sample forms of the subsample predictability tests this delivers a real-time monitoring procedure for the emergence of a regime of predictive ability of a regressor for stock returns data. Because our detection procedure is based on a sequence of subsample tests, we need to avoid the issue of spurious detections outlined in Inoue and Rossi (2005) by controlling the false positive detection rate for the detection procedure. We outline implementation rules, based on the number of consecutive rejections that need to be observed before a predictive regime is signalled by the procedure, pre-set by the practitioner at the start of the monitoring period, which control the overall false positive detection rate of the monitoring procedure.

The remainder of the paper is organised as follows. Section 2 outlines the time-varying predictive regression model which forms the basis for our analysis. Section 3 outlines some relevant motivating statistical theory. Section 4 details our proposed methodology for detecting windows of predictability and for dating any predictive regimes that are found to exist, showing how to implement a real-time detection procedure. We discuss how to control the false positive detection rate of the predictive regime detection procedure in practical applications. Section 5 reports the results from a series of Monte Carlo simulations to investigate the finite sample behaviour of our proposed predictive regime detection procedure. Section 6 presents an applied investigation into the predictability of the one month-ahead equity premium on the S&P Composite index. Section 7 concludes.

2 The Predictive Regime Model

We assume a relationship between the equity premium, y_t, and a single predictor variable[1] x_t that can be described by the following data generation process (DGP),

$$y_t = \mu_y + \sum_{j=1}^{n} \beta_j d_t(e_j, m_j) x_{t-1} + \epsilon_{y,t}, \quad t = 1, ..., T \tag{1}$$

where the (putative) predictor is generated by

$$\begin{aligned} x_t &= \mu_x + s_{x,t}, \quad t = 0, ..., T &(2)\\ s_{x,t} &= \rho s_{x,t-1} + \epsilon_{x,t}, \quad t = 1, ..., T &(3) \end{aligned}$$

with $s_{x,0}$ an $O_p(1)$ random variable and where $d_t(e_j, m_j)$ is a dummy variable defined such that $d_t(e_j, m_j)$ takes the value 1 for $m_j > 0$ consecutive values of t, ending with $t = e_j$. The innovation vector $\epsilon_t := [\epsilon_{y,t}, \epsilon_{x,t}]'$, where the notation "$x := y$" denotes that x is defined by y, is assumed to be a martingale difference sequence (mds). Specific conditions on ϵ_t are given below.

In the context of (1), if $\beta_j \neq 0$, then so we have a *predictive regime* of y_t by x_{t-1} of length m_j observations running from $t = e_j - m_j + 1$ through to $t = e_j$. The model in (1) allows for $n \geq 0$ such predictive regimes. Consistent with the discussion in the introduction and Paye and

[1] To keep our presentation as transparent as possible we will outline our procedure for the case of a single predictor. However, the approach we outline can be readily extended to the case where multiple predictors feature in (1). Here individual subsample t-statistics, of the form discussed in section 4.1, could be considered for each of the predictor variables, along with corresponding joint parameter heteroskedasticity-robust regression F-statistics. Moreover, although we focus on the case where a constant term is included in both the predictive regression model (1) and in the DGP for x_t in (2), the procedure we outline would also be valid for a more general deterministic component, such as a polynomial or broken deterministic trend, appearing in both components provided it is included in the test regression in (7) and the t-statistic, $\tau_{e,m}$, in (8) is commensurately re-defined.

Timmermann (2006) and Timmermann (2008), we have in mind scenarios where such regimes are relatively scarce and short-lived so that both the number of predictive regimes, n, and their durations, m_j, $j = 1, ..., n$, are taken to be small relative to the sample size, T. We assume $e_j < e_{j+1} - m_{j+1}$ such that the regimes where predictability holds are ordered (i.e. $d_t(e_1, m_1)$ is the earliest regime) and non-overlapping. Our proposed predictive regime detection procedure will consider the quantities e_j and m_j which delimit the start and end dates of the predictive regimes, and the number of regimes, n, to be unknown to the practitioner. Outside of these n predictive regimes the slope parameter in (1) is zero and the DGP is such that $y_t = \mu_y + \epsilon_{y,t}$ and, hence, y_t is unpredictable (in mean) due to the mds property assumed for $\epsilon_{y,t}$. Where $n = 0$ in (1), y_t is therefore unpredictable at all time periods.

The innovation vector ϵ_t is assumed to be a mds with finite fourth order moments and unconditional covariance matrix given by

$$E(\epsilon_t \epsilon_t') = \begin{bmatrix} \sigma_{y,t}^2 & r_{xy}\sigma_{y,t}\sigma_x \\ r_{xy}\sigma_{y,t}\sigma_x & \sigma_x^2 \end{bmatrix}$$

where $|r_{xy}| < 1$. This setup allows unconditional heteroskedasticity in $\epsilon_{y,t}$ while keeping the unconditional correlation between $\epsilon_{y,t}$ and $\epsilon_{x,t}$ constant at r_{xy}. Conditional heteroskedasticity, such as GARCH or stationary autoregressive stochastic volatility, is permitted in both $\epsilon_{y,t}$ and $\epsilon_{x,t}$. As regards the AR(1) process in (3), the predictive regime detection procedures we propose in this paper are valid regardless of whether $\rho = 1$ (a unit root predictor) or $|\rho| < 1$ (a stationary predictor). Moreover, ρ is also allowed to be T-dependent such as occurs, for example, in cases where the predictor is strongly persistent displaying either local or moderate deviations from a unit root; for full sample predictability tests directed at the latter, see Kostakis *et al.* (2015). The AR(1) specification is not in fact critical for our analysis, and it could be generalized to a higher order autoregressive process without affecting the validity of our proposed procedures; indeed, more generally, ϵ_t could validly be allowed to follow a stable linear process, albeit it is standard in the predictive regression literature to assume that ϵ_{yt} is serially uncorrelated.

In what follows, to facilitate our later analysis of real-time monitoring for the emergence of predictive regimes, we make a distinction between the end of the monitoring period, which we denote by $t = E$, and the notional future end of the DGP for y_t, that is $t = T$, such that $E \leq T$.

3 Background Results for Predictive Regime Detection

To motivate our approach to predictive regime detection, first suppose we have a sample of time series observations z_t, $t = 1, ..., N$ from a stationary continuous distribution. Consider the maximum value taken by z_t over $t = 1, 2, ..., N$; that is, $\max_{t \in [1,N]} z_t$. The *value* of $\max_{t \in [1,N]} z_t$ is clearly a function of the distribution of z_t. But consider the *location* at which $\max_{t \in [1,N]} z_t$

is obtained, that is $M := \arg\max_{t\in[1,N]} z_t$. Then, since all possible locations are equally likely, $\Pr(M = 1) = \Pr(M = 2) = \cdots = \Pr(M = N) = 1/N$, irrespective of the distribution of z_t. Hence, M has a discrete uniform distribution. If we standardize M as $p_M := M/N$, then, for large N, we find that $p_M \sim U(0, 1)$, where $U(0, 1)$ is the continuous uniform distribution on the interval $[0, 1]$. Hence,

$$\begin{aligned}\lim_{N\to\infty} \Pr(p_M \in [0, 1-\alpha]) &= 1-\alpha \\ \lim_{N\to\infty} \Pr(p_M \in [1-\alpha, 1]) &= \alpha.\end{aligned}$$

We can use the foregoing result in a slightly different form. Consider the maximum value of z_t in each of the two intervals $t = 1, ..., \lfloor(1-\alpha)N\rfloor$ and $t = \lfloor(1-\alpha)N\rfloor + 1, ..., N$, where $\lfloor\cdot\rfloor$ denotes the integer part of its argument; that is, $\max_{t\in[1,\lfloor(1-\alpha)N\rfloor]} z_t$ and $\max_{t\in[\lfloor(1-\alpha)N+1,N\rfloor]} z_t$, respectively, noting that only one of these can coincide with $\max_{t\in[1,N]} z_t$. Then,

$$\lim_{N\to\infty} \Pr\left(\max_{t\in[\lfloor(1-\alpha)N\rfloor+1,N]} z_t > \max_{t\in[1,\lfloor(1-\alpha)N\rfloor]} z_t\right) = \alpha. \tag{4}$$

Note that this result follows due to the large sample uniformity of the location of the maximum, and hence the probability that $\max_{t\in[1,N]} z_t$ is located in the latter interval $t \in [\lfloor(1-\alpha)N\rfloor+1, N]$, i.e. that $\max_{t\in[\lfloor(1-\alpha)N\rfloor+1,N]} z_t > \max_{t\in[1,\lfloor(1-\alpha)N\rfloor]} z_t$, is simply the limit ratio of the length of the latter interval $(N - \lfloor(1-\alpha)N\rfloor)$ to the total length of the two intervals together $(\lfloor(1-\alpha)N\rfloor + N - \lfloor(1-\alpha)N\rfloor = N)$; that is,

$$\lim_{N\to\infty} \frac{N - \lfloor(1-\alpha)N\rfloor}{N} = \alpha.$$

Now, instead of looking at the maximum value of z_t, consider instead the maximum number of *contiguous* values of z_t that exceed some threshold value, c say, where we assume that c is such that $0 < \Pr(z_t > c) < 1$. Using $1(\cdot)$ to denote the indicator function, let $R_t := 1(z_t > c)$ and define the following measure over $t = L$ to $t = U$ with $U \geq L$:

$$R(L, U) := (U - L + 1) \prod_{t=L}^{U} R_t. \tag{5}$$

Notice that when $R(L, U)$ is non-zero, its value, $U - L + 1$, represents the length of a sequence of contiguous exceedances. The maximum length of contiguous exceedances over $t = 1, ..., N$ is then $\max_{L,U\in[1,N]} R(L, U)$, which will depend on the distribution of z_t. If, however, we consider the *location* of the maximum length of contiguous exceedances, i.e. $(M_L, M_U) := \arg\max_{L,U\in[1,N]} R(L, U)$, this does not depend on the distribution of z_t as all possible locations for the pair (M_L, M_U) are equally likely. Paralleling the uniform distribution arguments leading to (4), we find that

$$\lim_{N\to\infty} \Pr\left(\max_{L,U\in[\lfloor(1-\alpha)N\rfloor+1,N]} R(L,U) > \max_{L,U\in[1,\lfloor(1-\alpha)N\rfloor]} R(L,U) \right) = \alpha. \tag{6}$$

Again, the intuition is that due to the large sample uniformity of the location of the maximum length of exceedances, the probability that $\max_{L,U\in[1,N]} R(L,U)$ is located in $L,U \in [\lfloor(1-\alpha)N\rfloor+1,N]$, i.e. that $\max_{L,U\in[\lfloor(1-\alpha)N\rfloor+1,N]} R(L,U) > \max_{L,U\in[1,\lfloor(1-\alpha)N\rfloor]} R(L,U)$, is the limit ratio of the length of the latter interval $(N-\lfloor(1-\alpha)N\rfloor)$ to the total length of the two intervals $(\lfloor(1-\alpha)N\rfloor + N - \lfloor(1-\alpha)N\rfloor = N)$, which is α.

We will subsequently employ the uniform probability arguments provided in the section to control the false positive rejection rate of the predictive regime detection procedure we propose.

4 Predictive Regime Detection

4.1 Subsample Regression t-statistics

We are interested in detecting the presence of a predictive regime for the variable y_t in real-time and propose a way of doing this utilizing subsample regression t-statistics. To that end, consider first selecting a subsample of m observations running from $t = e - m + 1$ to $t = e$, where m is chosen by the practitioner, and run the (generic) ordinary least squares [OLS] regression,

$$y_t = a + bx_{t-1} + u_t, \qquad t = e - m + 1, ..., e. \tag{7}$$

We then calculate the regression t-statistic, based around a heteroskedasticity-robust variance estimate (see White, 1982), for the significance of x_{t-1} in (7); that is,

$$\tau_{e,m} := \frac{\hat{b}}{\sqrt{\hat{V}(\hat{b})}} \tag{8}$$

where

$$\begin{aligned} \hat{b} &:= \frac{\sum_{t=e-m+1}^{e}(x_{t-1}-\bar{x}_{-1})(y_t-\bar{y})}{\sum_{t=e-m+1}^{e}(x_{t-1}-\bar{x}_{-1})^2}, \quad \hat{V}(\hat{b}) := \frac{\sum_{t=e-m+1}^{e}(x_{t-1}-\bar{x}_{-1})^2\hat{u}_t^2}{\{\sum_{t=e-m+1}^{e}(x_{t-1}-\bar{x}_{-1})^2\}^2} \\ \hat{u}_t &:= (y_t-\bar{y}) - \hat{b}(x_{t-1}-\bar{x}_{-1}) \\ \bar{y} &:= m^{-1}\sum_{t=e-m+1}^{e} y_t, \quad \bar{x}_{-1} := m^{-1}\sum_{t=e-m+1}^{e} x_{t-1}. \end{aligned}$$

Provided an appropriate critical value is available to the practitioner, a test for a predictive regime holding between y_t and x_{t-1} for the given subsample $t = e-m+1, ..., e$ can then be based on $\tau_{e,m}$. As a particular example, suppose we have data available for $t = 1, ..., T^* + m$; a test for the presence of a predictive regime in the last m available sample observations would therefore be based on the statistic $\tau_{T^*+m,m}$. Standard regime detection tests, such as those outlined in Paye and Timmermann (2006) use asymptotic distribution theory to approximate the test's critical value but this approximation is based on the assumption that the sample window m used in constructing the test statistic is a positive fraction of the sample size, T. This assumption is clearly not consistent with our aim which is to detect predictive regimes of short duration. Moreover, even if we were to assume m to be a function of the sample size T, the resulting limiting distribution of $\tau_{e,m}$ will depend on nuisance parameters present in the DGP in (1)-(3) including the degree of persistence of the predictor variable, x_t, and the correlation, r_{xy}, between $\epsilon_{y,t}$ and $\epsilon_{x,t}$. Consequently, without knowledge of these nuisance parameters, valid asymptotic critical values for the test could not be obtained in any case.

An alternative approach that we will use in this paper is based on the subsampling method for obtaining critical values developed in Andrews (2003) and Andrews and Kim (2006). Here the asymptotic justification for the procedure is based on the scenario where $T \to \infty$ but crucially, and as in our setting, the sample window, m, remains finite. Applied to the $\tau_{e,m}$ statistic, the Andrews-type approach, which we will outline in detail in section 4.2 below, involves comparing $\tau_{e,m}$ with critical values obtained by subsampling from some subset (this data subset is referred to as a *training period* in the end-of-sample testing example based on $\tau_{T^*+m,m}$ given above) of those sample data not used in calculating $\tau_{e,m}$. This approach delivers tests which, by design, are robust to the nuisance parameters in (1)-(3). Heuristically, this holds because the estimated critical values are obtained from an empirical distribution function that, for large T, has the same functional dependence on those nuisance parameters as does the distribution function of $\tau_{e,m}$ itself. Importantly, if no predictability holds throughout the data subset used to estimate these critical values, then the resulting test based on $\tau_{e,m}$ and these estimated critical values is a valid test for the null hypothesis of no predictability against the alternative of predictability, in the context of the subsample of m observations running from $t = e - m + 1$ to $t = e$.

The discussion above relates to a one-shot predictability test based on $\tau_{e,m}$. However, our goal is to develop a real-time monitoring procedure for the emergence of an end-of-sample predictive regime. To that end, we will construct a sequence of $\tau_{e,m}$ statistics, of the form given in (8), calculated for each possible end-of-subsample date $e = T^* + m, ..., E$, recalling that E denotes the end of the monitoring period, a parameter set by the practitioner. The predictive regime detection procedure we propose will be based on a subset of the resulting sequence of statistics.

4.2 The Detection Procedure

The first step of our predictive regime detection approach is to determine a critical value to use in connection with the sequence of $\tau_{e,m}$, $e = T^* + m, ..., E$, statistics. These will be estimated from an initial training period in which it will be assumed that no predictive regime occurs;

further discussion relating to where this assumption might be violated is given at the end of section 4.3. To that end, suppose there is a T^* such that $T^* := \lfloor \lambda T \rfloor$ for some $\lambda \in (0,1)$ where it holds that $T^* < e_1 - m_1 + 1$. We consider $t = 1, ..., T^*$ as the training period where predictive regimes are assumed absent.

As discussed in section 4.1, using the sequence of $\tau_{e,m}$ statistics that make use of data within this training period, i.e. $\tau_{e,m}$ for $e = m+1, ..., T^*$, we will follow the approach of Andrews (2003) and Andrews and Kim (2006) and calculate an empirical critical value for a significance level π, say. We denote this empirical critical value by cv_π. Assuming we want to perform upper one-tail testing (so that a rejection indicates a significant positive predictive relationship between y_t and x_{t-1}), then cv_π is defined such that $cv_\pi := \tau_{(\lfloor (1-\pi)(T^*-m) \rfloor)}$ where $\tau_{(j)}$, $j = 1, ..., T^* - m$ are the ascending order statistics of $\tau_{e,m}$, $e = m+1, ..., T^*$ (that is, $\tau_{(j+1)} > \tau_{(j)}$ for $j = 1, ..., T^* - m - 1$).[2] Under the conditions placed on (1), it follows from Andrews (2003) and Andrews and Kim (2006) that cv_π is a consistent estimate for the true π significance level critical value as $T \to \infty$.

Next, we start our monitoring period by calculating the first statistic $\tau_{e,m}$ which does not utilize any of the training period data; that is, $\tau_{T^*+m,m}$ (which uses data from $t = T^* + 1$ to $t = T^* + m$), and compare this with the training period critical value cv_π. Then, we move forwards one period, calculating $\tau_{T^*+m+1,m}$, again comparing the statistic with cv_π. We proceed sequentially in this manner, comparing $\tau_{e,m}$, $e = T^* + m, T^* + m + 1, ...$, with cv_π as we move forwards in time, and using $R_e = 1(\tau_{e,m} > cv_\pi)$ to record whether or not each test statistic in the sequences exceeds the critical value. Note that when $\beta_j = 0$ for all j in (1), then, for large T^*, the $\tau_{e,m}$ will be stationary across $e = \{m+1, ..., T^*\} \cup \{T^* + m, T^* + m + 1, ...\}$ (the m length gap can effectively be ignored), so the critical value cv_π calculated from $\tau_{e,m}$ for $e = m+1, ..., T^*$ is appropriate for $\tau_{e,m}$ for $e = T^* + m, T^* + m + 1, ...$ also.

To reliably detect a predictive regime, such that the false positive detection rate [FPR] of the procedure can be properly controlled, we do not simply take a single exceedance $R_e = 1$ to be sufficient evidence for an identified predictive regime. Rather we consider identifying a predictive regime when there is a contiguous sequence of exceedances that exceed some minimum length requirement. Specifically, for $U \geq L$, let

$$R(L,U) := (U - L + 1) \prod_{e=L}^{U} R_e$$

so that, when $R(L,U)$ is non-zero, it gives the length of contiguous exceedances between $e = L$ and $e = U$; cf. (5). We then determine that a predictive regime is present when $R(L,U) > m^*$ for some choice of $m^* > 1$.

Notice that, as a result, the first time period at which it would be possible to detect a predictive regime is $t = T^* + m + m^*$, because this is the first occasion where $R(L,U)$ can exceed

[2] The same general approach can also be used for lower one-tail or two-tail testing.

m^* (here $R(T^*+m, T^*+m+m^*) = (m^*+1)\prod_{e=T^*+m}^{T^*+m+m^*} R_e$). We then continue to apply this detection procedure as we move forwards in time, up to our end-of-monitoring date, $t = E$. Clearly, to be able to detect a predictive regime, it must be true that $E \geq T^* + m + m^*$, and, for a sufficiently large E, it is clearly possible for our procedure to detect multiple predictive regimes within the time span $T^*+m, ..., E$. In the next subsection, based on the arguments of section 3, we suggest a data-based method to choose m^* and analyse the impact of E on the overall FPR of the procedure.

4.3 Choice of m^* and the False Positive Detection Rate

We now consider issues pertaining to the FPR of our proposed procedure; that is, the probability of incorrectly identifying at least one predictive regime when in fact none exists and where the monitoring has been run out to E. Adapting the result of (6) to a statement regarding the location of the longest contiguous sequence of exceedances R_e, we can write

$$\lim_{T^*, E \to \infty} \Pr\left(\max_{L,U \in [T^*+m, E]} R(L,U) > \max_{L,U \in [m+1, T^*]} R(L,U) \right) = \alpha \tag{9}$$

provided, in the notation of (6), that

$$\begin{aligned} \lfloor (1-\alpha)N \rfloor &= T^* - m & (10) \\ N - \lfloor (1-\alpha)N \rfloor &= E - (T^*+m) + 1 & (11) \end{aligned}$$

with, trivially, $N = T^* - m + E - (T^* + m) + 1 = E - 2m + 1$. We note that (9) is a statement regarding the limiting probability of the longest contiguous sequence of exceedances lying in the monitoring period as opposed to the training period. For large N (and, hence, large T^* and E), (10)-(11) imply that $\alpha/(1-\alpha) = (E-(T^*+m)+1)/(T^*-m)$. This can be solved to give a finite sample approximation for α as follows

$$\alpha = \frac{E - T^* - m + 1}{E - 2m + 1}. \tag{12}$$

The practical implication of this result is that if we set m^* to be the longest contiguous sequence of exceedances in the training period; that is, we set

$$m^* = \widehat{m}^* := \max_{L,U \in [m+1, T^*]} R(L,U) \tag{13}$$

then, in large samples, the FPR of the resulting monitoring procedure run up to E is given by α. Here, α is a monotonically increasing function of E since

$$\frac{\partial \alpha}{\partial E} = \frac{T^* - m}{(E - 2m + 1)^2} > 0.$$

Hence, other things being equal, the longer the monitoring period, the greater the likelihood of spuriously finding a predictive regime. For any given monitoring horizon E, the result in (12) delivers an approximation to the empirical FPR that would be obtained in practice when setting $m^* = \widehat{m}^*$. By way of illustration, suppose we set $T^* = 400$ and $m = 30$, Figure 1 shows this approximation to the FPR as a function of E.

