Detecting Regimes of Predictability in the U.S. Equity Premium

Robert Taylor

University of Essex

Co-authors:
David Harvey (University of Nottingham)
Stephen Leybourne (University of Nottingham)
Robert Sollis (Newcastle University)

This research is funded by the ESRC under award number ES/R00496X/1.

March 2019

Introduction

- Predictive regressions play an important role in empirical economics. In financial economics it is of interest whether current information on variables such as dividend yields or interest spreads contain information about future (excess) stock price returns (e.g. Campbell and Shiller, 1988, JF).
- An important practical problem with performing such predictive regressions with financial applications is that in many cases the regressor is highly persistent, whereas the dependent variable is close to white noise. For example, stock price returns or exchange rate changes appear to be approximately white noise, whereas predictors like dividend yields or interest rate differentials exhibit persistence behaviour akin to that of a unit root or near unit root autoregressive process.
- As shown by Elliott and Stock (1994,ET), the conventional t-statistic in the predictive regression can suffer from severe size distortions in such cases.

Introduction

- Consider testing $H_0 : \beta = 0$ (i.e. y_t unpredictable by x_{t-1}) in the predictive regression

$$y_t = \alpha + \beta x_{t-1} + \epsilon_t, \ \ t = 1, ..., T$$

where y_t is local-to-white noise (e.g. returns) and x_t is local-to-unit root (e.g. dividend yield).

- A number of papers have focused on developing asymptotically valid tests of this hypothesis, allowing for an unknown local-to-unity parameter in x_t and unknown correlation between ϵ_t and the innovations to the x_t process, e.g.:
  - Cavanagh *et al.* (1995,ET) (Bonferroni bounds that yield conservative tests)
  - Campbell and Yogo (2006,JFE) (point optimal t-test)
  - Breitung and Demetrescu (2015,JoE) (variable addition and IV)
  - Kostakis, Magdalinos and Stamatogiannis (2015,Review of Financial Studies) (IVX).

Introduction

- These methods are designed to test the null of predictability against the alternative that x_{t-1} is predictive for y_t over the whole sample.
- However, if stock returns are predictable, then it seems likely it would be a time-varying phenomenon; eg, 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 supportive of this view. Eg, 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.

Introduction

- Timmermann (2008) argues stock returns are not predictable for most time points but that there are ‘pockets in time’ where evidence of local predictability is seen.
- 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 low power to detect these short-lived predictive episodes.

- 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 Bai-Perron structural break tests. 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.

Introduction

- However, the statistical methods used by these authors are based on methods not designed for use with highly persistent, endogenous predictors.
- Moreover, 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.

Introduction

- In this paper we develop new methods to investigate the stability of predictive regression models for the U.S. equity premium. As putative predictors we consider various commonly used traditional macroeconomic and financial variables, and a range of technical analysis rules where only price or volume data is used to predict returns.
- The methods we develop are designed to detect relatively short windows of predictability arising from shifts in the parameter on the predictor variable in the predictive regression. 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).

Introduction

- 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.

Introduction

- 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.
- 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.

Introduction

- 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 develop 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 Predictive Regression Model

- We assume a relationship between the equity premium, y_t, and a single predictor variable x_t described by the DGP

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

- The (putative) predictor is generated by

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

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 Predictive Regression Model

- The innovation vector $\epsilon_t := [\epsilon_{y,t}, \epsilon_{x,t}]'$ is assumed to be an MDS with finite fourth order moments and unconditional covariance matrix given by

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

  where $|r_{xy}| < 1$.
- This allows unconditional heteroskedasticity in $\epsilon_{y,t}$ and/or $\epsilon_{x,t}$ 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 $\epsilon_{x,t}$.

The Predictive Regression Model

- 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.

The Predictive Regression Model

- If $\beta_j \neq 0$ in (1), 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 $t = e_j$. (1) allows for $n \geq 0$ such predictive regimes.
- 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 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.
- 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.

Background Results for Predictive Regime Detection

- To motivate, suppose we have a sample of time series observations z_t, $t = 1, ..., N$ from a *stationary* continuous distribution. Consider $\max_{t \in [1,N]} z_t$; its *value* is clearly a function of the distribution of z_t.
- But, consider the *location* at which $\max_{t \in [1,N]} z_t$ occurs, $M := \arg\max_{t \in [1,N]} z_t$. As 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,

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

Background Results for Predictive Regime Detection

- Consider the maximum value of z_t in each of the two intervals $t = 1, ..., \lfloor (1-\alpha)N \rfloor$ and $t = \lfloor (1-\alpha)N \rfloor + 1, ..., N$; ie, $\max_{t \in [1, \lfloor (1-\alpha)N \rfloor]} z_t$ 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,

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

  due to the large sample uniformity of the location of the maximum.
- 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 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 ($\lfloor (1-\alpha)N \rfloor + N - \lfloor (1-\alpha)N \rfloor = N$); ie,

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

Background Results for Predictive Regime Detection

- Consider next the maximum number of *contiguous* values of z_t that exceed some threshold value, c say, where we assume that c is such that $0 < \Pr(z_t > c) < 1$.
- 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.