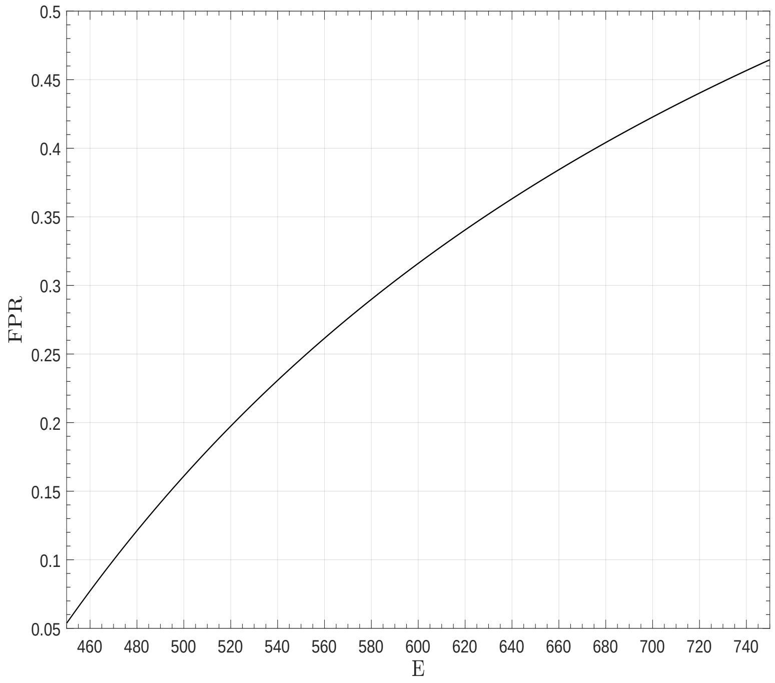


Figure 1. FPR as a function of E

If, for example, we wish to monitor out to $E = 680$, then the FPR will be about 0.40. We can also rearrange (12) as

$$E = \frac{T^* + m - 1 - \alpha(2m - 1)}{1 - \alpha} \tag{14}$$

which is useful if we wish to know the maximum monitoring horizon E such that the FPR is controlled at α. For the current illustration, if we wish to control this rate to $\alpha = 0.20$, then (14) shows us that E should be no more than about 520 (which is also apparent from Figure 1).

Notice that none of the foregoing material in this subsection appears to relate directly to the choice of significance level π at which the individual $\tau_{e,m}$ tests calculated in the monitoring sequence are run. In fact, the dependence is implicit because π influences the lengths of the contiguous rejections: the larger is π, the smaller is cv_π and the longer we would expect the sequences of contiguous rejections to be. This, in turn, will influence the value that $\widehat{m}^*$ in (13) takes.

An implication of the previous paragraph is that any sensible threshold value could in principle be used in place of the critical value estimated from the training period. A benefit of the estimated critical value approach is that where the training period contains no predictive regimes each individual test in our monitoring sequence can be interpreted marginally as a test for predictability in that particular subsample. Moreover, suppose, in contradistinction to our maintained assumption so far, that one or more short duration predictive regimes in (1) are present within the chosen training period. Although the large (in T) sample properties of the estimated critical value would be unaffected by this, for a given finite length training period, if (for example) positive predictability regimes existed in the training period then so we would expect both cv_{π} and $\widehat{m}^*$ to increase relative to the case where no predictability is present in the training period. We might therefore anticipate some reduction in the ability of the procedure to detect genuine predictive regimes present in the monitoring period due to the increase in $\widehat{m}^*$. We will explore the impact on our proposed procedure of a predictive regime holding in the training period as part of our Monte Carlo simulation study in section 5.

Our discussion in this section has assumed for simplicity that there is no separation between the data period used for the training period and the data used for monitoring, with the former spanning $t = 1, ..., T^*$ and the latter starting at $t = T^* + 1$. More generally, the last time period included in the training sample could be $T^* - k$ for some $k > 0$, thereby allowing for a separation between the training period and the start of the monitoring period. This might be relevant in cases where a predictability regime was thought to have occurred towards the end of the training period, so that the training period could be redefined to exclude this regime. In this case the right hand side of (10) becomes $T^* - m - k$ (while (11) remains unchanged), so the expressions for α and E in (12) and (14) become, respectively,

$$\alpha = \frac{E - T^* - m + 1}{E - 2m + 1 - k} \qquad \text{and} \qquad E = \frac{T^* + m - 1 - \alpha(2m - 1 + k)}{1 - \alpha}. \tag{15}$$

Although not consistent with the interpretation we are placing on the DGP in (1), it is also possible in practice that the training period could potentially contain longer periods of predictability, including the case where predictability holds throughout the training period. In the latter case, and as discussed in the Introduction, a test based on $\tau_{e,m}$ and the estimated critical values from the training period is a test for structural change in the slope parameter of the predictive regression in the subsample $t = e - m + 1, ..., e$ relative to its value in the training period (this is because using the estimated critical values from the training period acts to re-centre the t-statistic, $\tau_{e,m}$, in (7) about the value of the slope parameter in the training period, rather than about zero). For example, a rejection based on an upper tail critical value estimated from the training period would indicate a statistically significant increase in the magnitude of the slope parameter on x_{t-1} (and, hence, in the strength of the predictability of y_t by x_{t-1}) in the subsample $t = e - m + 1, ..., e$, *vis-à-vis* the value of the slope parameter in the training period. In practical applications, we recommend applying standard full-sample predictability tests to the training period to investigate whether the assumption of no predictability holds in the training period and this will be done in the empirical data analysis undertaken in section 6.

4.4 Dating of Predictive Regimes

Our proposed procedure allows detection of at least one predictive regime before the end-of-monitoring date E. To consider the possible dating of predictive regime(s) for a generic choice of m^*, which could be $\widehat{m}^*$ in (13), let D denote an $E \times 1$ vector of zeros. Then, for $e = T^* + m + m^*, ..., E$, if $\prod_{k=e-m^*}^{e} R_k = 1$, set $D_{e-m^*}, ..., D_e$ to 1. That is, for all end-of-window dates e that form part of a contiguous run of at least m^*+1 exceedances R_e, we set the eth element of D to one. Now suppose that D has h consecutive 1s in positions $e = j, ..., j+h-1$. Since the first exceedance is represented by R_j, which is based on data over the period $j - m + 1, ..., j$, we might consider $j - m + 1$ to represent a feasible start date for the predictive regime. With R_{j+h-1} representing the final exceedance, and this being based on data over the period $j - m + h, ..., j + h - 1$, we would consider $j + h - 1$ to represent a feasible end date for the predictive regime. By this categorisation, then, a given predictive regime covers the contiguous set of dates $j - m + 1, ...,$ $j + h - 1$.

In some sense, this set of dates is liberal, or *weak*, since it is possible that the predictive regime started after $j - m + 1$ and ended before $j + h - 1$; for example, only the later data used in R_j may be responsible for triggering that exceedance, and only the earlier data used in R_{j+h-1} responsible for triggering that exceedance. We might therefore consider an alternative dating approach where the predictive regime is characterised by the subset of dates for which every time the date is present in the test data, an exceedance is obtained. This subset, which we refer to as *strong*, is the contiguous set of dates $j, ..., j-m+h$; note that if $h \leq m-1$, the strong set will be empty. In situations where more than one predictive regime has been detected, it is possible that weak dates associated with consecutive regimes can overlap, though this possibility cannot arise for the strong dates.

5 Finite Sample Properties of the Monitoring Procedure

In this section Monte Carlo simulations are used to study the finite sample properties of our real-time predictive regime monitoring procedure, employing the data-based procedure for controlling the FPR defined in (13). In total we present the results from eight sets of simulation experiments based on the DGP given by (1)-(3). In all of the experiments we set $\mu_y = \mu_x = 0$ (without loss of generality) in the simulation DGP, and use negatively correlated standard normal error terms $\epsilon_{y,t} \sim N(0,1)$, $\epsilon_{x,t} \sim N(0,1)$, with $r_{x,y} = -0.90$.[3] All of the simulation experiments and the

[3] In predictive regression models for the equity premium employing valuation ratios as predictors (e.g. the dividend-price ratio, earnings-price ratio) the relevant error terms are strongly negatively correlated, hence our choice of $r_{x,y} = -0.90$.

empirical application in section 6 employ the upper-tailed version of our procedure.[4] In each simulation experiment the total sample size T and the date when monitoring starts $(T^* + m)$ are the same as in the empirical application $T = 493$ and $T^* + m = 302$, and in all cases $m = 30$.[5] All of the experiments are undertaken using MATLAB, employing the Mersenne Twister random number generator function and 5,000 replications.

The first set of experiments study the power of our procedure to detect a single predictive regime as a function of $\beta_1 = \{0.10, 0.20, 0.30, 0.40, 0.50, 1.00\}$ for $\rho = \{0.965, 0.975, 0.985, 0.995\}$, setting $\pi = 0.10$.[6] When $\beta_1 = 0$ (so that $n = 0$ and, hence, there are no predictive regimes in the data) the detection frequency obtained from the simulations is equivalent to an empirical FPR and we also report simulation results for this case. In the first set of experiments we assume a short monitoring period that ends at $E = 328$, which given the values used for T^* and m is consistent with $\alpha = 0.10$ (this can be verified using (12)). Therefore when $\beta_1 = 0$, because we are using the data-based $\widehat{m}^*$ for controlling the FPR to α the empirical FPR obtained should be approximaely equal to 0.10. If a predictive regime does occur during the monitoring period, then the power of our procedure to detect its presence will depend not only on how long the relevant predictive regime continues for (m_1) and its strength (measured by the magnitude of β_1), but also on when the predictive regime occurs relative to the start of monitoring. To investigate this issue in more detail, separate results are computed for five different predictive regime start dates: (a) $t = 287$ (15 observations before the start of monitoring), (b) $t = 297$ (5 observations before the start of monitoring), (c) $t = 302$ (at the same time as the start of monitoring), (d) $t = 307$ (5 observations after the start of monitoring), (e) $t = 317$ (15 observations after the start of monitoring). In each case the length of the predictive regime in the DGP is set to $m_1 = 30$.[7]

In empirical applications, whilst there might be a particular reason for favouring a short monitoring period, for predictive regimes that start towards the end of a short monitoring period the power of our procedure to detect their presence could be significantly improved if we monitor for a longer period of time. To investigate this issue in more detail, in the second set of experiments we repeat the first set of experiments employing the same simulation DGP and

[4] For the majority of the macroeconomic and financial variables and for all of the technical analysis indicators used in the empirical application in section 6 financial theory suggests a positive relationship with the equity premium. For those of the macroeconomic and financial variables where financial theory suggests a negative relationship with the equity premium (e.g. interest rates) we use $-x_{t-1}$ rather than x_{t-1} when testing for a predictive regime so that an upper-tailed test is applicable. This is consistent with recent research on detecting equity premium predictability using orthodox t-tests (e.g. Campbell and Thompson, 2008; Neely *et al.*, 2014).

[5] Our full sample of data used for the equity premium application below is monthly and covers the period December 1974 to December 2015 (hence $T = 493$), and in the application we monitor from January 2000 (hence $T^* + m = 302$). In addition to $m = 30$, in the empirical application results are also computed for $m = 20$ and $m = 60$.

[6] This range of values for ρ and β_1 was chosen following a preliminary analysis of the data used for the empirical application in section 6. Typically when AR(1) models are estimated for the traditional predictors used in section 6 (e.g. the valuation ratios), the fitted slope parameters lie in the range 0.965-0.998, and the majority of the $\hat{\beta}$ values obtained from subsample predictive regressions with $m = 30$ lie in the range 0.10-1.00.

[7] Therefore in these experiments $m = m_1$. In the third and fourth sets of experiments, discussed in more detail below, we investigate the performance of our monitoring procedure when the values of m and m_1 differ.

predictive regime dates, but extending the monitoring period to $E = 362$ which is consistent with $\alpha = 0.20$. Hence the empirical FPR obtained from the simulations in this case (when $\beta_1 = 0$) should be approximately equal to 0.20.

The third set of experiments study the power of our procedure to detect a single predictive regime as a function of its length, m_1, for $E = 328$ and for the predictive regime dates used in the previous experiments. The AR(1) parameter for the predictor's DGP in (3) is set to $\rho = 0.995$, $\beta_1 = \{0.25, 0.50, 0.75, 1.00\}$, while the other parameters are set to the values used for the previous experiments. Results are computed for $m_1 = \{10, 15, ..., 60\}$ in steps of 5 observations. We would expect that increases in m_1 will lead to increases in power, although the extent of the increase will depend on the length of the monitoring period and the location of the predictive regime in the DGP. In the fourth set of experiments we repeat the third set of experiments employing the same DGP and predictive regime dates, but now extending the monitoring period to $E = 362$.

The first four sets of experiments assume no predictability over the training period. As discussed in section 4.3, our procedure can still be used for detecting predictive regimes during the monitoring period if predictability exists during the training period, albeit the FPR and power of the procedure could be affected. Recall from section 4.3 that setting $m^* = \widehat{m}^*$, the longest contiguous sequence of $\tau_{e,m}$ exceedances over the training period, controls the (theoretical) FPR of our monitoring procedure to α. If our procedure is applied to data where a regime of positive predictability exists in the DGP during the training period, the number of contiguous right-tailed $\tau_{e,m}$ exceedances over the training period, and therefore the value of $\widehat{m}^*$ used for monitoring, are likely to be larger than the values obtained if the DGP had contained no predictability over the training period but was otherwise identical. It follows straightforwardly in this case that the power of our procedure to detect a predictive regime over the monitoring period will be reduced relative to the case of no predictability over the training period. The empirical FPR might also be affected by the presence of predictability over the training period as it is also a function of $\widehat{m}^*$.

The fifth through eighth sets of experiments investigate this issue in more detail. In these experiments we repeat the first four sets of experiments again using the DGP given by (1)-(3), but in addition to the original predictive regime at locations (a)-(e), an earlier predictive regime is imposed in the DGP during the relevant training periods. Specifically, the full DGP for each set of experiments contains two predictive regimes (i.e. we set $n = 2$ in (1)), where the first predictive regime is set to occur during the training period at $t = \lfloor T^*/2 \rfloor$, and we set $m_1 = 15$ and $\beta_1 = 0.25$ (hence the associated predictive regime in the training period continues for 15 observations). The second predictive regime mirrors the original predictive regime in the first four sets of experiments. The length of this second regime, m_2, and the strength of the predictability, β_2, are set to the same values as the relevant parameters in the first four sets of experiments (m_1 and β_1, respectively). Note that in the fifth through eighth sets of experiments the predictive regime in the training period is relatively short (being half the length of the predictive regime in the monitoring period for first two sets of experiments). It is particularly important to assess the finite sample performance of our procedure when there is a short predictive regime in the training period, since short predictive regimes are more difficult to identify than long predictive regimes. If a long predictive regime exists over the initial training period chosen by a researcher using our procedure, then it is more likely that the researcher would be aware of its presence (e.g. via a preliminary analysis of the data). The researcher might choose to continue monitoring using the critical value cv_π and $\widehat{m}^*$ computed from the initial training period data, but take into account the presence of the earlier predictive regime when interpreting the results obtained, or alternatively they might choose to select a different training period before using our procedure for monitoring; cf. the discussion in section 4.3 relating to equation (15).

As discussed in section 1, an attractive feature of our monitoring procedure is that, for sufficiently large T, in addition to being robust to any degree of contemporaneous correlation of the error terms in the DGP, it is also robust to conditional and/or unconditional heteroskedasticity, and to non-Gaussian errors. To investigate how well these robustness properties hold in finite samples, in additional experiments we repeated a selection of the simulation experiments discussed above using the same DGPs but for a range of error distributions and heteroskedasticity patterns for $\epsilon_{y,t}$ in (1), specifically: (i) $t(10)$ error terms; (ii) $t(5)$ error terms; (iii) normally distributed GARCH(1,1) error terms with GARCH parameters $\alpha_0 = 0.10$, $\alpha_1 = 0.10$, $\beta = 0.80$; (iv) $t(5)$ GARCH(1,1) error terms with the same GARCH parameters, and (v) $t(5)$ error terms with an unconditional volatility shift during the monitoring period from $\sigma_y = 1$ to $\sigma_y = 2$. In each case very similar results were obtained to the results from the original experiments reported here. In the interest of brevity the detailed results from these additional experiments are therefore relegated to an on-line supplementary appendix to this paper available from `www.sites.google.com/view/pr-supplemental`

5.1 Results

The results from the first set of experiments are given in Figure 2. Recall that the end of the monitoring period for this set of experiments is chosen using (11) to be consistent with a FPR of $\alpha = 0.10$. Therefore when $\beta_1 = 0$ we would expect the predictive regime detection frequency computed from our simulations to be close to 0.10. It can be seen that each of the curves indeed starts from approximately 0.10 consistent with the theoretical results in section 4.3. For cases (a)-(c) when the predictive regime starts before or at the same time as the start of monitoring, power rises rapidly with β_1. For cases (d) and (e) when the predictive regime starts after the start of monitoring, a higher proportion of the subsamples used when computing $\tau_{e,m}$ will be data from the period of the DGP when no predictability exists. Furthermore, in these two cases monitoring ends shortly after the predictive regime starts (e.g. for case (e), monitoring ends 11 observations after the predictive regime starts). Therefore, as expected, power rises with β_1 at a lower rate than for cases (a)-(c) and ultimately flattens out at a lower value.

The results from the second set of experiments are given in Figure 3. As expected, when the monitoring period is extended to $E = 361$ the predictive regime detection frequency as a function of β_1 increases. Indeed the results are now virtually identical for each of the predictive regime start dates considered here and all of the curves flatten out quickly as β_1 increases. This reflects the fact that because of the longer monitoring period, each set of sequential $\tau_{e,m}$ statistics now includes a run of statistics computed using subsamples where a high proportion of each subsample is data from when predictability exists in the DGP. When $\beta_1 = 0$ the empirical FPR increases to approximately 0.20, again as expected.

Consider next the results from the third set of experiments given in Figure 4. As expected, power initially increases with m_1. Note that for case (a) the curve flattens out at between 0.85 and 0.95 (depending on the value of β_1) when $m_1 = 40$. For cases (b)-(e) the curve flattens out earlier and at a lower value. This pattern reflects the fact that as we move from cases (a) to (e), because the predictive regime starts progressively later in the sample, the value of m_1 such that the end of the predictive regime lies beyond the end of the monitoring period E gets smaller. Hence for case (e), the curve is relatively flat for $m_1 > 10$ because the monitoring period ends 11 observations after the start of the predictive regime in the DGP. Therefore, in this case, further increases in m_1 above 10 do not lead to any further increase in the power of our procedure to detect the predictability regime.

For the fourth set of experiments the results given in Figure 5 show that power as function of m_1 is now very similar, irrespective of when the predictive regime occurs. With the monitoring period finishing later in the sample there are sufficient observations in the monitoring period for increases in m_1 to translate through to increases in power before the monitoring period ends. In all of the cases (a)-(e) the curve indicates that our procedure has a very high probability of successfully detecting a predictive regime when $m_1 \geq 40$.

The results for the fifth through eighth set of experiments are given in Figures 6-9. We find that, as expected, due to the presence of a predictive regime during the training period, in each of the individual experiments the longest contiguous sequence of $\tau_{e,m}$ exceedances over the training period, and therefore the value of $\widehat{m}^*$ selected for monitoring, is on average slightly larger than the corresponding value in the first four sets of experiments. As a result, the power curves are generally lower in these experiments than the corresponding curves in the first four sets of experiments. When $\beta_2 = 0$ and $E = 328$ (consistent with $\alpha = 0.10$), the detection frequency in Figure 6 is approximately 0.05. When $\beta_2 = 0$ and $E = 362$ (consistent with $\alpha = 0.20$), the detection frequency in Figure 7 is approximately 0.10. Similarly, it can be seen in Figures 6 and 7 that for $\beta_2 > 0$, the curves are approximately 0.05-0.10 lower than the corresponding curves in Figures 2 and 3. The curves in Figure 6 for $E = 328$ are sensitive to where the second predictive regime is located. However it can be seen in Figure 7 that as in Figure 3, extending the monitoring period to $E = 362$ reduces the sensitivity of the curves to the exact location of the predictive regime.

We find in Figure 8 that power as a function of the length of the predictive regime (here m_2) also falls by approximately 0.05-0.10 compared to the original results in Figure 4. The curves in Figure 9 are also approximately 0.05-0.10 lower for smaller values of m_2 than the curves in Figure 5. Notice, however, that for larger values of m_2 (e.g. $m_2 \geq 50$) and larger values of β_2 there is virtually no loss of power relative to the original results and as before power is close to unity. Intuitively, when the predictive regime in the monitoring period lasts for a relatively long period of time compared with the predictive regime within the training period, and the predictability is relatively strong, it dominates the negative impact on power of the increase in $\widehat{m}^*$ caused by the predictive regime within the training period.