Background Results for Predictive Regime Detection

- 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.

Subsample Regression t-statistics

- Consider selecting a subsample of m observations running from $t = e - m + 1$ to $t = e$, m chosen by the practitioner, and run the (generic) OLS regression,

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

- Calculate the heteroskedasticity-robust regression t-statistic for the significance of x_{t-1} in (7),

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

where

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

Subsample Regression t-statistics

- Given an appropriate critical value, a test for a predictive regime holding between y_t and x_{t-1} for the subsample $t = e - m + 1, ..., e$ can be based on $\tau_{e,m}$. Eg, with data for $t = 1, ..., T^* + m$ a test for a predictive regime in the last m sample observations would 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 based on the assumption that m is a fraction of the sample size, T. This assumption is clearly not consistent with the aim to detect predictive regimes of short duration.
- Even if we were to assume m to be a function of the sample size T, the limiting distribution of $\tau_{e,m}$ will depend on nuisance parameters present in the DGP; in particular, the degree of persistence of x_t and the correlation, r_{xy}, between $\epsilon_{y,t}$ and $\epsilon_{x,t}$. Without knowledge of these nuisance parameters, valid asymptotic critical values for the test could not be obtained in any case.

Subsample Regression t-statistics

- We will use an alternative approach 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, this approach involves comparing $\tau_{e,m}$ with critical values obtained by subsampling from a *training period* of the data not used in calculating $\tau_{e,m}$.
- This approach delivers tests which, by design, are robust to the nuisance parameters in the DGP discussed above. 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.

Subsample Regression t-statistics

- Importantly, if no predictability holds in the training period, 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 $e = T^* + m, ..., E$, where $E \leq T$ is used to denote 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.

The Detection Procedure

- The first step is to determine a critical value to use with the sequence of $\tau_{e,m}$, $e = T^* + m, ..., E$, statistics from an initial training period, $t = 1, ..., T^*$ where $T^* := \lfloor \lambda T \rfloor$ for some $\lambda \in (0, 1)$, and where $T^* < e_1 - m_1 + 1$.
- Using the sequence of $\tau_{e,m}$ statistics that make use of data within this training period, i.e. $\tau_{e,m}$ for $e = m + 1, ..., T^*$, calculate an empirical critical value for a significance level π, say. Denote this empirical critical value by cv_π.
- 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 $T \to \infty$.

The Detection Procedure

- Next, start the monitoring period by calculating the first statistic $\tau_{e,m}$ which does not use any of the training period data; ie, $\tau_{T^*+m,m}$ (which uses data from $t = T^* + 1$ to $t = T^* + m$), and compare this with the training period critical value cv_π.
- Then, move forwards one period, calculating $\tau_{T^*+m+1,m}$, again comparing the statistic with cv_π.
- Proceed sequentially in this manner, comparing $\tau_{e,m}$, $e = T^* + m, T^* + m + 1, ...$, with cv_π as we move forwards in time.
- Use $R_e := 1(\tau_{e,m} > cv_\pi)$ to record whether or not each test statistic in the sequences exceeds the critical value.

The Detection Procedure

- 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. 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 $e = U$. We then determine that a predictive regime is present when $R(L,U) > m^*$ for some choice of $m^* > 1$.

The Detection Procedure

- 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 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 $T^* + m, ..., E$.
- We next discuss a data-based method to choose m^* and how it relates to the associated FPR of the monitoring procedure.

Choice of m^* and the False Positive Detection Rate

- Now consider the FPR of our proposed procedure; that is, the probability of incorrectly identifying at least one predictive regime when in fact none exists, when 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

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

where adapting to the notation of (6),

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

and $N \equiv T^* - m + E - (T^* + m) + 1 = E - 2m + 1$.

Choice of m^* and the False Positive Detection Rate

- 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

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

- 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.

Choice of m^* and the False Positive Detection Rate

- Here, α is a monotonically increasing function of E since

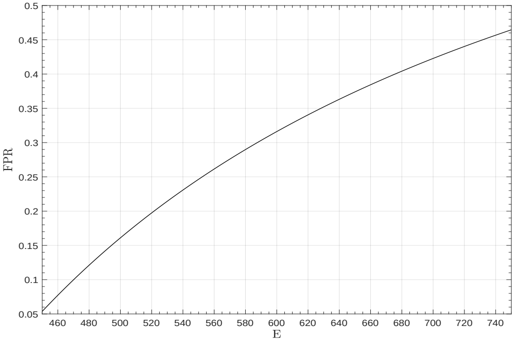
$$\frac{\partial \alpha}{\partial E} = \frac{T^* - m}{(E - 2m + 1)^2} > 0.$$

  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 $m^* = \widehat{m}^*$.