Overall, the results obtained from these simulation experiments illustrate that our method for controlling the FPR associated with our procedure works very well for the type of sample sizes and subsample sizes that might be used in empirical applications. The simulation results show that our procedure has very good power for detecting a predictive regime that occurs shortly before or around the same time as the start of monitoring, even when the monitoring period is short. Naturally, for predictive regimes that start towards the end of a short monitoring period the power of our procedure will be lower. Increasing the monitoring period increases the power of our procedure, *ceteris paribus*, enabling these predictive regimes to be better detected, albeit subject to an increased FPR. The power of our procedure also increases with the length of the predictive regime, up to an upper threshold that depends on the location of the predictive regime relative to the end of the monitoring period. The simulation results from the fifth through eighth sets of experiments show that our procedure is relatively robust to the presence of a short period of predictability within the training period. Although the power of the procedure falls in this case compared with the results obtained when there is no predictability within the training period, the impact is relatively small for the examples considered here and the results obtained suggest that our procedure should still be useful for detecting predictive regimes in empirical applications in cases where this situation arises.

6 Empirical Application

6.1 Data and Preliminary Analysis

The dataset used for the empirical application of our monitoring procedure consists of monthly observations on the equity premium for the S&P Composite index calculated using CRSP's month-end values and on 20 different predictors for the period 1974:12-2015:12 ($T = 493$). We define the equity premium as in Goyal and Welch (2008) and Neely *et al.* (2014) as the log return on the value-weighted CRSP stock market index minus the log return on the risk-free Treasury bill: $y_t = log(1 + R_{m,t}) - log(1 + R_{f,t})$ where $R_{m,t}$ is the CRSP return and $R_{f,t}$ is the Treasury bill return. Ten of the predictors are traditional macroeconomic and financial variables (MFVs) and ten are binary technical analysis indicators (TAIs) also used by Neely *et al.* (2014) in their analysis of equity premium predictability. The TAIs used are four moving average indicators (MAIs), two momentum indicators (MOIs), and four on-balance volume (OBV) indicators. The four moving-average rule indicators ($MAI_{s,l,t}$) are,

$$MAI_{s,l,t} := \begin{cases} 1, & \text{if } MA_{s,t} \geq MA_{l,t}, \text{ indicating a buy signal} \\ 0, & \text{otherwise,} \end{cases}$$

where $MA_{j,t} := (1/j) \sum_{i=0}^{j-1} P_{t-i}$ for $j = \{s, l\}$ and $s = \{1, 2\}$, $l = \{9, 12\}$ and where P_t is the level of the S&P Composite index. The two l-period momentum rule indicators ($MOI_{l,t}$) are,

$$MOI_{l,t} := \begin{cases} 1, & \text{if } P_t \geq P_{t-l}, \text{ indicating a buy signal} \\ 0, & \text{otherwise,} \end{cases}$$

where $l = \{9, 12\}$. The four on-balance volume rule indicators ($OBV_{s,l,t}$) are,

$$OBV_{s,l,t} := \begin{cases} 1, & \text{if } MA^{OBV}_{s,t} \geq MA^{OBV}_{l,t}, \text{ indicating a buy signal} \\ 0, & \text{otherwise,} \end{cases}$$

where $MA^{OBV}_{j,t} := (1/j) \sum_{i=0}^{j-1} obv_{t-i}$ for $j = \{s, l\}$ and $s = \{1, 2\}$, $l = \{9, 12\}$, and, $obv_t := \sum_{k=1}^{t} VOL_k D_k$, where VOL_k is trading volume for the S&P Composite index in period k and D_k is a binary variable,

$$D_t := \begin{cases} 1, & \text{if } P_t \geq P_{t-1} \\ -1, & \text{otherwise.} \end{cases}$$

The data used to construct the equity premium and the predictors are taken from the updated monthly data set on Amit Goyal's website (`www.hec.unil.ch/agoyal/`) which is an extended version of the data set used by Welch and Goyal (2008). The traditional MFVs are in log form (as in Goyal and Welch, 2008; Neely *et al.*, 2014) and each of the predictors is lagged one period. A full list of the predictors is given in Table 1. Graphs of the excess returns and the MFVs are given in Figures 10(a)-10(k). A graph of the S&P Composite index and the TAIs are given in Figure 11 (note that here the TAI's and index are scaled to fit on the same graph).

We begin with a preliminary analysis of our data set using some popular orthodox methods for detecting predictability. Table 2 reports, for each predictor variable considered, the estimated slope parameter ($\hat{\beta}$), a right-tailed Newey-West t-test of significance (t_{NW}) and the standard and adjusted R^2 values for orthodox bivariate regression models applied to the full sample of data using OLS for parameter estimation. For both the MFVs and the TAIs, consistent with many of the previous empirical studies discussed in section 1 very little evidence of predictability is provided by the t_{NW} tests run at conventional significance levels and in all cases the R^2 values are under 1%. It is important to recognize that although popular in studies of equity premium predictability, orthodox t-tests (including t_{NW}) can be misleading in this case because of the highly persistent lagged regressors used (see again the discussion in Section 1). Therefore also reported in Table 2 is the IV_{comb} test of Breitung and Demetrescu (2015). The asymptotic null distribution of this test statistic is standard normal, such that the test is valid, irrespective of the persistence of the predictor and any heteroskedasticity present in the errors. As discussed in Remark 4 of Breitung and Demetrescu (2015,p.364), the IV_{comb} test can only be validly implemented as a two-tailed test, even where the sign of β under predictability is known *a priori*. For the MVFs there is no statistically significant evidence of predictability from IV_{comb} at conventional significance levels, and only a single rejection at the 10% significance level for the TAIs.[8]

Recall that in outlining our monitoring procedure in section 4 we assumed in generating the empirical critical value, cv_{π}, that there was no predictability over the training periods. To assess how this assumption sits with our data sets we apply the same methods used for obtaining the full sample results in Table 2 to the training periods employed in the monitoring application below. Although we present the results for all of the methods used in Table 2, to assess the presence of predictability in these training periods we focus on the IV_{comb} test. For the monitoring application below, our initial choice of training periods is 12/74-05/98 (for $m = 20$), 12/74-07/97 (for $m = 30$), and 12/74-01/95 (for $m = 60$). These are the implied training periods given by $T^* = 302 - m$, where observation $t = 302$ is the date at which monitoring starts in the application below, 01/00. If there is statistically significant evidence of predictability for an initial choice of training period, but this is thought to be due to a period of predictability towards the end of that training period, then we recommend ending the training period at an earlier date so as to reduce the likelihood it contains predictability. Thus, the final training periods employed when monitoring could finish earlier than the initial choice of training period; see the discussion in section 4.3 relating to equation (15).[9]

Our preliminary analysis of the data over the implied training periods reveals that for the two interest rate series st_{t-1} and lt_{t-1}, and for the bond yield spread dsp_{t-1}, there is statistically significant evidence of predictability at conventional significance levels from IV_{comb} for one or

[8] Financial theory suggests negative predictive power for st_{t-1}, lt_{t-1}, $ntis_{t-1}$ and inf_{t-1}. We therefore multiply each of these predictors by -1 so that a right-sided test (excepting the IV_{comb} test which, as discussed above, is implemented as a two-tailed test) is appropriate for detecting predictability. See footnote 4 for further details.

[9] If predictability is present during the training period, as the simulations in Section 5 demonstrate, our procedure can still be useful for detecting positive predictability over the monitoring period. Note that if negative predictability exists over the training period and a predictability regime change is detected using the upper-tailed version of our procedure, we cannot conclude that the change is to a period of positive predictability without further analysis, because it could be due to a change to a period of no predictability.

more values of m. Furthermore, the rejections obtained do not appear to be driven by predictability at the end of these implied samples. Therefore, in the monitoring application below we continue to use the implied training periods for these three predictors despite the rejections from IV_{comb}. Statistically significant evidence of predictability from IV_{comb} is also obtained for $ntis_{t-1}$, for all values of m. In this case, we find that predictability is concentrated in the data from 01/92 through to the end of the training periods. Hence, for this predictor and for all values of m, we end the relevant training periods at 12/91 in the monitoring application below. For all of the other MFV and TAI predictors no statistically significant evidence of predictability is found from IV_{comb} using the implied training periods. The full set of results from the preliminary analysis of the data over the training samples (using the adjusted training sample for $ntis_{t-1}$) are given in Table 3 and Table 4 for the MFVs and TAIs respectively. Note from Table 3 that there is also evidence of predictability for inf_{t-1} from t_{NW} for the training periods associated with $m = \{20, 30\}$. However, this variable is a highly persistent process and so t_{NW} is not an appropriate test to use. Instead IV_{comb} should be used, and doing so reveals no statistically significant evidence of predictability at the 10% level.

6.2 Monitoring Results

6.2.1 Monitoring Results: Macroeconomic and Financial Variables

We assume that a practitioner applies our real-time monitoring procedure for the presence of predictive regimes from 01/00 (thus in all cases $T^* + m = 302$). Results are presented assuming that the monitoring continues through to the final observation in the data set, 12/15. In real-world applications it is not envisaged that our procedure would be used for continuous monitoring over anything like such a long period, but it is helpful to present the results through to 12/15 to illustrate the relationship between the length of the monitoring period and the FPR. Results are computed for $m = \{20, 30, 60\}$, for both 10% and 5% level estimated critical values, i.e. cv_π for $\pi = \{0.10, 0.05\}$.

For each predictor and value of m considered, the value of $\widehat{m}^*$ of (13) (the longest contiguous sequence of exceedances in the training period) and the number of predictive regimes detected are given in Table 5. For all of the predictors and values of m in Table 5 where one or more regimes are detected using the 10% level critical values ($\pi = 0.10$), graphs of the subsample $\tau_{e,m}$ values along with the relevant critical values are given in Figure 12. For presentation purposes we do not display $\tau_{e,m}$ over the entire training period and instead start the horizontal axis five years before the end of each training period. Also indicated on these graphs are the end of the training period T^*, the date when monitoring starts $T^* + m$, the date of the first significant rejection for the i-th predictive regime j_i, the date at which the i-th predictive is detected $j_i + m^*$, the FPR as a function of E (computed using (12)), the weak set of predictive regime dates and, where relevant, the strong set of predictive regime dates. Observe from section 4.4 that if more than one predictive regime is detected it is possible that the weak dates associated with consecutive predictive regimes can overlap, although no such possibility can arise for the strong dates. Notice also that if a predictive regime does exist for the data being examined and is detected by our procedure, but the contiguous rejections continue for a relatively short period of time, then the strong set of dates will typically be empty.

Neely *et al.* (2014) investigate differences in predictability between macroeconomic recession and expansion periods by computing separate R^2 statistics for predictive regression models using the NBER indicator of recessions and expansions to partition the relevant data. They find that for both the MFVs and TAIs predictability is substantially higher over recessions than over expansions. In the light of these findings it is interesting to compare the subsample $\tau_{e,m}$ values over the monitoring period with the NBER indicator to see if our procedure finds a similar pattern of support for predictability over the business cycle. Hence, the NBER indicator is also plotted in Figure 12. There are two US recessions over the monitoring period 01/00-12/15: one short recession in early 2001 (specifically, March 2001-November 2001), and one major recession associated with the global financial crisis (December 2007-June 2009).

It can be seen in the first half of Table 5 that, as expected, the largest number of contiguous rejections over the training period for each predictor, $\widehat{m}^*$, is sensitive to the value of π used (the test statistic significance level), being larger for $\pi = 0.10$ than for $\pi = 0.05$. The second half of Table 5 shows that in total, either one, two or three predictive regimes are detected for six of the ten MFVs when $\pi = 0.10$ and for five of the MFVs when $\pi = 0.05$. Note that the number of predictive regimes detected varies depending on the predictor and the value of m used when computing $\tau_{e,m}$. For ep_{t-1} and lt_{t-1}, one or more predictive regimes are detected for all values of m considered. For bm_{t-1}, a single predictive regime is detected when $m = 20$ and $m = 60$, while for dy_{t-1} and dp_{t-1} two or three predictive regimes are detected when $m = 60$ (but for these predictors no regimes are detected when $m = 20$ or $m = 30$). For dsp_{t-1} a single predictive regime is detected only when $m = 20$.

Consider next the graphical results for the MFVs given in Figure 12 that correspond to the results in Table 5. Figure 12(a) shows that for dy_{t-1}, the $\tau_{e,m}$ test first rejects at 06/01, and further contiguous rejections confirm that a predictive regime appears to have occurred in 02/02 with the associated weak set of dates given by 07/96-05/02. This first predictive regime is consistent with the period of the dot-com bubble, being detected towards the end of the boom with the associated predictive regime ending shortly after the crash. As the monitoring period continues a further rejection occurs in 03/05, with confirmation of a second predictive regime in 11/05. The weak set of dates in this case are 04/00-02/06 which therefore overlap with the weak set of dates for the first predictive regime. The FPR associated with the first predictive regime in this case is approximately 0.12. Because it occurs later in the monitoring period the FPR associated with the second predictive regime is higher, being approximately 0.28. As expected, the sequential $\tau_{e,m}$ values in Figure 12(b) are similar to those obtained for

dy_{t-1} in Figure 12(a), although for dp_{t-1} three predictive regimes are detected. The first and second regimes cover approximately the same sub-periods as the two regimes detected when dy_{t-1} is used as a predictor. The third predictive regime in this case occurs much later in the monitoring period, with the initial rejection in 02/14 and confirmation of a predictive regime in 01/15. According to the weak sets of dates for each predictive regime, in total dp_{t-1} had predictive power for the equity premium over the majority of the monitoring period. Notice that the breakdown of predictability during the monitoring period over 03/06-02/09 correlates with the declining equity premium that occurred as a consequence of the global financial crisis (see Figure 10(a) for a graph of the equity premium over this period). According to the weak set of dates our results suggest that although the predictive relationship between dp_{t-1} and the equity premium was not present at the height of the global financial crisis, it reappeared again in early 2009.

For ep_{t-1}, recall from Table 4 that when $m = 20$ a single predictive regime is detected and Figure 12(c) shows that this occurs in 02/04. The associated FPR is approximately 0.16 and the weak set of dates spans the period 04/02-02/04. Therefore in this case the identified predictive regime does not continue beyond the point at which our procedure detects its presence. Figure 12(d) shows that when $m = 30$ the first predictive regime is detected in 01/04, the second in 03/08, and the third in 09/15. Similar to the results for $m = 20$, in this case the first predictive regime ends two months after its detection, while the second and third predictive regimes end immediately following their detection. When $m = 60$ a single predictive regime is detected in 12/04 and the weak set of dates are 09/98-06/06. The results for bm_{t-1} in Figure 12(f) ($m = 20$) and Figure 12(g) ($m = 60$) are similar to the results for dp_{t-1} in the sense that in both cases the predictive regime is detected early in the monitoring period around the time of the dot-com bubble, and the predictive regime ends one month later.

For lt_{t-1} with $m = 20$ Figure 12(h) shows that the single predictive regime referred to in Table 4 is detected in 09/03 and the weak predictive regime dates are 04/01-03/04. In Figure 12(i) for lt_{t-1} with $m = 30$ similar results are obtained. In Figure 12(j) for lt_{t-1} with $m = 60$ two predictive regimes are detected, the first in 08/05 and the second in 11/12. Recall from the results of our preliminary analysis of the data over the training periods reported in Table 3 that for this predictor, statistically significant evidence of predictability is detected during all three training periods by the IV_{comb} test statistic. Thus, the monitoring results for this predictor illustrates the power of our procedure to detect changes in predictability regimes over a monitoring period when predictability also exists during the relevant training period. In Figure 12(k) for dsp_{t-1} with $m = 20$ a single predictive regime is detected in 12/12 and the weak set of dates in this case cover the period 09/10-01/13. Interestingly, this period is one in which the Federal Reserve was operating a policy of quantitative easing. Our results suggest that this may have influenced the predictive relationship between the yield spread and the equity premium at this time.

We can observe that in nearly all of the results obtained for the MFVs, the contiguous run of

$\tau_{e,m}$ rejections associated with each predictive regime ends shortly after the regime is detected. A consequence of this is that the strong set of dates for each predictor is empty. Furthermore, from a practical perspective this is an important result because it suggests that although investors using our procedure in real-time would have been able to detect predictability in these cases, there would have been very little time after the point of detection to exploit the predictability before it no longer existed, at least according to the results from our dating procedure. Paye and Timmermann (2006) and Timmermann (2008) argue that if predictability reflects market inefficiencies then it is only ever likely to be a short-lived phenomenon because when it exists, investors will quickly allocate capital to exploit its presence. Our finding of short pockets of predictability that end quickly after being detected is entirely consistent with this view.

Consider now the sequential $\tau_{e,m}$ values in these graphs relative to the NBER indicator of recessions and expansions. Interestingly it can be seen that in some cases there is evidence suggesting that, consistent with the findings in Neely *et al.* (2014), predictability is stronger during recessions than during expansions, but it is not a pattern obtained for all of the predictors. Consider for example the graphs for ep_{t-1} in Figure 12(c)-12(e). The test statistic $\tau_{e,m}$ peaks during both the 2001 and 2008-2009 recessions for all three values of m. However in Figures 12(h)-12(j) for lt_{t-1} the test statistics falls during or at the start of both recessions. For dy_{t-1} and dp_{t-1} in Figures 12(a) and 12(b) the test statistic $\tau_{e,m}$ peaks during the first recession but reaches a minimum during the 2008-2009 recession.[10]

6.2.2 Monitoring Results: Technical Analysis Indicators

Table 6 shows the value of $\widehat{m}^*$ of (13) for each of the TAIs and the number of predictive regimes detected by our procedure. Whilst the $\widehat{m}^*$ values reported in the first half of Table 6 are broadly similar to those obtained for the MFVs in Table 5, on average they are slightly higher for the TAIs. It can be seen in the second half of Table 6 that for all of the TAIs either one, two, or three predictive regimes are detected for at least one of the values of m considered here. Interestingly, and consistent with the findings in Neely *et al.* (2014), we therefore find stronger evidence of predictability for the TAIs than for the MFVs. Notice that the TAI predictors are 0-1 dummy variables that will often take the same value for several consecutive observations, and consequently the subsample $\tau_{e,m}$ values can be undefined when the TAI does not change over the subsample. If $\tau_{e,m}$ is undefined during the monitoring period it simply means that at the relevant observation when this occurs the test statistic is uninformative about the presence of predictability, but the $\tau_{e,m}$ values that *are* defined can still be used for monitoring. However, a large number of undefined test statistics in the training period could have a detrimental impact on the finite sample performance of the procedure. For completeness, the results for $m = \{20, 30\}$ are

[10] We note that Neely *et al.* (2014) study a longer sample of the data than the sample used here that ends earlier (12/50-12/11) and their empirical work is fundamentally different to ours, being an *ex post* analysis of predictability (in-sample and out-of-sample) rather than a real-time monitoring application.

reported in Table 6, although for some of the TAIs undefined test statistics occur quite frequently over the training period with these values of m. In practice, we recommend using $m \geq 60$ when using our procedure with these particular TAIs to minimize the number of undefined test statistics over the training period. Alternatively, for a given value of m, reducing the value of l when constructing the TAIs will result in fewer undefined test statistics. In the application here we report results for $l = \{9, 12\}$ to be consistent with the regression-based analysis of TAIs in Neely *et al.* (2014), even though for some of the MOIs and OBV indicators with $m = 60$ and these values of l, $\tau_{e,m}$ is occasionally undefined over the training and/or monitoring period. For the MAIs with $l = \{9, 12\}$ and $m = 60$ there are no undefined test statistics.

For the TAIs where one or more predictive regimes are detected using the 10% level critical values ($\pi = 0.10$), graphs of the sequences of subsample $\tau_{e,m}$ values along with the relevant critical values are given in Figure 13. For brevity, the graphs are presented for $m = 60$ and $s = 1$ only.[11] Figure 13(a) shows that for $MAI_{1,9,t-1}$, the first predictive regime is detected in 06/04 and the second in 03/09, with FPRs of approximately 0.23 and 0.38, respectively. The first predictive regime is correlated with the dot-com bubble (although it is not detected until after the crash) and the second predictive regime is correlated with the global financial crisis. The weak set of dates for the predictive regimes are 10/98-02/06 and 07/03-08/13. Therefore the two identified predictive regimes are overlapping in this case. Notice that for the second predictive regime in this case the strong set of dates is not empty, and spans the period 06/08-09/08. Interestingly, for both predictive regimes the contiguous rejections continue for much longer after the predictive regimes are first detected than we found for the MFVs. Specifically, for the first predictive regime the contiguous rejections continue for 20 months. For the second predictive regime the contiguous rejections continue for over four years following the initial detection. Clearly for this TAI, the pockets of predictability which our procedure detects are considerably longer than for the MFVs. If the predictive regimes detected for the MFVs and TAIs considered here reflect market inefficiencies, then this finding suggests that investors were slower to exploit the inefficiencies for this TAI than for the MFVs, allowing the predictability to persist for a longer period. The results for $MAI_{1,12,t-1}$ in Figure 13(b) show that similar to $MAI_{1,9,t-1}$, two predictive regimes are detected in 04/04 and 03/09 respectively. However a short predictive regime is also detected earlier in the monitoring period when $MAI_{1,12,t-1}$ is used as a predictor. Specifically, for this regime the $\tau_{e,m}$ test statistic first rejects at 12/01 and a predictive regime is confirmed at 09/02. Hence this first predictive regime appears to be correlated with the dot-com bubble and crash.