Choice of m^* and the False Positive Detection Rate

To illustrate, suppose $T^* = 400$ and $m = 30$, Figure 1 below shows this approximation to the FPR as a function of E.



So, eg, if we monitor out to $E = 680$, the FPR will be about 0.40.

Choice of m^* and the False Positive Detection Rate

- 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 this appears to relate directly to the choice of significance level π. 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.

Choice of m^* and the False Positive Detection Rate

- In fact any sensible threshold value could be used. 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.
- If one or more short duration predictive regimes are present in the training period although the large (in T) sample properties of the estimated critical value would be unaffected, for a given finite length training period we would expect both cv_π and $\widehat{m}^*$ to increase relative to the case where the training period was non-predictive. 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 this using simulations later.

Choice of m^* and the False Positive Detection Rate

- Thus far we assumed no separation between the training and monitoring periods, 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 largest possible training period. This also provides a way to vary the FPR of the procedure based on $\widehat{m}^*$, for a given E. The foregoing expressions can be easily modified for this.

Choice of m^* and the False Positive Detection Rate

- It is also possible 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 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.
- In practice, we recommend applying standard full-sample predictability tests to the training period to investigate whether the assumption of no predictability appears valid.

Dating of Predictive Regimes

- Our key aim here is one of detecting the emergence of a predictive regime in real-time. However, one could conceive of an historical exercise in data analysis where we look to date any predictive episodes identified within the available range of sample data.
- A liberal, or weak, start date for any predictive episode signalled by our procedure is the first data point in the sample used for the first statistic in this set of contiguous exceedances. A corresponding weak end date for the episode is the last sample observation in the last of these contiguous exceedances. It is therefore possible for separate identified predictive episodes to overlap for weak dates.
- We could also 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. Such *strong* dates cannot overlap for separate identified predictive episodes, and could be an empty set for a given identified predictive episode.

Finite Sample Properties of the Monitoring Procedure

- We present the results from 8 sets of MC simulation experiments based on the DGP

$$
\begin{aligned}
y_t &= \sum_{j=1}^{n} \beta_j d_t(e_j, m_j) x_{t-1} + \epsilon_{y,t}, \;\; t = 1, ..., T, \; \epsilon_{y,t} \sim NIID(0,1) \\
x_t &= \rho x_{t-1} + \epsilon_{x,t}, \;\; t = 1, ..., T, \; \epsilon_{x,t} \sim NIID(0,1)
\end{aligned}
$$

  with $x_0 = 0$ and $r_{x,y} = -0.90$.
- Upper tailed tests throughout. MATLAB, 5000 MC reps.
- Set $T = 493$ and $T^* + m = 302$ consistent with our empirical application. Set window width $m = 30$.
- The ranges of values for ρ and β_1 considered were chosen following a preliminary analysis of the data used in the empirical application, as was $r_{x,y}$.
- For all of the experiments we report, graphs will be used to show the empirical frequencies with which at least one PR was signalled by the procedure taken across the whole monitoring period.

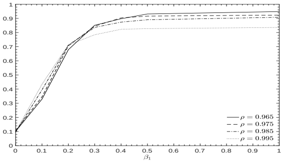
Finite Sample Properties of the Monitoring Procedure

- The first set of experiments study "power" to detect a single predictive regime [PR] 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 $\pi = 0.10$.
- First we assume a short monitoring period that ends at $E = 328$, consistent with an FPR at time E of $\alpha = 0.10$.
- NB if $\beta_1 = 0 \Rightarrow n = 0$ and there are no PRs, the detection frequency obtained is an empirical FPR at time E.
- If a PR occurs in the monitoring period, detection power depends on the length of the regime (m_1), its strength (β_1), and on when the PR starts relative to the start of monitoring. For the latter we considered five different PR start dates: (a) $t = 287$ (15 observations before the start of monitoring), (b) $t = 297$ (5 obs before the start of monitoring), (c) $t = 302$ (at the same time as the start of monitoring), (d) $t = 307$ (5 obs after the start of monitoring), (e) $t = 317$ (15 obs after the start of monitoring). In each case $m_1 = 30$.

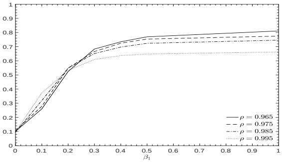
Finite Sample Properties of the Monitoring Procedure

- The "power" curves indeed all start from approximately 0.10, consistent with the theoretical FPR at the end of the monitoring period of 0.10.
- For cases (a)-(c) when the PR 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 PR 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.

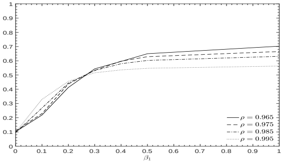
PR detection frequency for different values of ρ: $T = 493$, $T^* + m = 302$, $E = 328$, $m_1 = 30$, $m = 30$



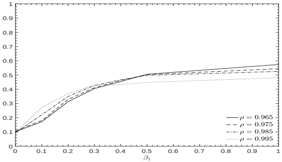
(a) 15 observations before the start of monitoring



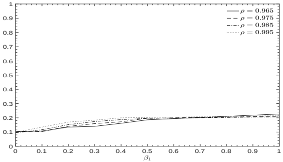
(b) 5 observations before the start of monitoring



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

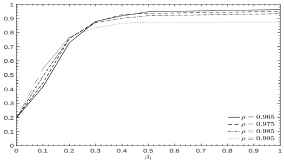


(d) 5 observations after the start of monitoring

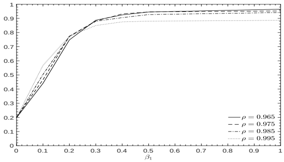


(e) 15 observations after the start of monitoring

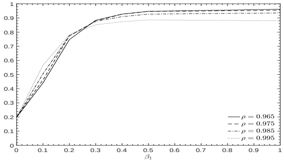
PR detection frequency for different values of ρ: $T = 493$, $T^* + m = 302$, $E = 362$, $m_1 = 30$, $m = 30$



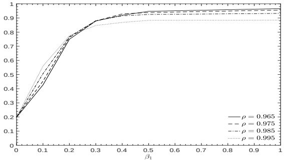
(a) 15 observations before the start of monitoring



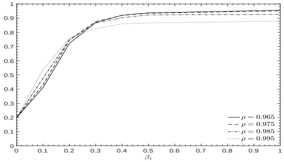
(b) 5 observations before the start of monitoring



(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

Finite Sample Properties of the Monitoring Procedure

- In empirical applications, whilst there might be a particular reason for favouring a short monitoring period, for PRs that happen to 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 predictive regime dates, but extending the monitoring period to $E = 362$ which is consistent with a FPR at the end of the monitoring period of $\alpha = 0.20$.

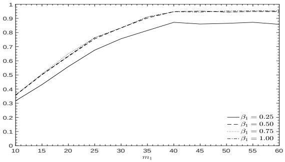
Finite Sample Properties of the Monitoring Procedure

- As expected, when the monitoring period is extended to $E = 362$, PR detection frequency as a function of β_1 increases.
- Indeed the results are now virtually identical for each of the PR start dates considered here and all of the curves flatten out quickly as β_1 increases.
- This is because due to 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 at the end of monitoring increases to approximately 0.20, again as expected.
- Notice, comparing (a) for $E = 328$ and $E = 362$, that except for very small values of β_1, "power" is almost unrelated to the FPR.

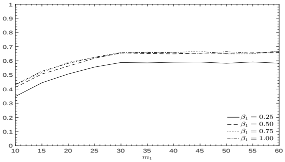
Finite Sample Properties of the Monitoring Procedure

- The third set of experiments study detection power as a function of its length, m_1, for $E = 328$ and for the PR dates used in the previous cases. We consider $m_1 = \{10, 15, ..., 60\}$. We set $\rho = 0.995$ and consider $\beta_1 = \{0.25, 0.50, 0.75, 1.00\}$.
- In the fourth set of experiments we repeat the third set, but extending the monitoring period to $E = 362$.

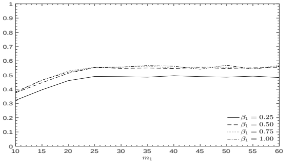
PR detection frequency as a function of m_1: $T^* + m = 302$, $E = 328$, $m = 30$, $\rho = 0.995$



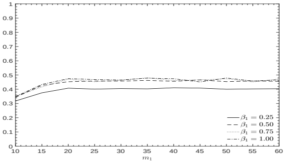
(a) 15 observations before the start of monitoring



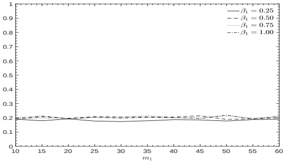
(b) 5 observations before the start of monitoring



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

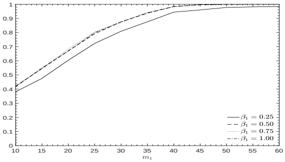


(d) 5 observations after the start of monitoring

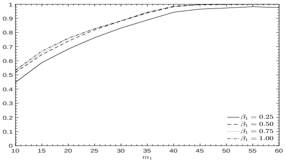


(e) 15 observations after the start of monitoring

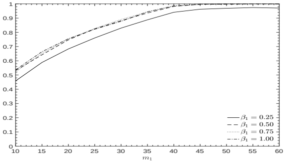
PR detection frequency as a function of m_1: $T^* + m = 302$, $E = 362$, $m = 30$, $\rho = 0.995$