The results for $MOI_{9,t-1}$ and $MOI_{12,t-1}$ in Figure 13(c) and 13(d) are similar to the results for the MAIs, in the sense that the first predictive regime identified is again correlated with the dot-com bubble, the second with the global financial crisis, and the contiguous rejections continue for a longer period of time after the initial predictive regimes are first detected than we

[11] Very similar results are obtained when $s = 2$.

found for the MFVs. For both of these predictors the strong set of dates for the first predictive regime is not empty. Because the first predictive regime is detected earlier in the monitoring period for the MOIs than for the MAIs, the associated FPR is lower, being approximately 0.12 for both of the MOIs. Notice also that for both of the MOIs the test statistic is undefined for the period 2000-2001 because the indicator has a fixed value over this period (it can be seen in Figure 11 that both of the MOIs were indicating a buy signal for most of the period 1995-2001). Figure 13(e) shows that for the volume-based indicator $OBV_{1,9,t-1}$ the single predictive regime referred to in Table 5 is detected in 05/10 and similar to the other TAIs the contiguous rejections continue for several years after the initial detection date. For $OBV_{1,12,t-1}$ in Figure 13(f), the predictive regime is detected six-months earlier in 12/09 and the contiguous rejections continue through to 08/13. In this case $\tau_{e,m}$ is undefined between 1999-2000.

Neely *et al.* (2014) find that similar to the MFVs, predictive regression models with TAIs have larger R^2 values for the NBER recession periods than for the expansion periods suggesting stronger in-sample predictability during recessions. Interestingly there is also evidence supporting this argument in our results. Consider, for example, Figure 13(b) for $MAI_{1,12,t-1}$. The first predictive regime is detected shortly after the 2001 recession and the second predictive regime is detected during the 2008-2009 recession. Notice also that for this predictor $\tau_{e,m}$ exceeds the relevant critical value line over the 2001 recession, but this is not recognised as a predictive regime because the contiguous run of rejections does not exceed the $\widehat{m}^*$ threshold for this predictor (given in Table 6). For both of the MOIs, predictive regimes are detected during or shortly after the 2001 and 2008-2009 recessions, and for the volume-based indicators predictive regimes are detected shortly after the 2008-2009 recession.

7 Conclusions

In this paper we have developed a new real-time monitoring procedure for detecting the emergence of predictive regimes. Our proposed method is based on the sequential application of standard heteroskedasticity-robust (predictive) regression t-statistics for predictability to end-of-sample data. Critical values for each test in the sequence are provided using subsampling methods applied to data in a training period. Unlike tests based on conventional full sample regression t-statistics for predictability, this renders the tests robust to both the degree of persistence and endogeneity of the regressors in the predictive regression. They are also robust to certain forms of heteroskedasticity in the shocks. A predictive regime is deemed to have occurred once a certain number of consecutive t-statistics in the sequence have exceeded this estimated critical value. This number can be set by the practitioner and we have discussed a data-based procedure for choosing it, based on the longest run of exceedances in the training period, such that the false positive rate of the monitoring procedure can be controlled, for a given monitoring period length. Where the presence of a predictive regime is signalled by our procedure, we have also proposed procedures for dating the period over which the predictive relationship held. We have applied our proposed monitoring approach to investigate for the presence of regime changes in the predictability of the U.S. equity premium at the one-month horizon by traditional macroeconomic and financial variables, and by binary technical analysis indicators. Our results suggest that the one-month ahead equity premium has displayed episodes of temporary predictability and that these episodes could have been detected in real-time by practitioners using our proposed methodology.

Acknowledgements

Taylor gratefully acknowledges financial support provided by the Economic and Social Research Council of the United Kingdom under research grant ES/R00496X/1.

References

Andrews, D.W.K. (2003). End-of-sample instability tests, Econometrica, 71, 1661-1694.

Andrews, D.W.K. and Kim, J-S. (2006). Tests for cointegration breakdown over a short time period, Journal of Business & Economic Statistics, 24, 379-393.

Ang, A. and Bekaert, G. (2007). Stock return predictability: is it there? Review of Financial Studies, 20, 651-707.

Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. Econometrica, 66, 47-78.

Bai, J. and Perron, P. (2003). Computation and analysis of multiple structural change models, Journal of Applied Econometrics, 18, 1-22.

Boudoukh, J.R., Michaely, M., Richardson, P. and Roberts, M.R. (2007). On the importance of measuring payout yield: Implications for empirical asset pricing, Journal of Finance, 62, 877-915.

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

Brock, W., Lakonishok, J. and LeBaron, B. (1992). Simple technical trading rules and the stochastic properties of stock returns, Journal of Finance, 47, 1731-1764.

Campbell, J.Y. (1987). Stock returns and the term structure, Journal of Financial Economics 18, 373-400.

Campbell, J.Y. and Shiller, R.J. (1988a). Stock prices, earnings and expected dividends, Journal of Finance, 43, 3, 661-676.

Campbell, J. and Shiller, R.J. (1988b). The dividend-price ratio and expectations of future dividends and discount factors, Review of Financial Studies, 1, 195-228.

Cochrane, J.H. (2008). The dog that did not bark: a defense of return predictability, Review of Financial Studies, 21, 1533-1575.

Fama, E.F. (1981). Stock returns, real activity, inflation, and money, American Economic Review, 71, 545-565.

Fama, E.F. (1990). Stock returns, expected returns, and real activity, Journal of Finance, 45, 1089-1108.

Goyal, A. and Welch, I. (2003). Predicting the equity premium with dividend ratios, Management Science, 49, 639-654.

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

Hudson, R., Dempsey, M. and Keasey, K. (1996). A note on the weak form efficiency of capital markets: the application of simple technical trading rules to UK stock prices - 1935 to 1994, Journal of Banking and Finance, 20, 1121-1132.

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

Keim, D.B. and Stambaugh, R.F. (1986). Predicting returns in the stock and bond markets, Journal of Financial Economics, 17, 357-390.

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

Lettau, M. and Ludvigsson, S. (2001). Consumption, aggregate wealth, and expected stock returns. Journal of Finance, 56, 815-850.

Neely, C.J., Rapach, D.E., Tu, J. and Zhou, G. (2014). Forecasting the equity risk premium: the role of technical indicators, Management Science, 60, 1772-1791.

Nelson C.R. and Kim M.J. (1993). Predictable stock returns: the role of small sample bias, Journal of Finance, 48, 641-661.

Paye, B.S. and Timmermann, A. (2006). Instability of return prediction models, Journal of Empirical Finance, 13, 274-315.

Stambaugh, R.F. (1999). Predictive regressions, Journal of Financial Economics, 54, 375-421.

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

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

White, H. (1982). Maximum likelihood estimation of misspecified models, Econometrica, 50, 1–25.

Table 1. List of predictors used

Macroeconomic and financial variables (MFVs)
1. log dividend yield (dy_{t-1})
2. log dividend-price ratio (dp_{t-1})
3. log earnings-price ratio (ep_{t-1})
4. book-to-market ratio (bm_{t-1})
5. short term yield (st_{t-1})
6. long-term yield (lt_{t-1})
7. long-term - short-term yield spread ($sp_{t-1} = lt_{t-1} - st_{t-1}$)
8. BAA-AAA corporate bond yield spread (dsp_{t-1})
9. net equity expansion ($ntis_{t-1}$)
10. inflation (inf_{t-1})
Technical analysis indicators (TAIs)
1. 1-9 moving average rule ($MAI_{1,9,t-1}$)
2. 1-12 moving average rule indicator ($MAI_{1,12,t-1}$)
3. 2-9 moving average rule ($MAI_{2,9,t-1}$)
4. 2-12 moving average rule ($MAI_{2,12,t-1}$)
5. 9 period momentum rule ($MOI_{9,t-1}$)
6. 12 period momentum rule ($MOI_{12,t-1}$)
7. 1-9 on balance volume rule ($OBV_{1,9,t-1}$)
8. 1-12 on balance volume rule ($OBV_{1,12,t-1}$)
9. 2-9 on balance volume rule ($OBV_{2,9,t-1}$)
10. 2-12 on balance volume rule ($OBV_{2,12,t-1}$)

Table 2. Preliminary results for the full sample, 12/74-12/15

	$\hat{\beta}$	t_{NW}	IV_{comb}	$R^2(\%)$	$\bar{R}^2(\%)$
		MFVs			
dy_{t-1}	0.546	1.265	0.506	0.313	0.110
dp_{t-1}	0.576	1.300*	0.606	0.345	0.141
ep_{t-1}	0.424	0.743	1.072	0.225	0.022
bm_{t-1}	0.497	0.679	0.544	0.106	-0.098
st_{t-1}	0.042	0.743	0.350	0.116	-0.087
lt_{t-1}	0.036	0.513	0.398	0.056	-0.148
sp_{t-1}	0.108	0.826	-0.025	0.129	-0.075
dsp_{t-1}	0.135	0.216	0.064	0.021	-0.183
$ntis_{t-1}$	-0.005	-0.031	0.883	0.000	-0.204
inf_{t-1}	0.518	0.775	0.656	0.148	-0.056
		TAIs			
$MAI_{1,9,t-1}$	0.431	0.838	0.513	0.200	-0.004
$MAI_{1,12,t-1}$	0.647	1.125	1.142	0.415	0.212
$MAI_{2,9,t-1}$	0.453	0.870	1.126	0.215	0.012
$MAI_{2,12,t-1}$	0.802	1.490*	1.766*	0.648	0.445
$MOI_{9,t-1}$	0.370	0.635	0.209	0.136	-0.067
$MOI_{12,t-1}$	0.350	0.567	0.623	0.116	-0.088
$OBV_{1,9,t-1}$	0.491	1.011	0.045	0.269	0.065
$OBV_{1,12,t-1}$	0.679	1.281	0.150	0.488	0.285
$OBV_{2,9,t-1}$	0.759	1.503*	0.478	0.637	0.434
$OBV_{2,12,t-1}$	0.776	1.451*	0.761	0.642	0.439

Note. * denotes statistical significance at the 10% level. The critical value used for t_{NW} is 1.282. The critical value used for IV_{comb} is ± 1.645.

Table 3. MFVs: preliminary results for each training period used when monitoring with $m = \{20, 30, 60\}$

	$\hat{\beta}$	t_{NW}	IV_{comb}	$R^2(\%)$	$\bar{R}^2(\%)$
			$m = 20$		
dy_{t-1}	-0.328	-0.424	-0.132	0.061	-0.298
dp_{t-1}	-0.182	-0.243	0.226	0.019	-0.340
ep_{t-1}	0.026	0.041	-0.017	0.001	-0.358
bm_{t-1}	-0.237	-0.276	-0.233	0.027	-0.331
st_{t-1}	0.146	2.157*	2.440*	0.933	0.578
lt_{t-1}	0.179	1.624*	2.380*	0.762	0.406
sp_{t-1}	0.172	1.179	1.179	0.370	0.013
dsp_{t-1}	0.414	0.700	1.270	0.220	-0.138
$ntis_{t-1}$	0.362	2.717*	1.434	1.926	1.441
inf_{t-1}	1.290	1.902*	1.376	0.868	0.512
			$m = 30$		
dy_{t-1}	-0.262	-0.260	-0.086	0.031	-0.342
dp_{t-1}	-0.125	-0.129	0.175	0.007	-0.365
ep_{t-1}	0.084	0.122	0.008	0.005	-0.367
bm_{t-1}	-0.189	-0.204	-0.135	0.016	-0.355
st_{t-1}	0.146	2.120*	2.466*	0.943	0.575
lt_{t-1}	0.181	1.554*	2.442*	0.747	0.378
sp_{t-1}	0.186	1.276	1.144	0.436	0.066
dsp_{t-1}	0.477	0.776	1.409	0.283	-0.087
$ntis_{t-1}$	0.362	2.717*	1.434	1.926	1.441
inf_{t-1}	1.283	1.835*	1.333	0.854	0.485
			$m = 60$		
dy_{t-1}	1.513	1.088	0.890	0.651	0.234
dp_{t-1}	1.721	1.305	1.035	0.823	0.408
ep_{t-1}	0.569	0.770	0.632	0.226	-0.192
bm_{t-1}	0.854	0.787	0.675	0.272	-0.145
st_{t-1}	0.112	1.556*	2.405*	0.569	0.153
lt_{t-1}	0.108	0.838	2.200*	0.241	-0.177
sp_{t-1}	0.212	1.450*	1.027	0.602	0.186
dsp_{t-1}	1.084	1.758*	2.197*	1.316	0.903
$ntis_{t-1}$	0.362	2.717*	1.434	1.926	1.441
inf_{t-1}	0.928	1.244	0.929	0.446	0.029

Note. * denotes statistical significance at the 10% level. The critical value used for t_{NW} is 1.282. The critical value used for IV_{comb} is ± 1.645. The training periods are 12/74-05/98 (for $m = 20$), 12/74-07/97 (for $m = 30$), and 12/74-01/95 (for $m = 60$) for all predictors other than $ntis_{t-1}$. For $ntis_{t-1}$ the training periods are 12/74-12/91 for all values of m.

Table 4. TAIs: preliminary results for each training period used when monitoring with $m = \{20, 30, 60\}$

	$\hat{\beta}$	t_{NW}	IV_{comb}	$R^2(\%)$	$\bar{R}^2(\%)$
$m = 20$					
$MAI_{1,9,t-1}$	-0.503	-0.862	-0.660	0.276	-0.082
$MAI_{1,12,t-1}$	-0.043	-0.075	0.261	0.002	-0.357
$MAI_{2,9,t-1}$	-0.067	-0.126	0.245	0.005	-0.354
$MAI_{2,12,t-1}$	0.215	0.411	0.837	0.046	-0.313
$MOI_{9,t-1}$	-0.230	-0.386	0.275	0.053	-0.305
$MOI_{12,t-1}$	-0.447	-0.691	0.275	0.183	-0.175
$OBV_{1,9,t-1}$	0.380	0.692	0.756	0.157	-0.201
$OBV_{1,12,t-1}$	0.220	0.333	0.394	0.048	-0.310
$OBV_{2,9,t-1}$	0.533	0.859	0.961	0.293	-0.064
$OBV_{2,12,t-1}$	0.230	0.342	0.735	0.053	-0.305
$m = 30$					
$MAI_{1,9,t-1}$	-0.528	-0.896	-0.693	0.309	-0.062
$MAI_{1,12,t-1}$	-0.060	-0.104	0.236	0.004	-0.368
$MAI_{2,9,t-1}$	-0.086	-0.159	0.218	0.008	-0.364
$MAI_{2,12,t-1}$	0.201	0.380	0.819	0.040	-0.331
$MOI_{9,t-1}$	-0.250	-0.416	0.249	0.064	-0.308
$MOI_{12,t-1}$	-0.468	-0.717	0.248	0.204	-0.167
$OBV_{1,9,t-1}$	0.367	0.663	0.729	0.150	-0.222
$OBV_{1,12,t-1}$	0.206	0.309	0.360	0.043	-0.329
$OBV_{2,9,t-1}$	0.522	0.834	0.934	0.286	-0.084
$OBV_{2,12,t-1}$	0.215	0.318	0.707	0.048	-0.324
$m = 60$					
$MAI_{1,9,t-1}$	-0.792	-1.317	-1.090	0.703	0.288
$MAI_{1,12,t-1}$	-0.302	-0.506	-0.101	0.094	-0.324
$MAI_{2,9,t-1}$	-0.251	-0.448	0.016	0.068	-0.349
$MAI_{2,12,t-1}$	-0.030	-0.055	0.562	0.001	-0.418
$MOI_{9,t-1}$	-0.503	-0.829	-0.039	0.265	-0.153
$MOI_{12,t-1}$	-0.659	-0.997	-0.049	0.414	-0.003
$OBV_{1,9,t-1}$	0.130	0.228	0.460	0.019	-0.399
$OBV_{1,12,t-1}$	-0.026	-0.038	0.033	0.001	-0.418
$OBV_{2,9,t-1}$	0.299	0.463	0.643	0.096	-0.322
$OBV_{2,12,t-1}$	-0.019	-0.028	0.402	0.000	-0.418

Note. * denotes statistical significance at the 10% level. The critical value used for t_{NW} is 1.282. The critical value used for IV_{comb} is ± 1.645. The training periods are 12/74-05/98 (for $m = 20$), 12/74-07/97 (for $m = 30$), and 12/74-01/95 (for $m = 60$).

Table 5. MVFs: $\widehat{m}^*$ and number of predictive regimes detected

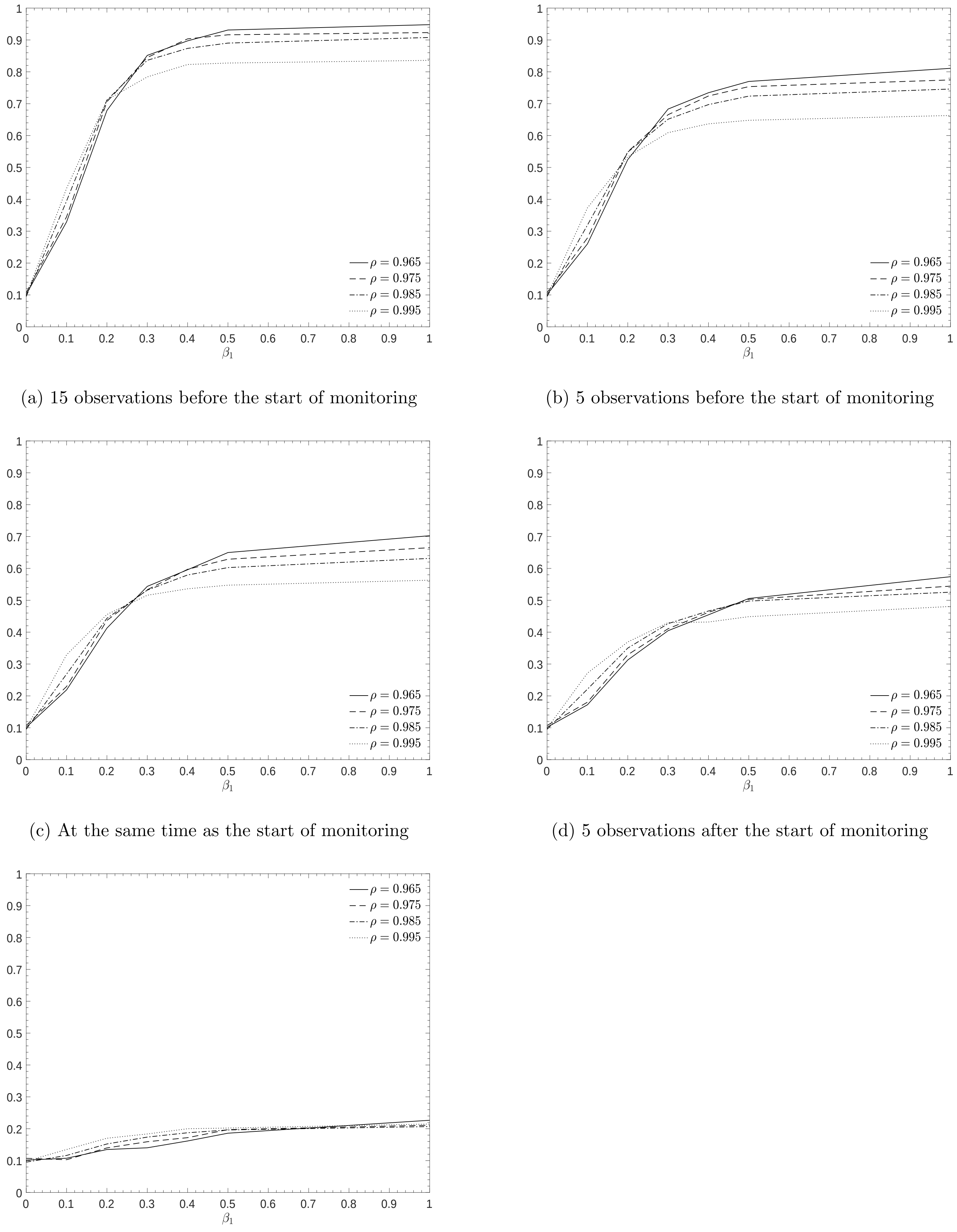
	$m = 20$		$m = 30$		$m = 60$	
π	0.10	0.05	0.10	0.05	0.10	0.05
			$\widehat{m}^*$			
dy_{t-1}	9	6	6	6	8	4
dp_{t-1}	7	2	11	3	11	4
ep_{t-1}	3	2	4	3	16	5
bm_{t-1}	10	4	14	4	8	3
st_{t-1}	9	8	17	9	5	3
lt_{t-1}	10	4	17	4	9	3
sp_{t-1}	8	4	16	5	15	4
dsp_{t-1}	8	6	8	7	5	4
$ntis_{t-1}$	7	5	6	3	10	5
inf_{t-1}	16	6	9	5	15	9
	Number of predictive regimes detected					
dy_{t-1}	0	0	0	0	2	2
dp_{t-1}	0	0	0	0	3	2
ep_{t-1}	1	2	3	0	1	0
bm_{t-1}	1	1	0	0	1	1
st_{t-1}	0	0	0	0	0	0
lt_{t-1}	1	3	1	1	2	1
sp_{t-1}	0	0	0	0	0	0
dsp_{t-1}	1	0	0	0	0	0
$ntis_{t-1}$	0	0	0	0	0	0
inf_{t-1}	0	0	0	0	0	0

Note. The training periods are 12/74-05/98 (for $m = 20$), 12/74-07/97 (for $m = 30$), and 12/74-01/95 (for $m = 60$) for all predictors other than $ntis_{t-1}$. For $ntis_{t-1}$ the training periods are 12/74-12/91 for all values of m.

Table 6. TAIs: $\widehat{m}^*$ and number of predictive regimes detected

	$m = 20$		$m = 30$		$m = 60$	
π	0.10	0.05	0.10	0.05	0.10	0.05
			$\widehat{m}^*$			
$MAI_{1,9,t-1}$	12	8	23	8	9	6
$MAI_{1,12,t-1}$	10	5	8	4	9	8
$MAI_{2,9,t-1}$	12	7	22	8	6	3
$MAI_{2,12,t-1}$	10	5	8	7	9	8
$MOI_{9,t-1}$	4	3	5	4	12	6
$MOI_{12,t-1}$	7	5	10	6	10	6
$OBV_{1,9,t-1}$	11	7	14	9	10	7
$OBV_{1,12,t-1}$	10	4	8	4	14	8
$OBV_{2,9,t-1}$	8	6	7	6	10	6
$OBV_{2,12,t-1}$	4	3	8	3	8	3
			Number of predictive regimes detected			
$MAI_{1,9,t-1}$	1	1	1	1	2	1
$MAI_{1,12,t-1}$	0	0	2	2	3	2
$MAI_{2,9,t-1}$	0	1	2	0	2	2
$MAI_{2,12,t-1}$	0	0	2	1	3	3
$MOI_{9,t-1}$	3	1	3	3	2	3
$MOI_{12,t-1}$	0	0	2	3	2	2
$OBV_{1,9,t-1}$	0	1	1	1	1	3
$OBV_{1,12,t-1}$	0	1	2	2	1	2
$OBV_{2,9,t-1}$	1	0	0	0	2	0
$OBV_{2,12,t-1}$	1	1	1	1	3	3

Note. The training periods are 12/74-05/98 (for $m = 20$), 12/74-07/97 (for $m = 30$), and 12/74-01/95 (for $m = 60$).



(a) 15 observations before the start of monitoring

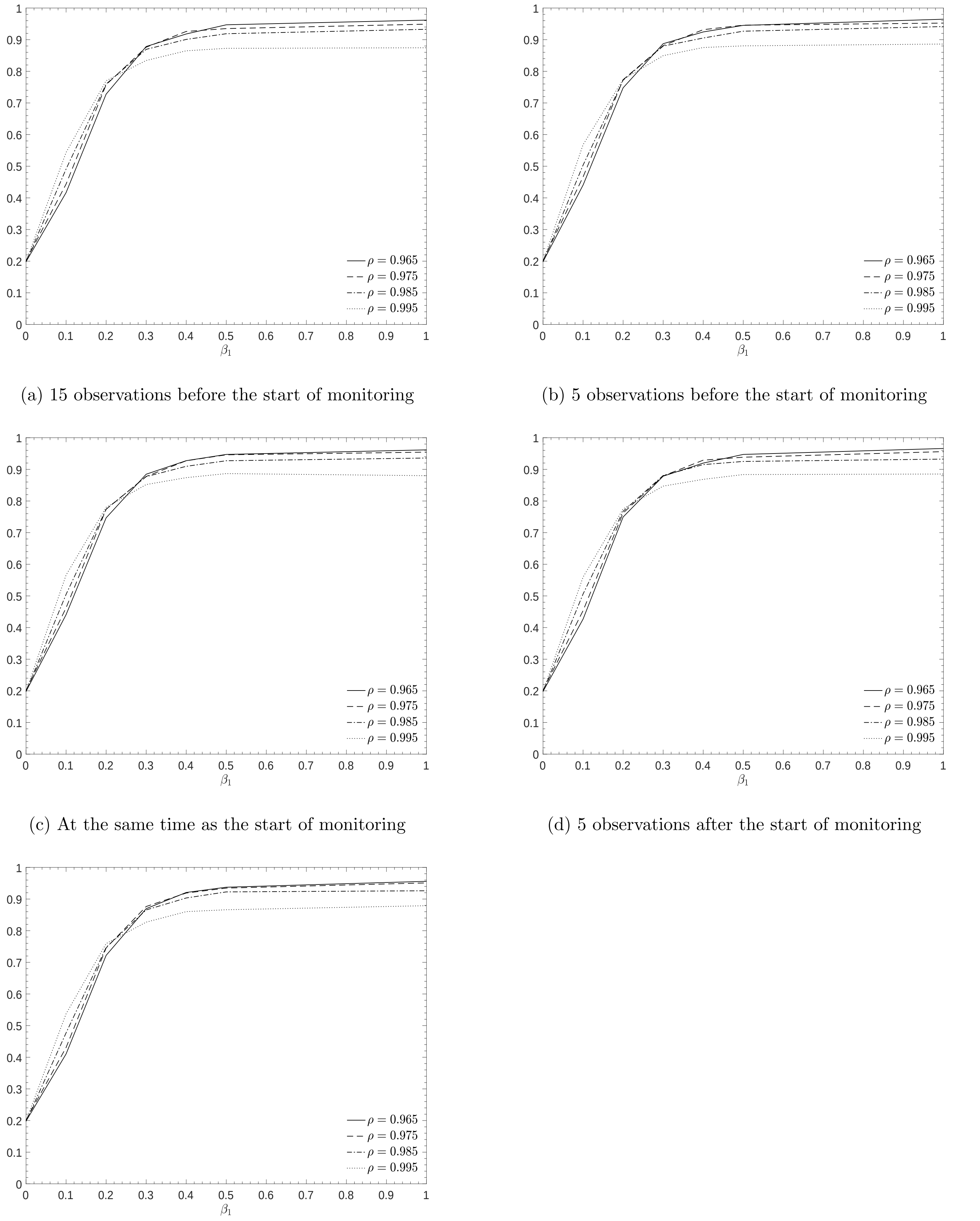
(b) 5 observations before the start of monitoring

(c) At the same time as the start of monitoring

(d) 5 observations after the start of monitoring

(e) 15 observations after the start of monitoring

Figure 2. Predictability regime detection frequency as a function of β_1 (predictability strength) for different values of ρ: $T = 493$, $T^* + m = 302$, $E = 328$, $m_1 = 30$, $m = 30$



(a) 15 observations before the start of monitoring

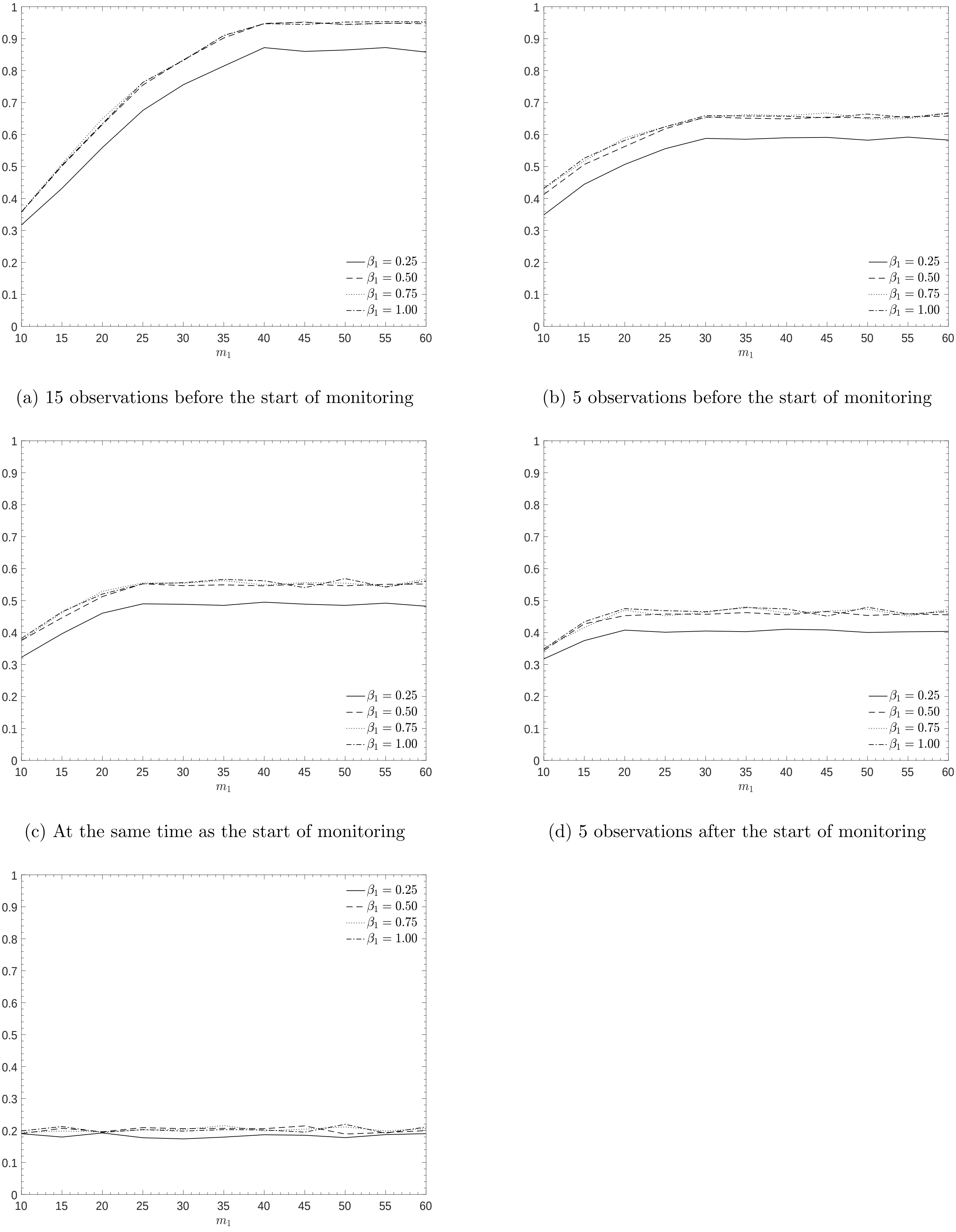
(b) 5 observations before the start of monitoring

(c) At the same time as the start of monitoring

(d) 5 observations after the start of monitoring

(e) 15 observations after the start of monitoring

Figure 3. Predictability regime detection frequency as a function of β_1 (predictability strength) for different values of ρ: $T = 493$, $T^* + m = 302$, $E = 362$, $m_1 = 30$, $m = 30$



(a) 15 observations before the start of monitoring

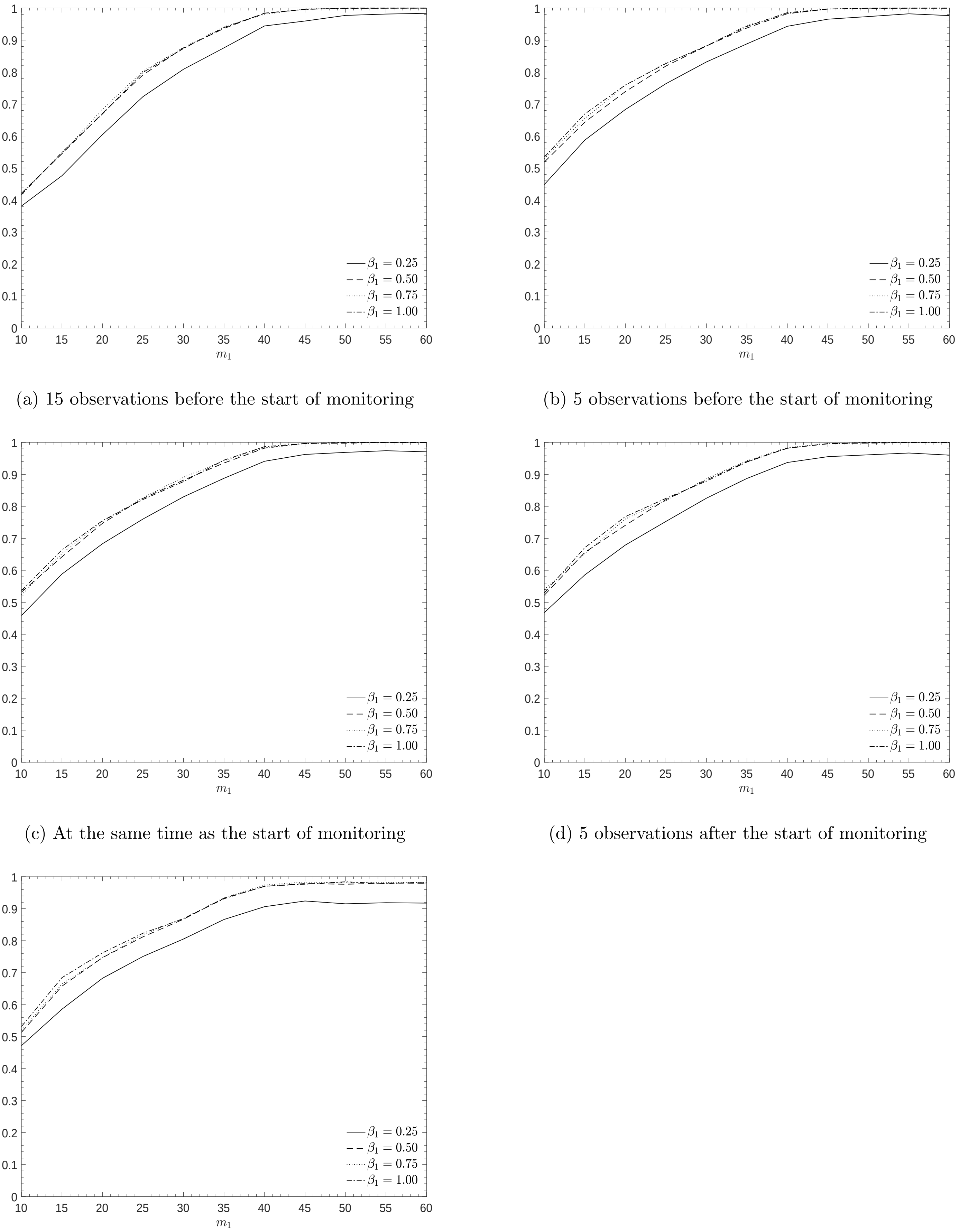
(b) 5 observations before the start of monitoring

(c) At the same time as the start of monitoring

(d) 5 observations after the start of monitoring

(e) 15 observations after the start of monitoring

Figure 4. Predictability regime detection frequency as a function of m_1 (predictability regime length) for different values of β_1: $T^* + m = 302$, $E = 328$, $m = 30$, $\rho = 0.995$



(a) 15 observations before the start of monitoring

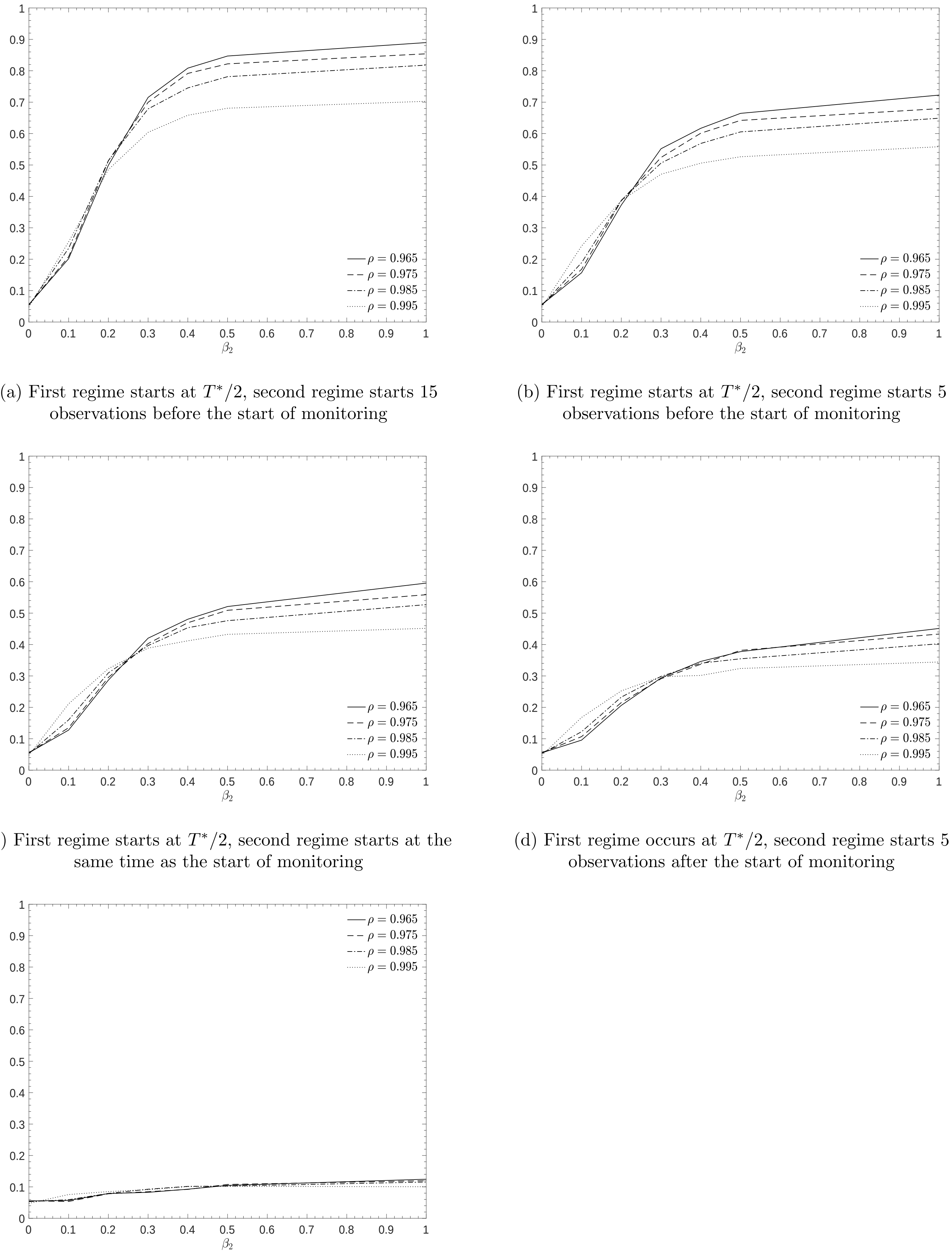
(b) 5 observations before the start of monitoring

(c) At the same time as the start of monitoring

(d) 5 observations after the start of monitoring

(e) 15 observations after the start of monitoring

Figure 5. Predictability regime detection frequency as a function of m_1 (predictability regime length) for different values of β_1: $T^* + m = 302$, $E = 362$, $m = 30$, $\rho = 0.995$



(a) First regime starts at $T^*/2$, second regime starts 15 observations before the start of monitoring

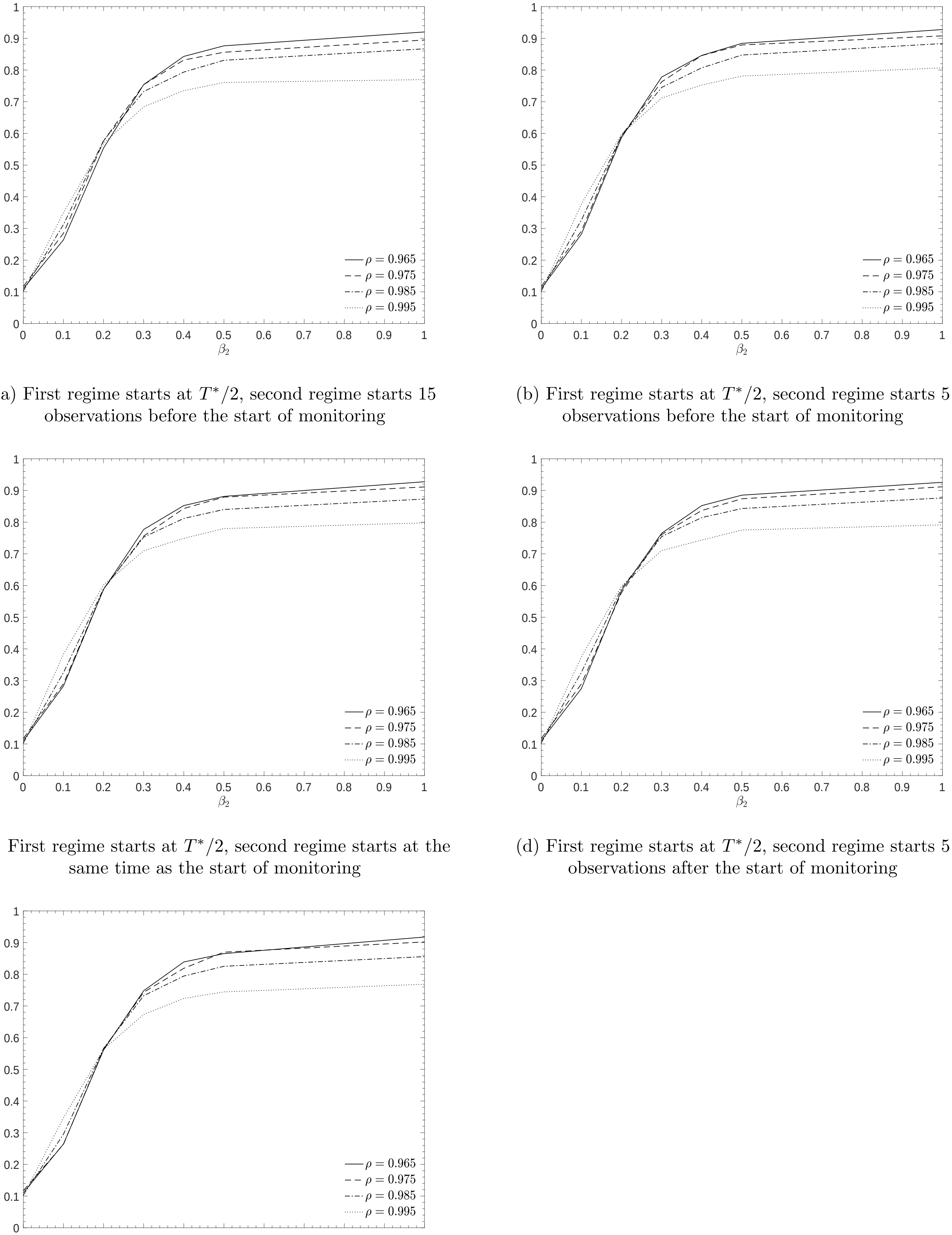
(b) First regime starts at $T^*/2$, second regime starts 5 observations before the start of monitoring