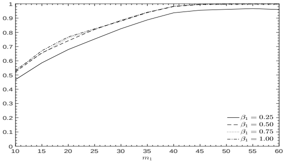
(a) 15 observations before the start of monitoring



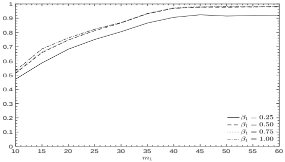
(b) 5 observations before the start of monitoring



(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

Finite Sample Properties of the Monitoring Procedure

$E = 328$

- For $E = 328$ power is seen to initially increase with m_1. 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), the PR starts progressively later and so the value of m_1 such that the end of the PR lies beyond 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 PR. Therefore, in this case, further increases in m_1 above 10 do not lead to any further increase in the power.

Finite Sample Properties of the Monitoring Procedure

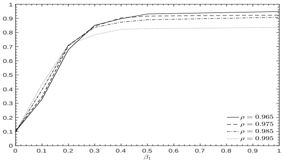
$E = 362$

- For $E = 362$ power, as function of m_1, is now very similar, irrespective of when the PR 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 $m_1 \geq 40$.

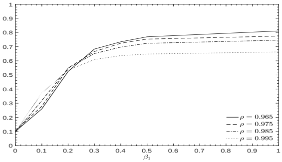
Finite Sample Properties of the Monitoring Procedure

- The fifth through eighth sets of experiments repeat the first four sets but with a PR in the training period at time $t = \lfloor T^*/2 \rfloor$ with $m_1 = 15$ and $\beta_1 = 0.25$. Hence, $n = 2$ in the DGP.
- This first PR is relatively short as these are more difficult to identify. A long PR in the training period would likely be detectable and could then be corrected for.
- The value of $\widehat{m}^*$ selected for monitoring, was on average found to be 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 in the first four sets of experiments. Generally around 5-10% drops for $E = 328$ but much less for $E = 362$.
- 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.

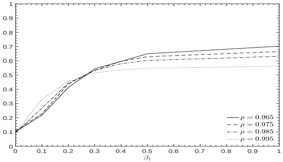
PR detection frequency for different values of ρ: $T = 493$, $T^* + m = 302$, $E = 328$, $m_1 = 30$, $m = 30$



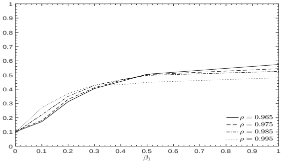
(a) 15 observations before the start of monitoring



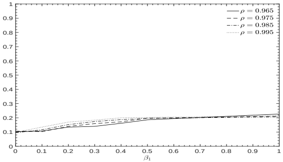
(b) 5 observations before the start of monitoring



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

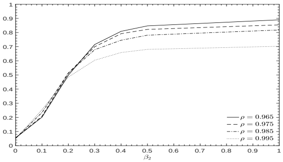


(d) 5 observations after the start of monitoring

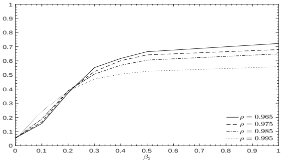


(e) 15 observations after the start of monitoring

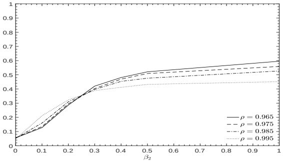
PR detection frequency for different values of ρ: $T = 493$, $T^* + m = 302$, $E = 328$, $m_1 = 15$, $m_2 = 30$, $m = 30$



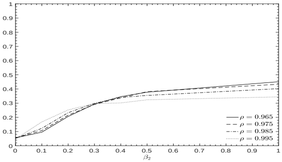
(a) 15 observations before the start of monitoring



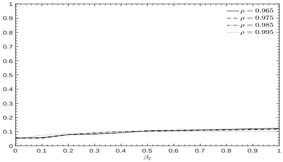
(b) 5 observations before the start of monitoring



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

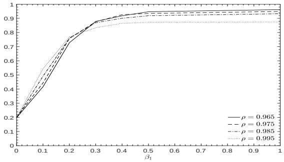


(d) 5 observations after the start of monitoring

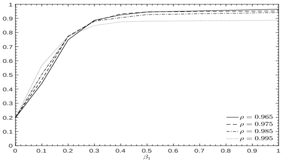


(e) 15 observations after the start of monitoring

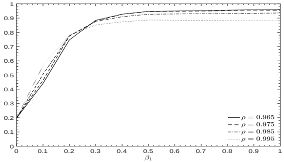
PR detection frequency for different values of ρ: $T = 493$, $T^* + m = 302$, $E = 362$, $m_1 = 30$, $m = 30$