(c) First regime starts at $T^*/2$, second regime starts at the same time as the start of monitoring

(d) First regime occurs at $T^*/2$, second regime starts 5 observations after the start of monitoring

(e) First regime starts at $T^*/2$, second regime starts 15 observations after the start of monitoring

Figure 6. Detection frequency for second predictability regime as a function of β_2 (predictability strength) for different values of ρ: $T = 493$, $T^* + m = 302$, $E = 328$, $m_1 = 30$, $m_2 = 30$, $m = 30$



(a) First regime starts at $T^*/2$, second regime starts 15 observations before the start of monitoring

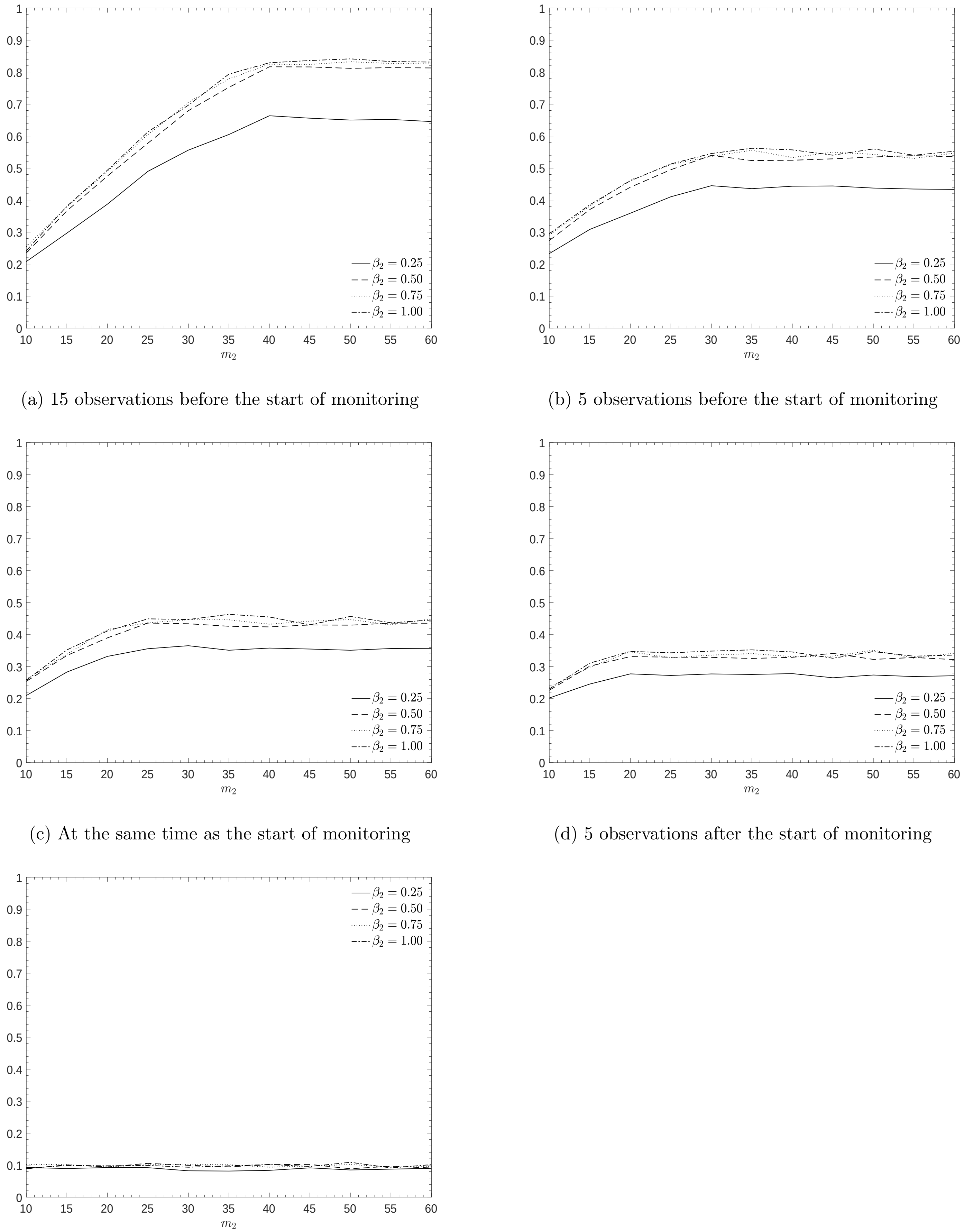
(b) First regime starts at $T^*/2$, second regime starts 5 observations before the start of monitoring

(c) First regime starts at $T^*/2$, second regime starts at the same time as the start of monitoring

(d) First regime starts at $T^*/2$, second regime starts 5 observations after the start of monitoring

(e) First regime starts at $T^*/2$, second regime starts 15 observations after the start of monitoring

Figure 7. Detection frequency for second predictability regime as a function of β_2 (predictability strength) for different values of ρ: $T = 493$, $T^* + m = 302$, $E = 362$, $m_1 = 30$, $m_2 = 30$, $m = 30$



(a) 15 observations before the start of monitoring

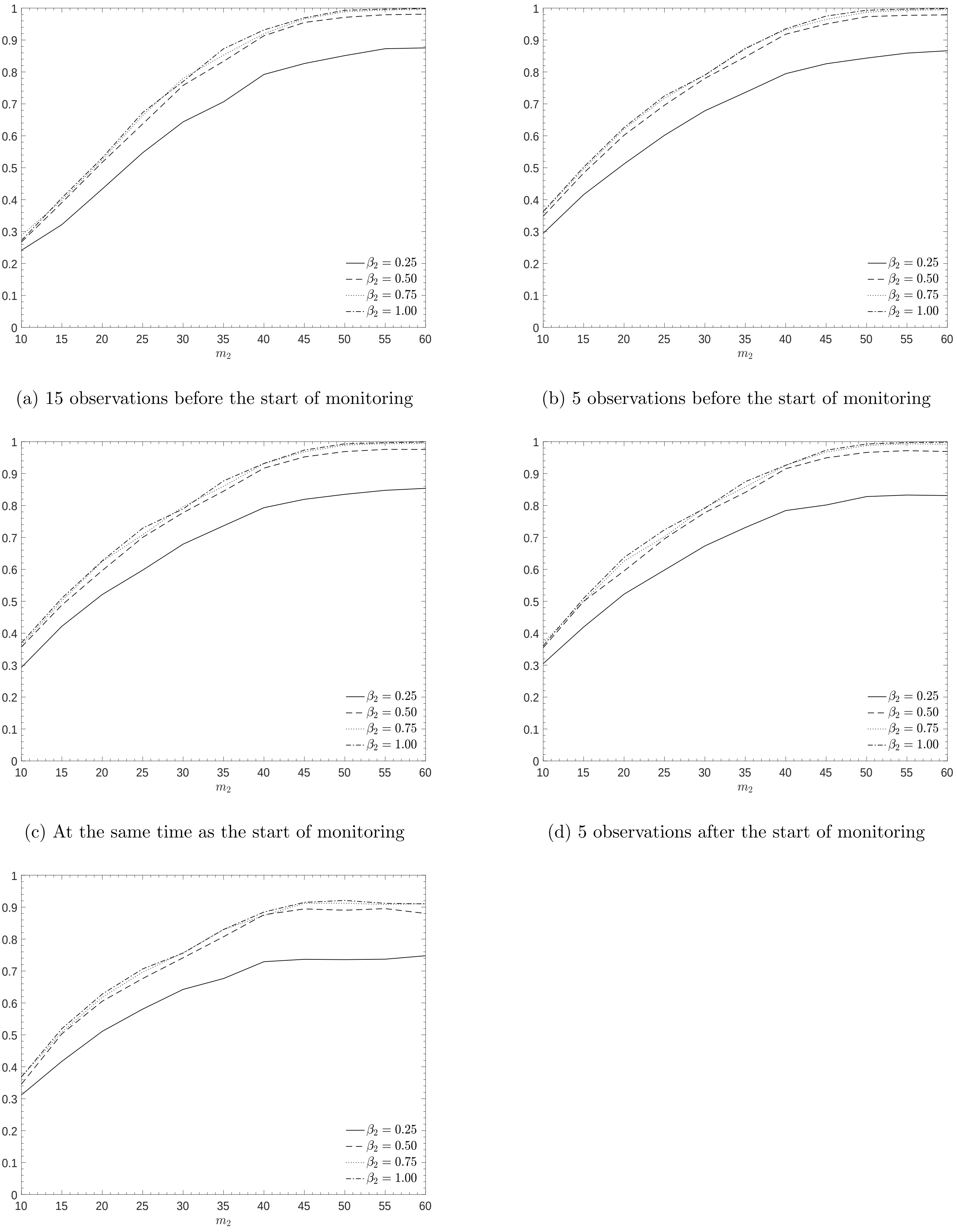
(b) 5 observations before the start of monitoring

(c) At the same time as the start of monitoring

(d) 5 observations after the start of monitoring

(e) 15 observations after the start of monitoring

Figure 8. Detection frequency for second predictability regime as a function of m_2 (predictability regime length) for different values of β_2: $T^* + m = 302$, $E = 328$, $m = 30$, $\rho = 0.995$



(a) 15 observations before the start of monitoring

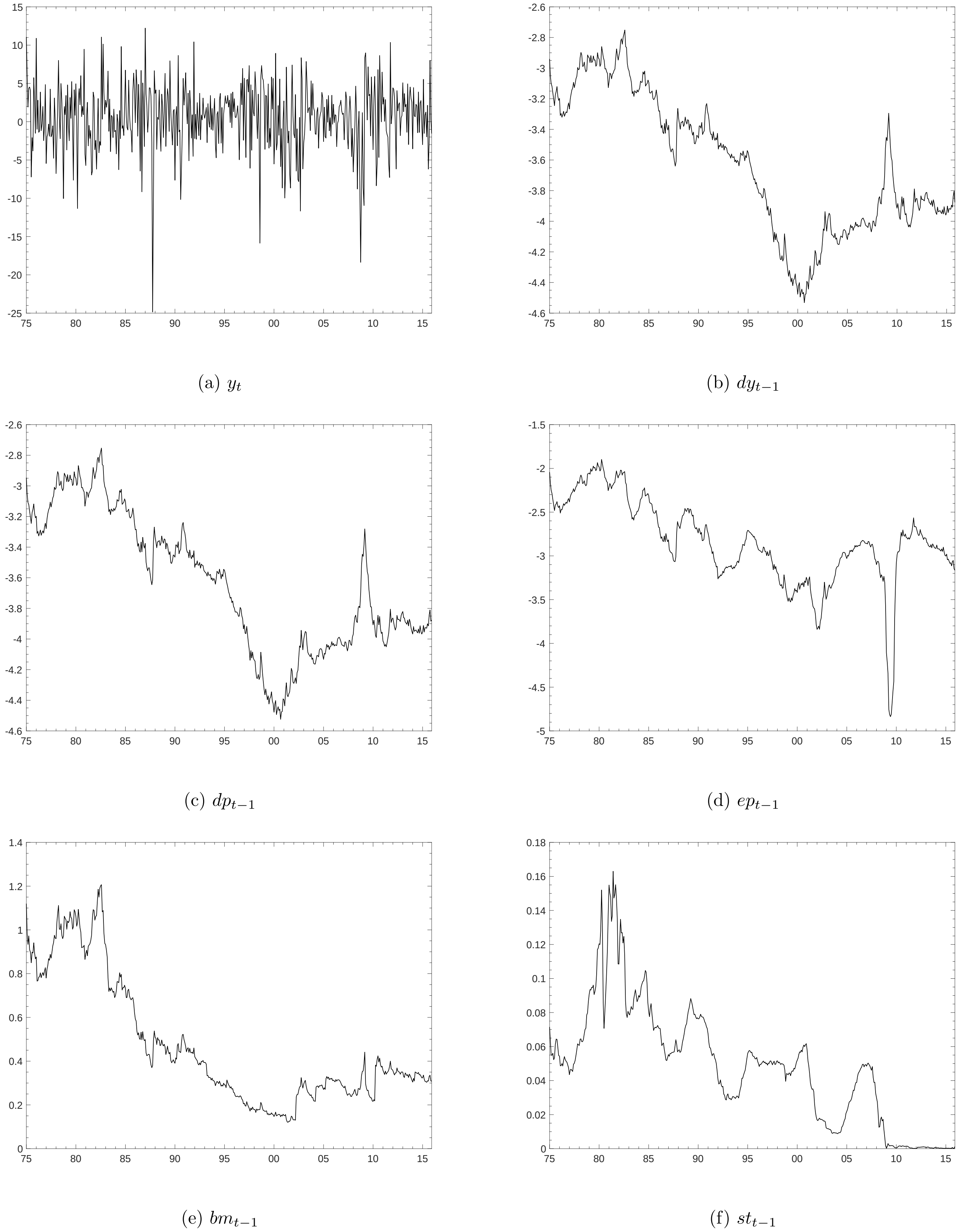
(b) 5 observations before the start of monitoring

(c) At the same time as the start of monitoring

(d) 5 observations after the start of monitoring

(e) 15 observations after the start of monitoring

Figure 9. Detection frequency for second predictability regime as a function of m_2 (predictability regime length) for different values of β_2: $T^* + m = 302$, $E = 362$, $m = 30$, $\rho = 0.995$



(a) y_t

(b) dy_{t-1}

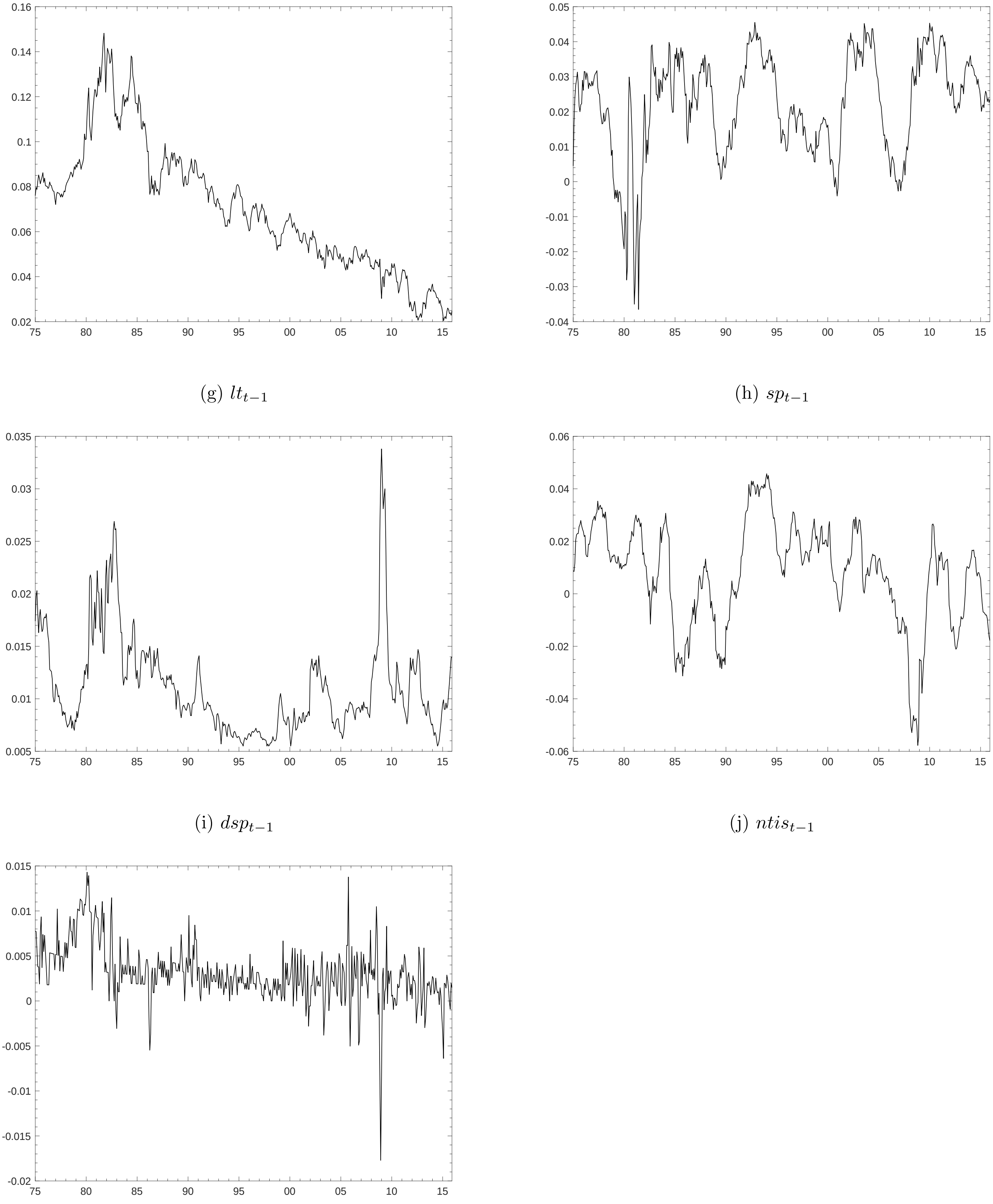
(c) dp_{t-1}

(d) ep_{t-1}

(e) bm_{t-1}

(f) st_{t-1}

Figure 10. Excess returns and MFVs



(g) lt_{t-1}

(h) sp_{t-1}

(i) dsp_{t-1}

(j) $ntis_{t-1}$

(k) inf_{t-1}

Figure 10 Continued. Excess returns and MFVs

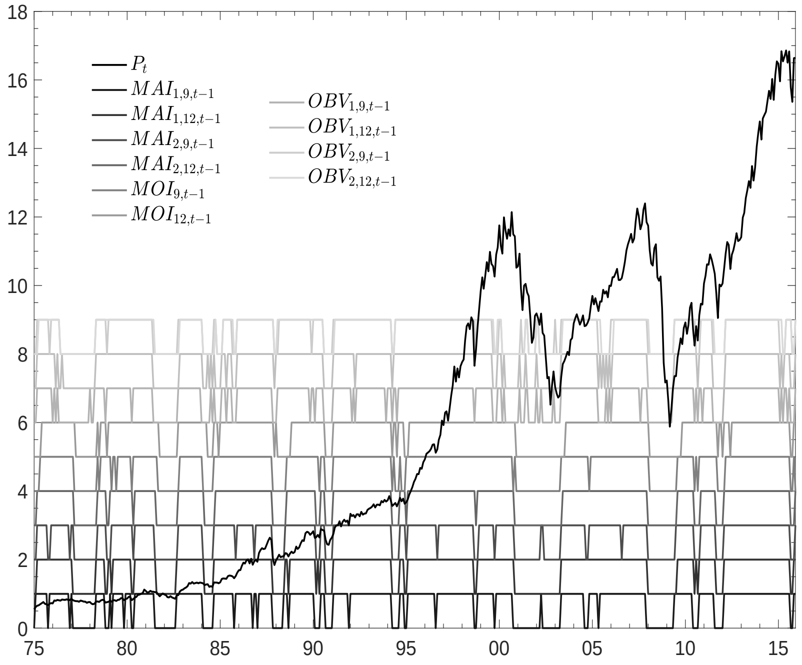
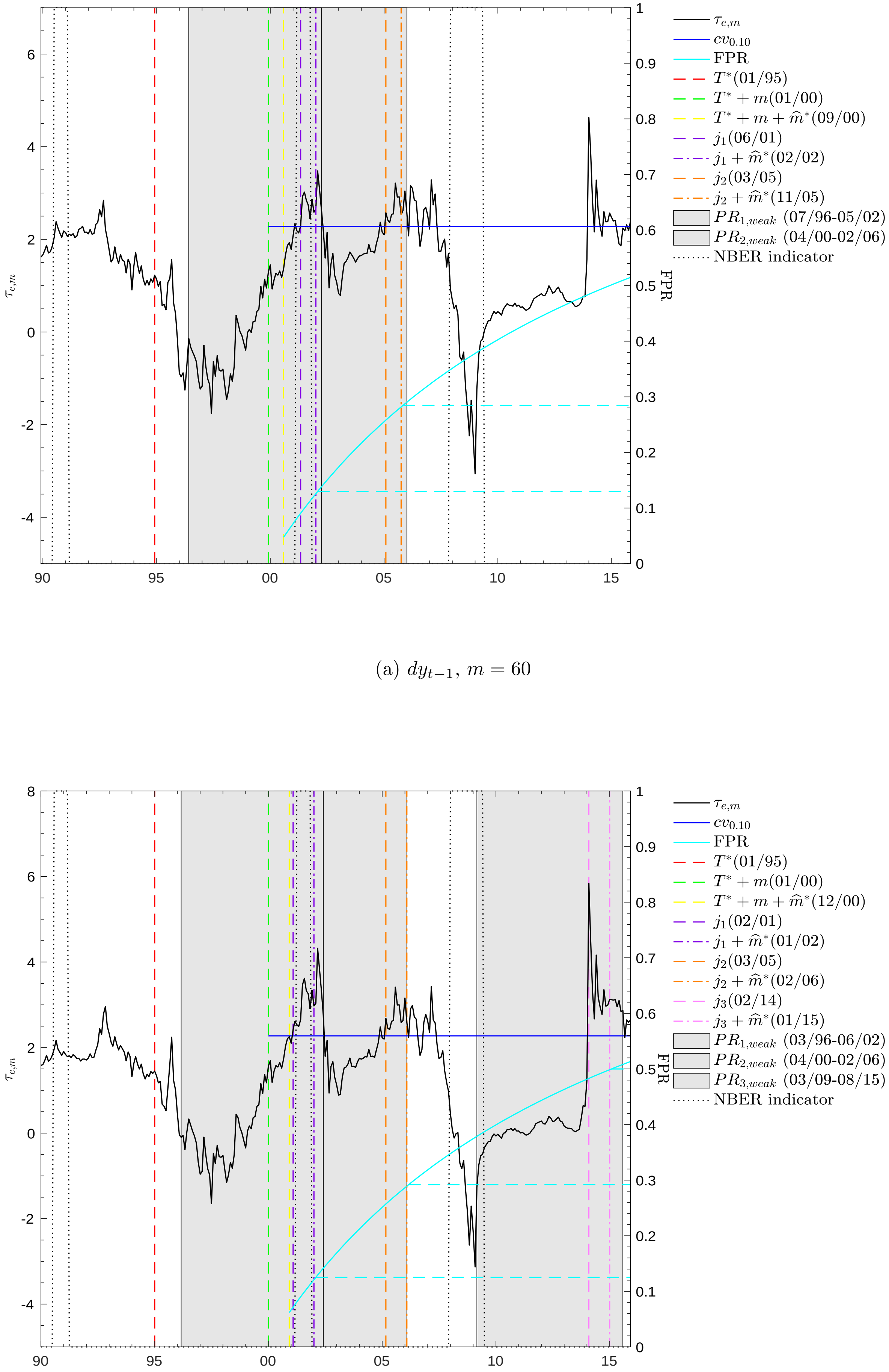