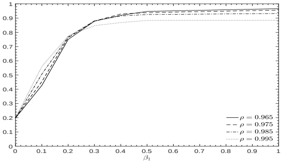
(a) 15 observations before the start of monitoring



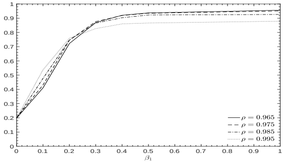
(b) 5 observations before the start of monitoring



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

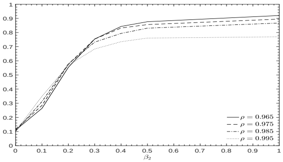


(d) 5 observations after the start of monitoring

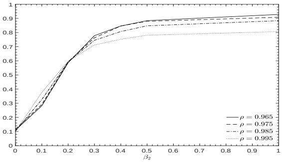


(e) 15 observations after the start of monitoring

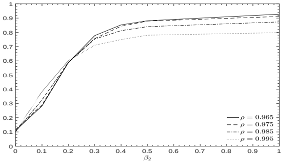
PR detection frequency for different values of ρ: $T = 493$, $T^* + m = 302$, $E = 362$, $m_1 = 15$, $m_1 = 30$, $m = 30$



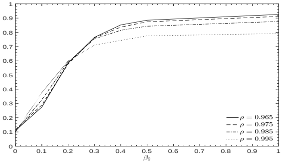
(a) 15 observations before the start of monitoring



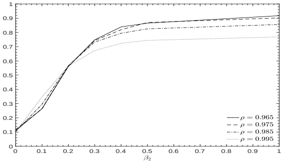
(b) 5 observations before the start of monitoring



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

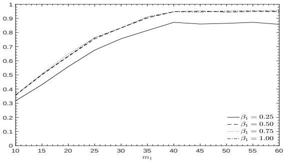


(d) 5 observations after the start of monitoring

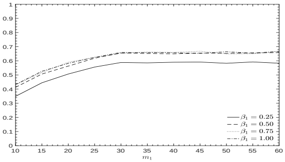


(e) 15 observations after the start of monitoring

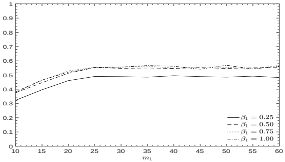
PR detection frequency as a function of m_1: $T^* + m = 302$, $E = 328$, $m = 30$, $\rho = 0.995$



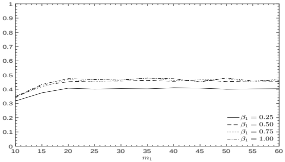
(a) 15 observations before the start of monitoring



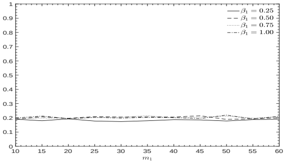
(b) 5 observations before the start of monitoring



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

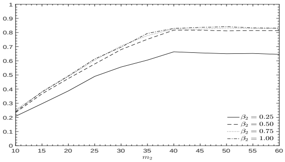


(d) 5 observations after the start of monitoring

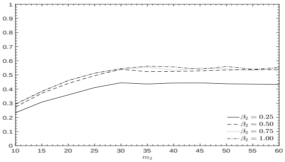


(e) 15 observations after the start of monitoring

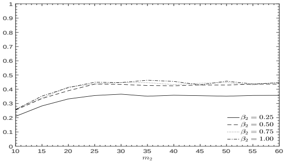
PR detection frequency as a function of m_2: $T^* + m = 302$, $E = 328$, $m_1 = 15$, $m = 30$, $\rho = 0.995$



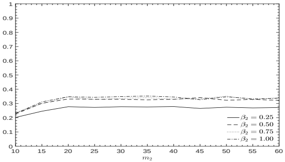
(a) 15 observations before the start of monitoring



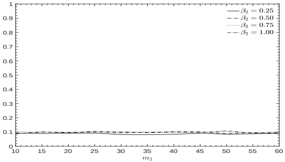
(b) 5 observations before the start of monitoring



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

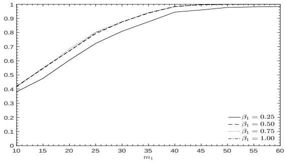


(d) 5 observations after the start of monitoring

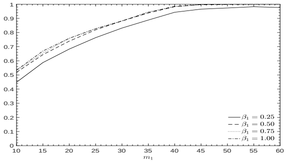


(e) 15 observations after the start of monitoring

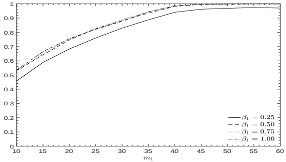
PR detection frequency as a function of m_1: $T^* + m = 302$, $E = 362$, $m = 30$, $\rho = 0.995$