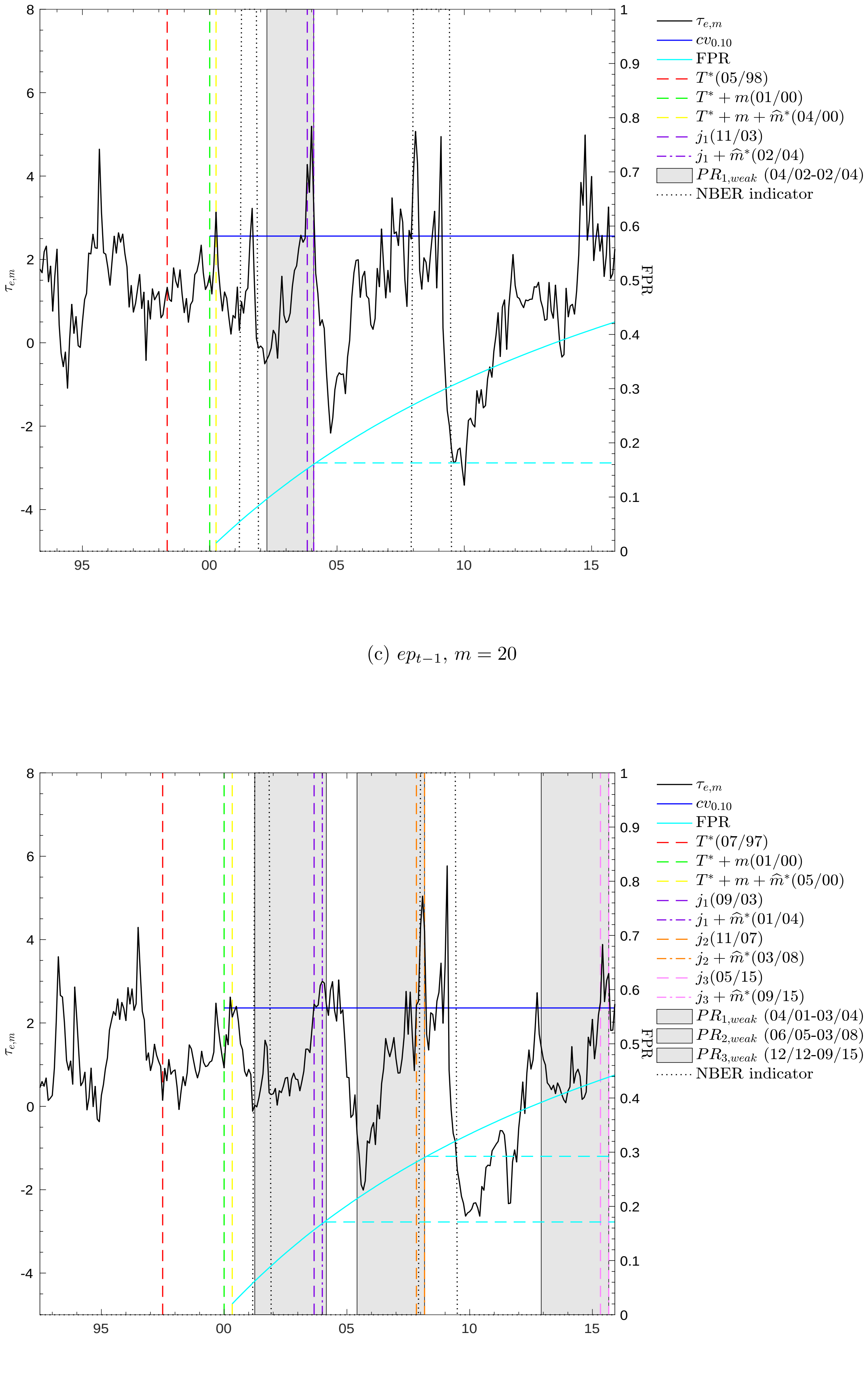
Figure 11. TAIs and the S&P Composite price index



(a) dy_{t-1}, $m = 60$

(b) dp_{t-1}, $m = 60$

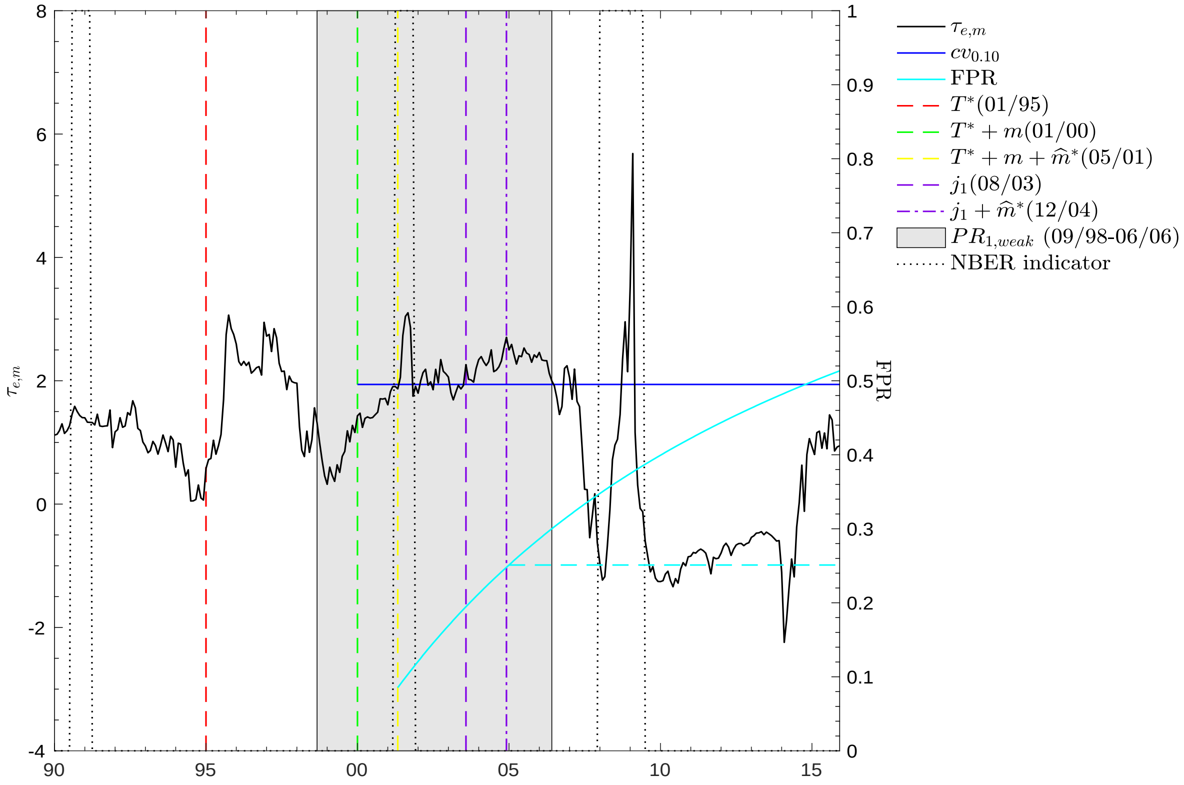
Figure 12. Monitoring results: MFVs



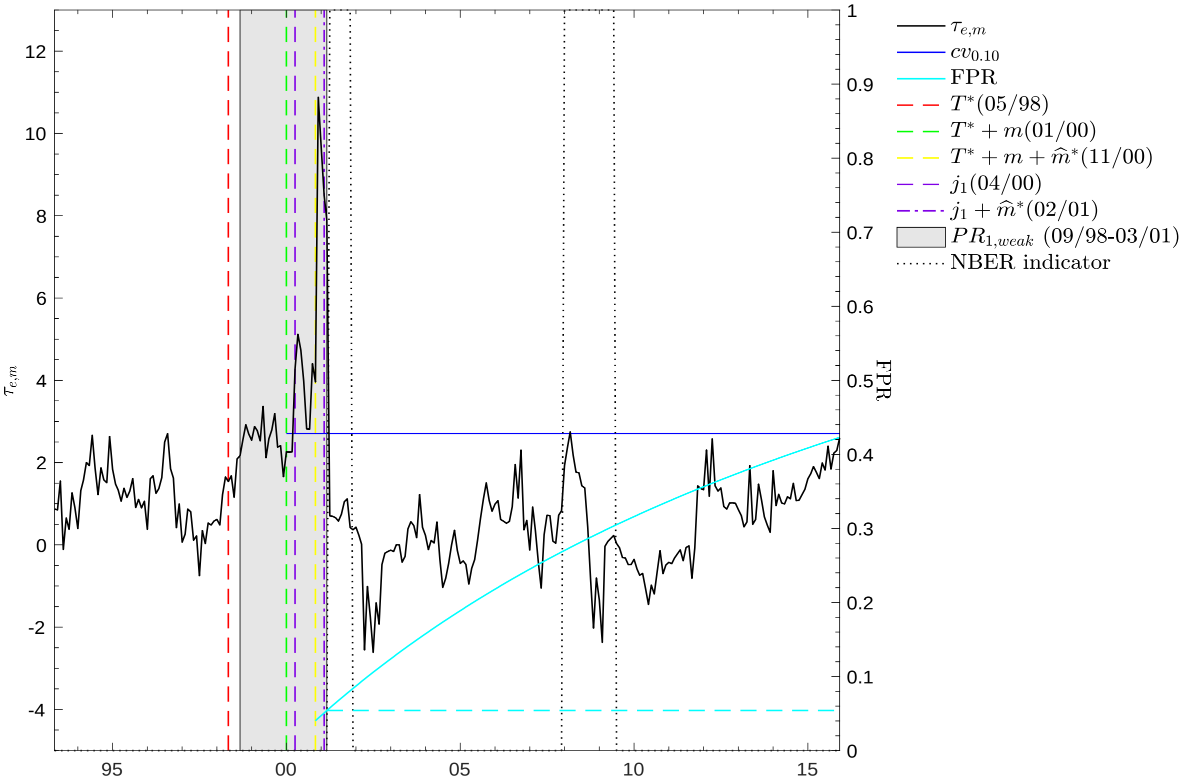
(c) ep_{t-1}, $m = 20$

(d) ep_{t-1}, $m = 30$

Figure 12 continued. Monitoring results: MFVs

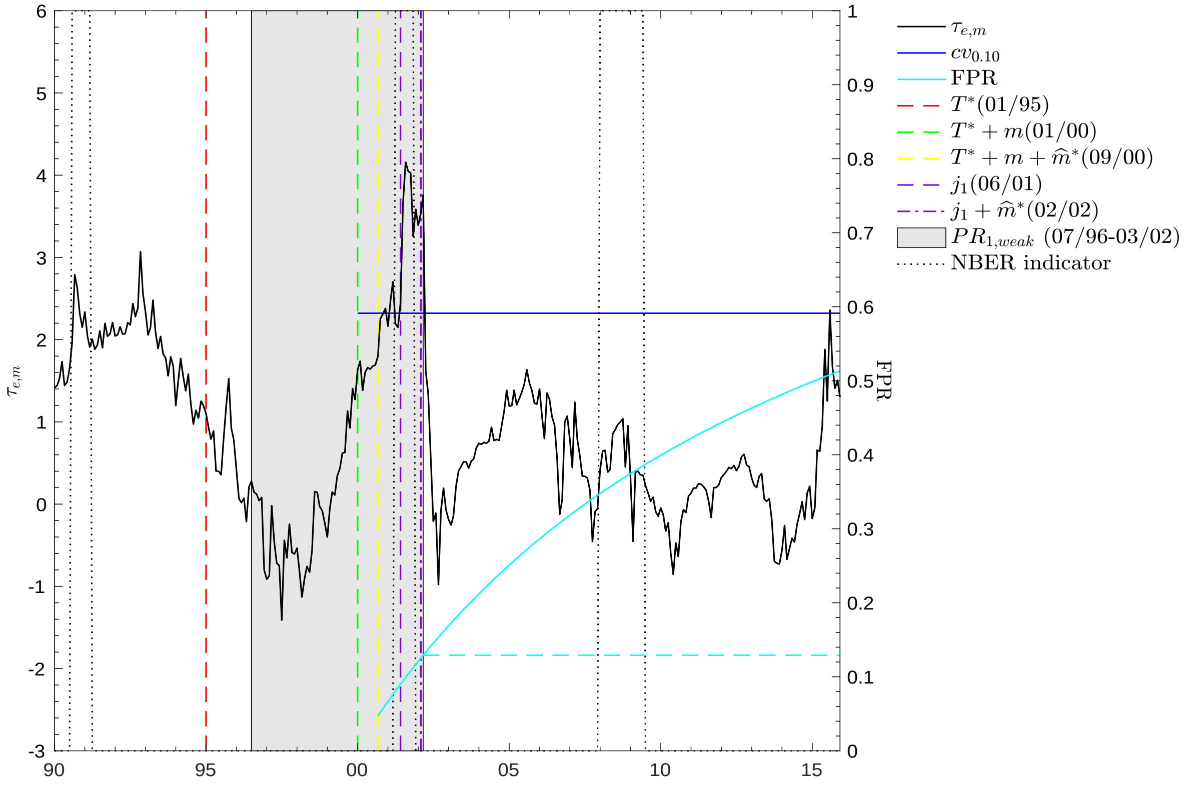


(e) ep_{t-1}, $m = 60$

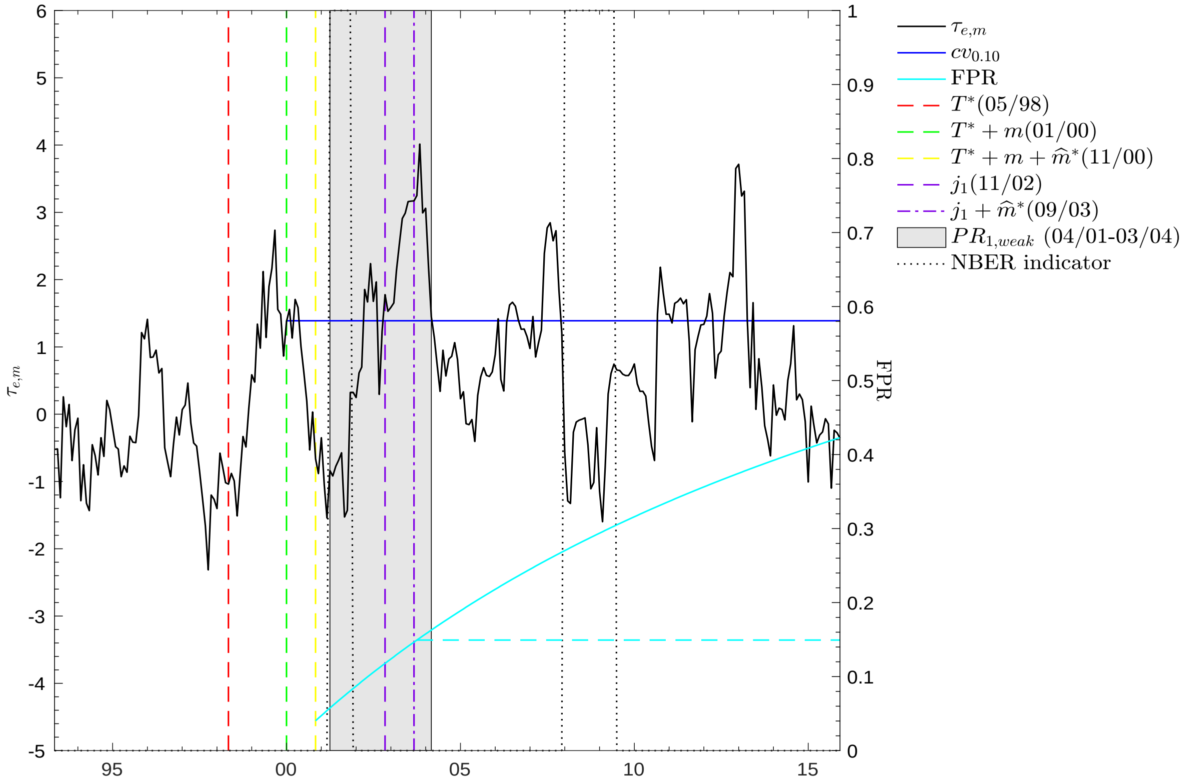


(f) bm_{t-1}, $m = 20$

Figure 12 continued. Monitoring results: MFVs

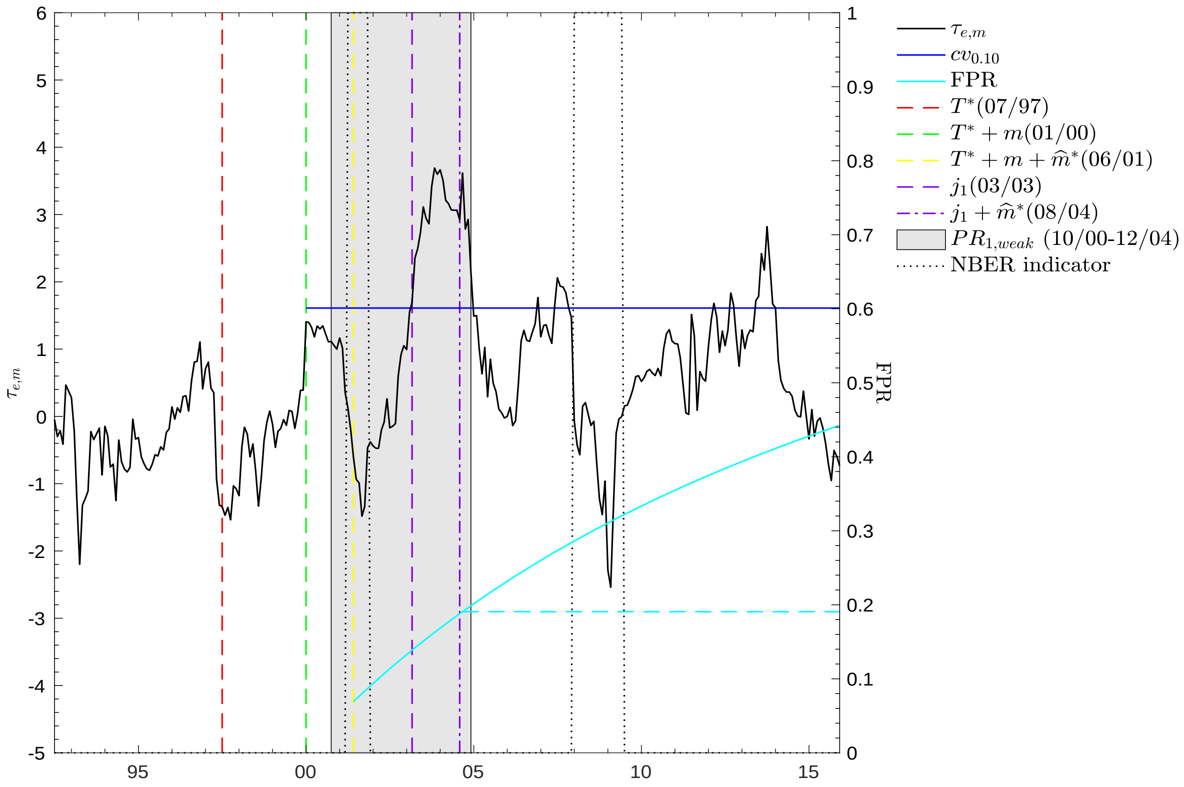


(g) bm_{t-1}, $m = 60$

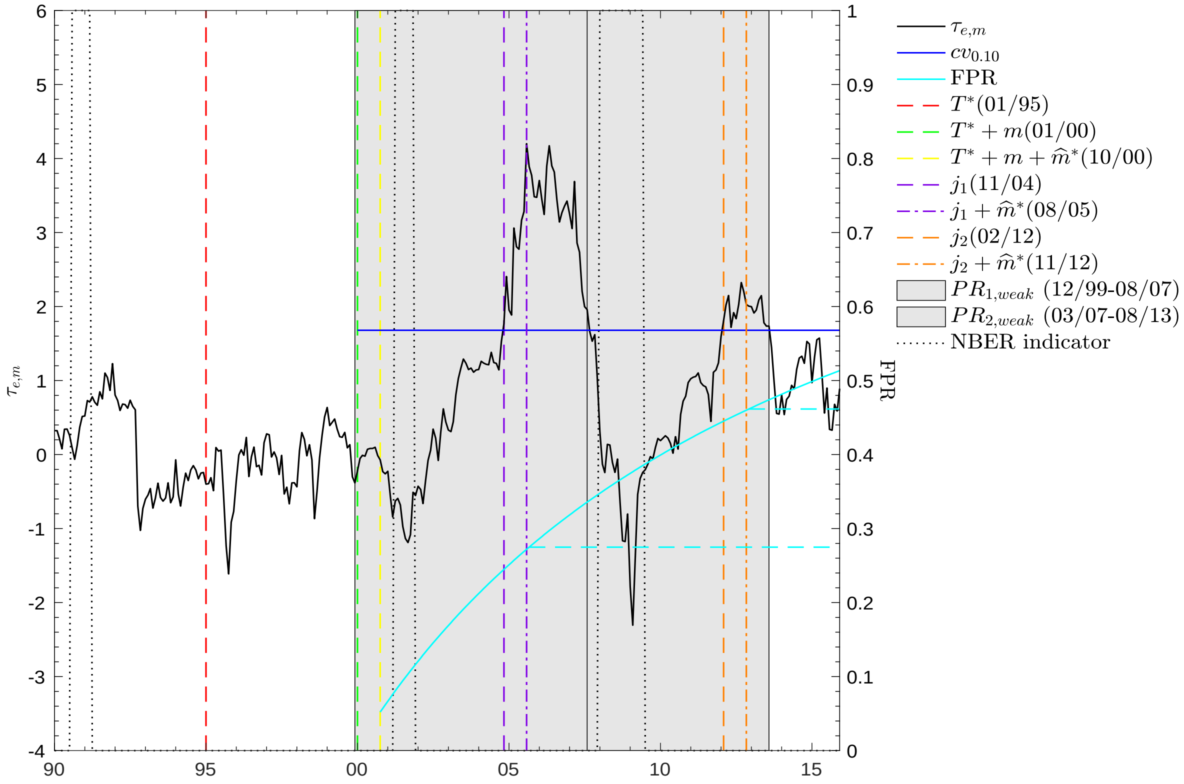


(h) lt_{t-1}, $m = 20$

Figure 12 continued. Monitoring results: MFVs

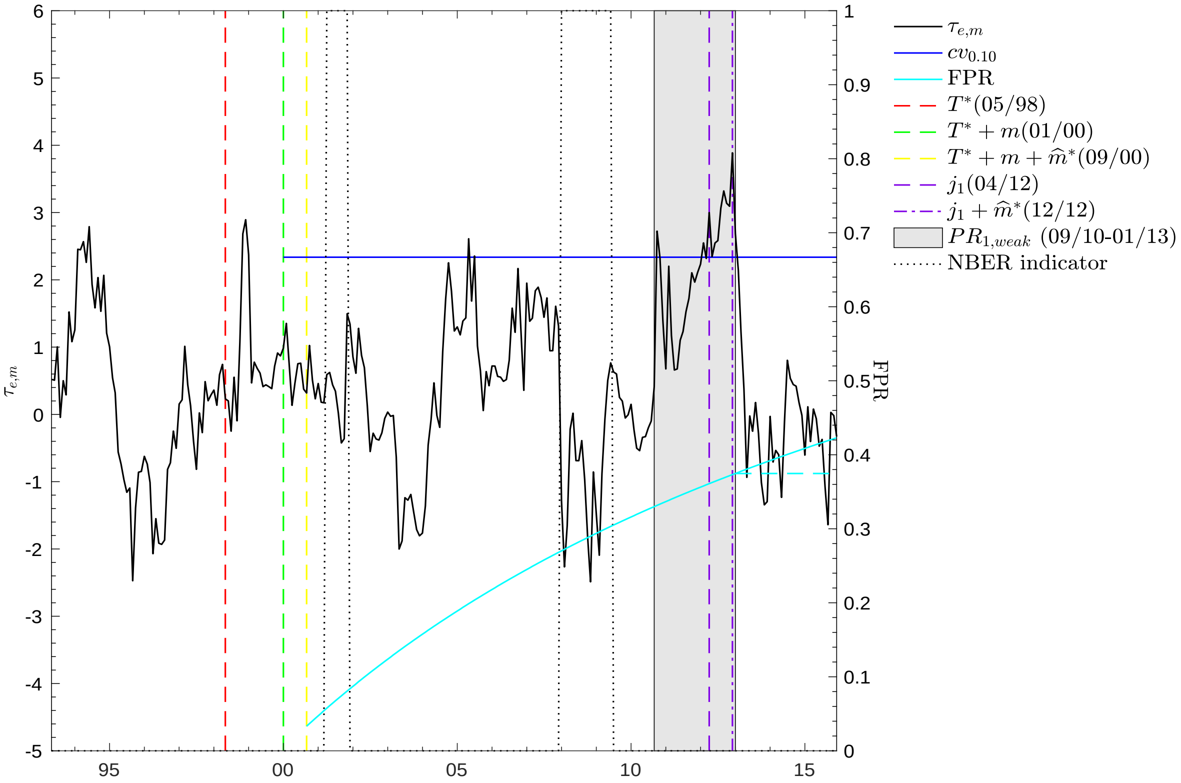


(i) lt_{t-1}, $m = 30$



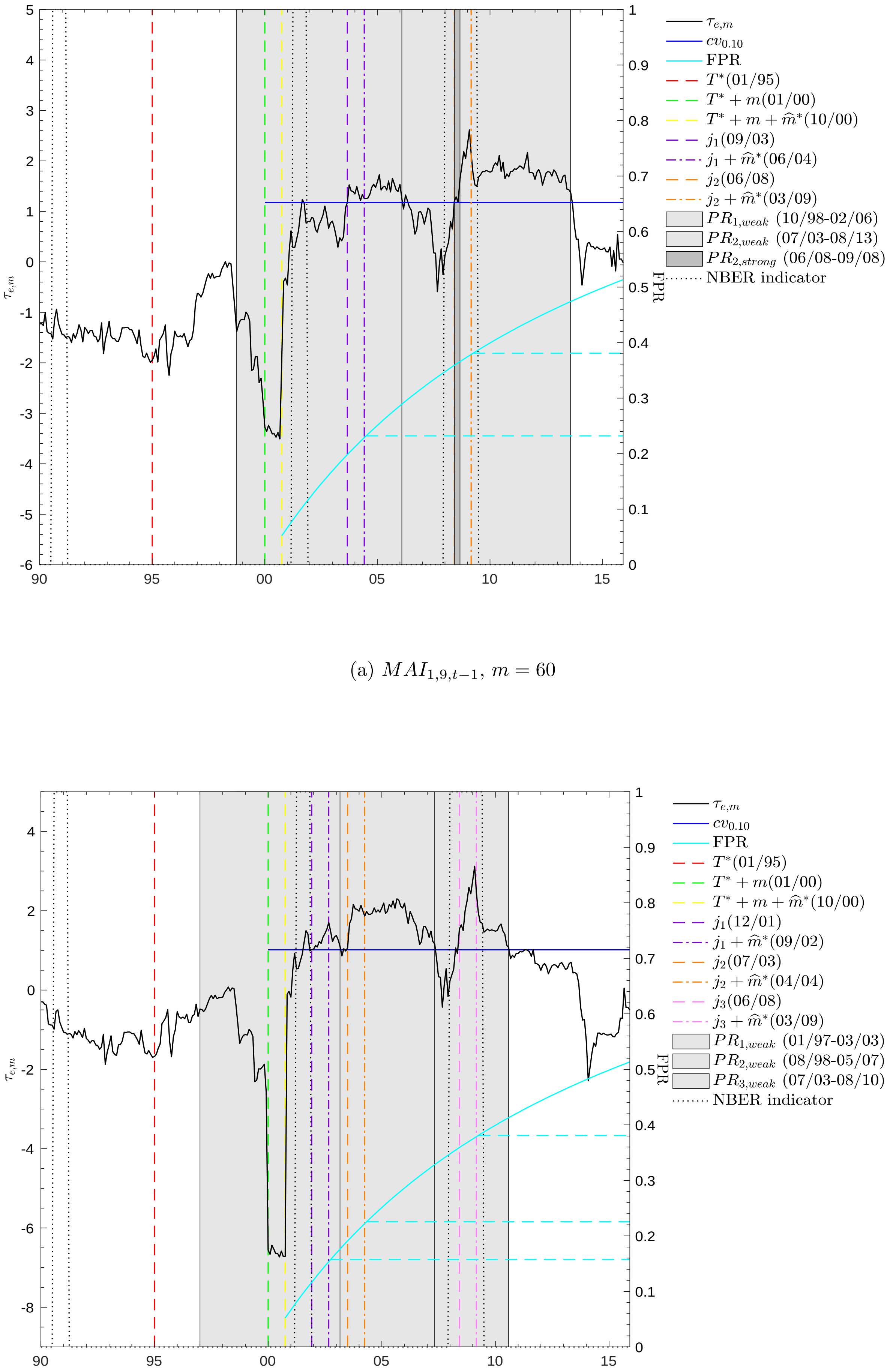
(j) lt_{t-1}, $m = 60$

Figure 12 continued. Monitoring results: MFVs



(k) dsp_{t-1}, $m = 20$

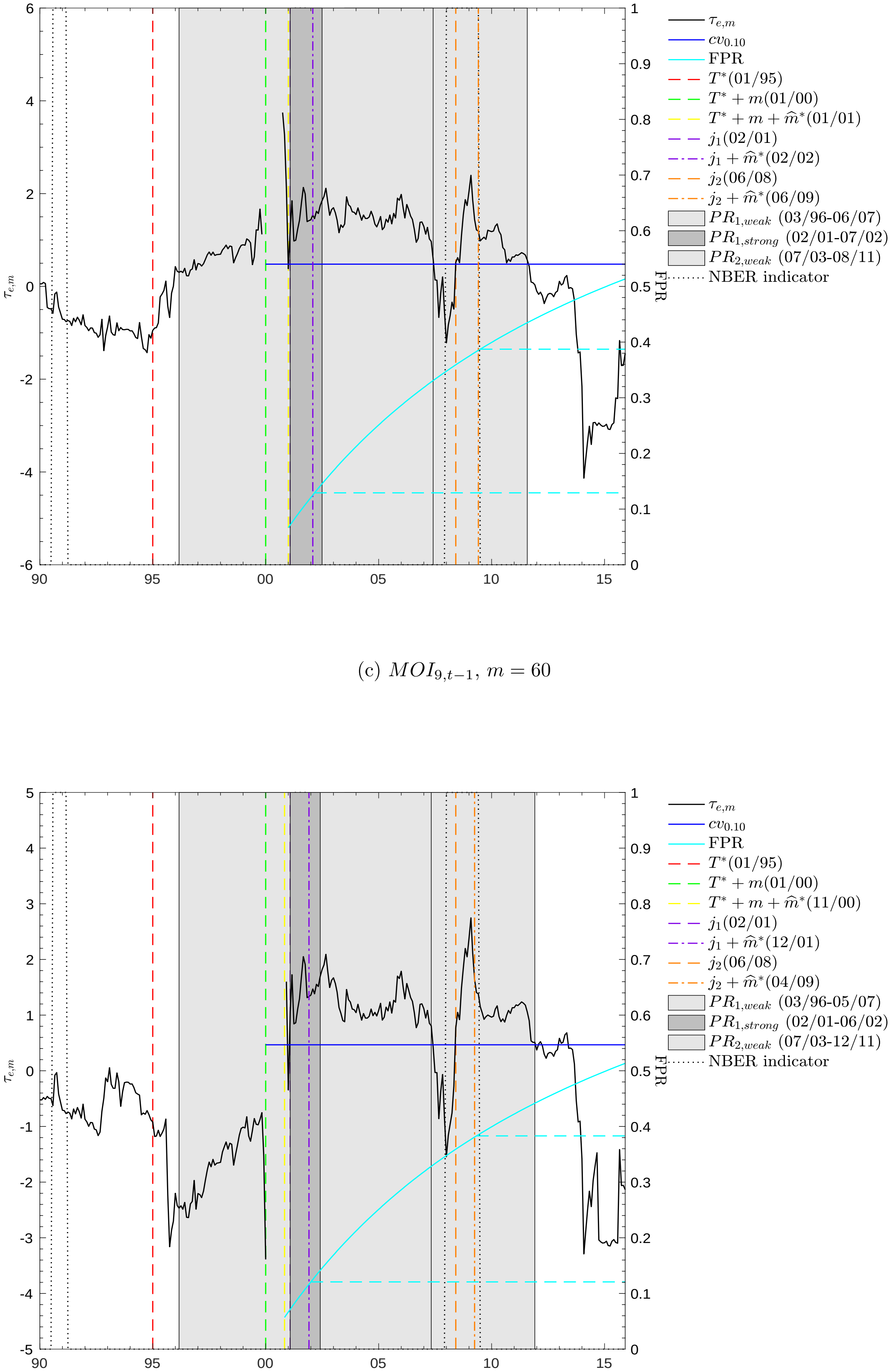
Figure 12 continued. Monitoring results: MFVs



(a) $MAI_{1,9,t-1}$, $m = 60$

(b) $MAI_{1,12,t-1}$, $m = 60$

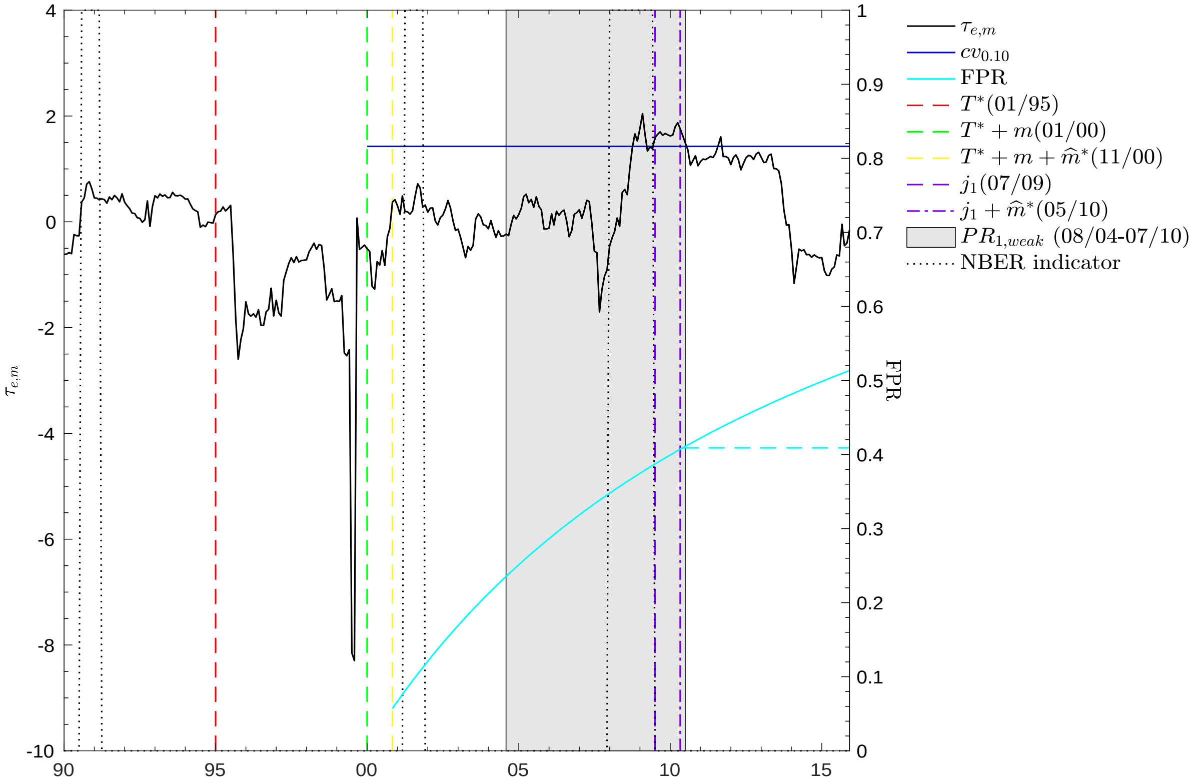
Figure 13. Monitoring results: TAIs



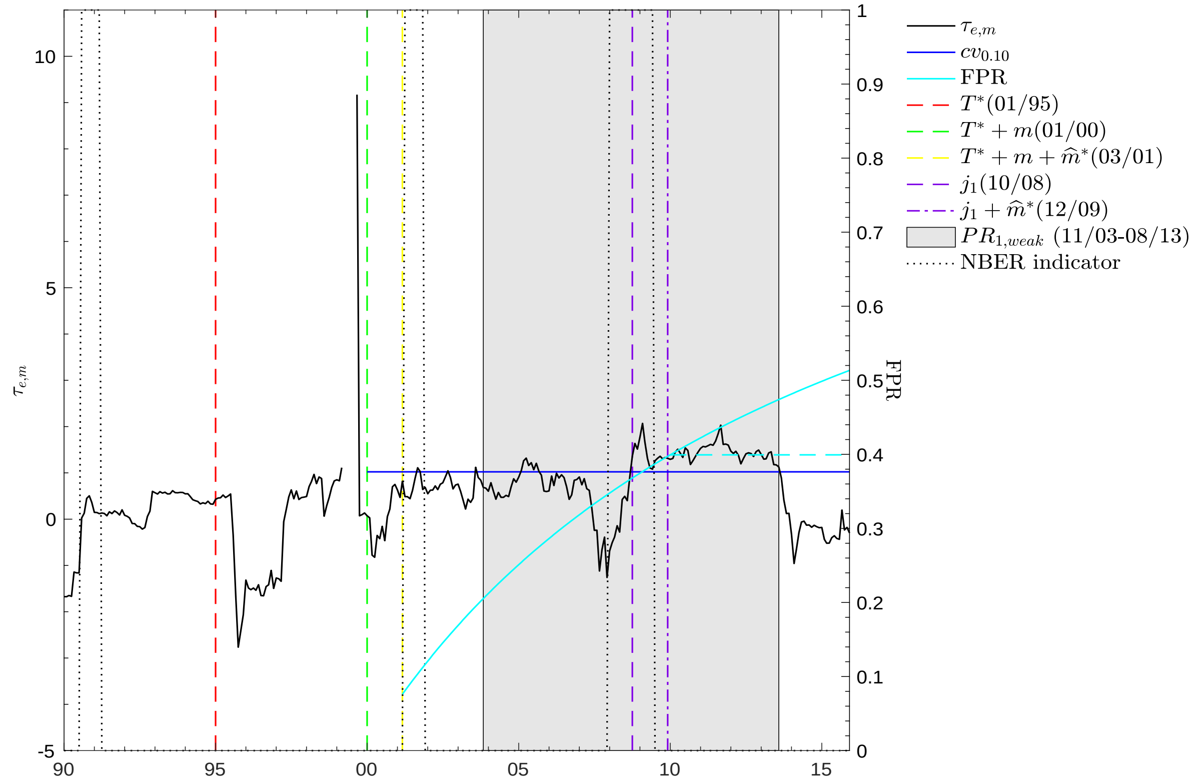
(c) $MOI_{9,t-1}$, $m = 60$

(d) $MOI_{12,t-1}$, $m = 60$

Figure 13 continued. Monitoring results: TAIs



(e) $OBV_{1,9,t-1}$, $m = 60$



(f) $OBV_{1,12,t-1}$, $m = 60$

Figure 13 continued. Monitoring results: TAIs