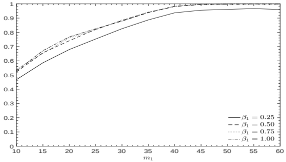
(a) 15 observations before the start of monitoring



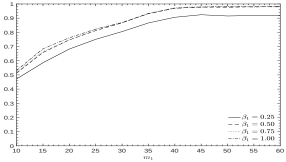
(b) 5 observations before the start of monitoring



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

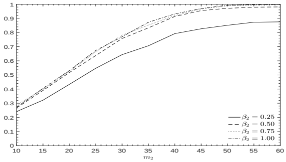


(d) 5 observations after the start of monitoring

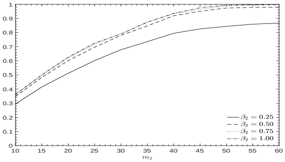


(e) 15 observations after the start of monitoring

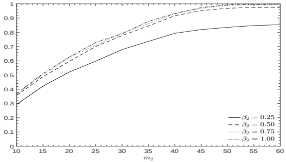
PR detection frequency as a function of m_2: $T^* + m = 302$, $E = 362$, $m_1 = 15$, $m = 30$, $\rho = 0.995$



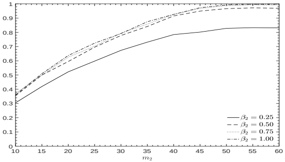
(a) 15 observations before the start of monitoring



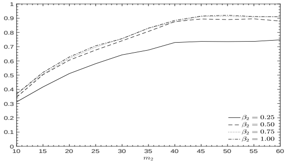
(b) 5 observations before the start of monitoring



(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

Finite Sample Properties of the Monitoring Procedure

- To investigate robustness to conditional and/or unconditional heteroskedasticity, and to non-Gaussian errors, 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.

Empirical Application - The Data

- The dataset used is 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.

Empirical Application - The Data

- 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.
- More recently Neely *et al.* (2014) have investigated the in-sample and out-of-sample predictive power of binary TAIs 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 TAIs have predictive power that matches or exceeds traditional MFVs used as predictors. They also show that combining information from TAIs and MFVs significantly improves equity risk premium forecasts versus using either type of predictor in isolation.

Empirical Application - The Data

- 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 $l = \{9, 12\}$.

Empirical Application - The Data

- The four on-balance volume rule indicators ($OBV_{s,l,t}$) are,

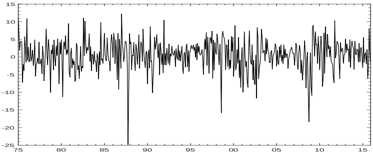
$$OBV_{s,l,t} := \begin{cases} 1, & \text{if } MA^{OBV}_{s,t} \geq MA^{OBV}_{l,t}, \text{ indicating a buy signal} \\ 0, & \text{otherwise,} \end{cases}$$

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,

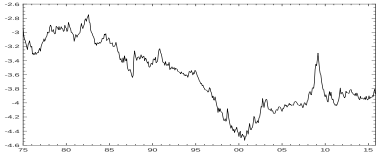
$$D_t := \begin{cases} 1, & \text{if } P_t \geq P_{t-1} \\ -1, & \text{otherwise.} \end{cases}$$

Empirical Application - The Data

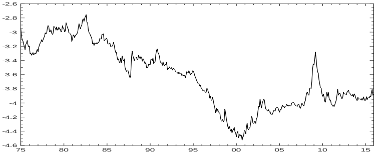
Macroeconomic and financial variables (MFVs)
1. log dividend yield (dy_{t-1})
2. log dividend-price ratio (dp_{t-1})
3. log earnings-price ratio (ep_{t-1})
4. book-to-market ratio (bm_{t-1})
5. short term yield (st_{t-1})
6. long-term yield (lt_{t-1})
7. long-term - short-term yield spread ($sp_{t-1} = lt_{t-1} - st_{t-1}$)
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 ($OBV_{2,12,t-1}$)



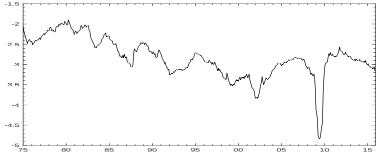
(a) y_t



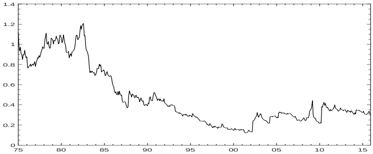
(b) dy_{t-1}



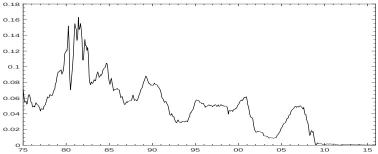
(c) dp_{t-1}



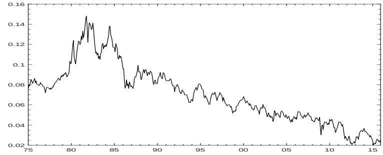
(d) ep_{t-1}



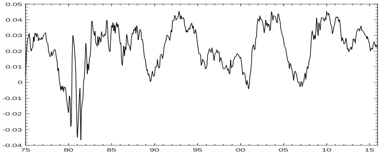
(e) bm_{t-1}



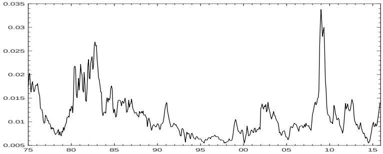
(f) st_{t-1}



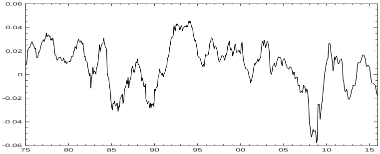
(g) lt_{t-1}



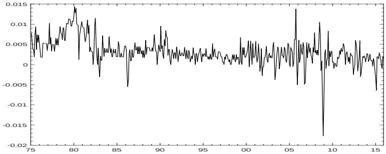
(h) sp_{t-1}



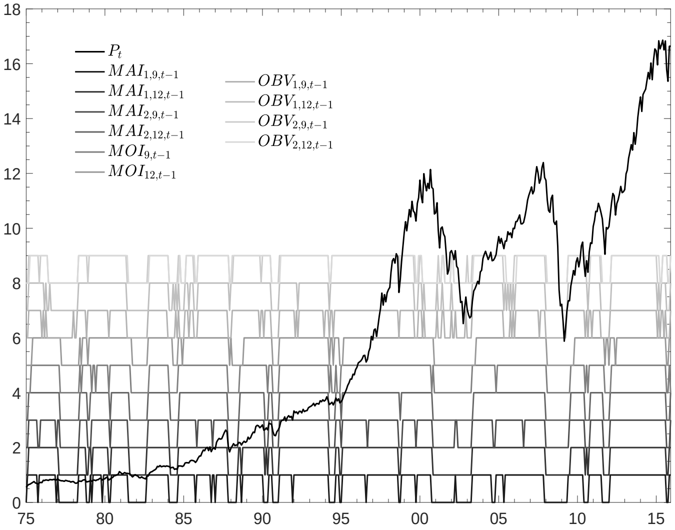
(i) dsp_{t-1}



(j) $ntis_{t-1}$



(k) inf_{t-1}



TAIs and the S&P Composite price index

Empirical Application - The Data

- 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.
- Financial theory suggests negative predictive power for st_{t-1}, lt_{t-1}, $ntis_{t-1}$ and inf_{t-1} and we therefore multiply each of these predictors by -1 so that a right-sided test is appropriate for detecting predictability.

Empirical Application - Preliminary Analysis

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 $\pm$ 1.645.

Empirical Application - Preliminary Analysis

- For both the MFVs and the TAIs, very little evidence of predictability is provided by full sample tests. Eg using the IV_{comb} test of Breitung and Demetrescu (2015) 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.
- We also checked the training samples for predictability. 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.

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

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

Empirical Application - Preliminary Analysis

- For the two interest rate series st_{t-1} and lt_{t-1}, and the bond yield spread dsp_{t-1}, there is statistically significant evidence of predictability at conventional significance levels from IV_{comb} for one or more values of m. The rejections do not appear to be driven by predictability at the end of these samples. Therefore, in the monitoring application 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.
- 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.

Empirical Application - Monitoring Results

- We assume that a practitioner applies our real-time monitoring procedure from 01/00, so in all cases $T^* + m = 302$. Results are presented assuming 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 $\widehat{m}^*$ and the number of predictive regimes detected are ...

Empirical Application - Monitoring Results MFVs

MFVs: $\widehat{m}^*$ 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.

Empirical Application - Monitoring Results MFVs

- It can be seen in the first half of the table 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 the table shows that in total, either one, two or three PRs 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 PRs detected varies depending on the predictor and the value of m used. For ep_{t-1} and lt_{t-1}, one or more PRs are detected for all values of m considered. For bm_{t-1}, a single PR is detected when $m = 20$ and $m = 60$, while for dy_{t-1} and dp_{t-1} two or three PRs 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 PR is detected only when $m = 20$.

Empirical Application - Monitoring Results TAIs

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 $m = 60$).

Empirical Application - Monitoring Results TAIs

- The $\widehat{m}^*$ values reported in the first half of Table 6 are broadly similar to those obtained for the MFVs, but on average are slightly higher.
- For all of the TAIs either one, two, or three predictive regimes are detected for at least one of the values of m considered. Consistent with the findings in Neely *et al.* (2014), we therefore find stronger evidence of predictability for the TAIs than for the MFVs.

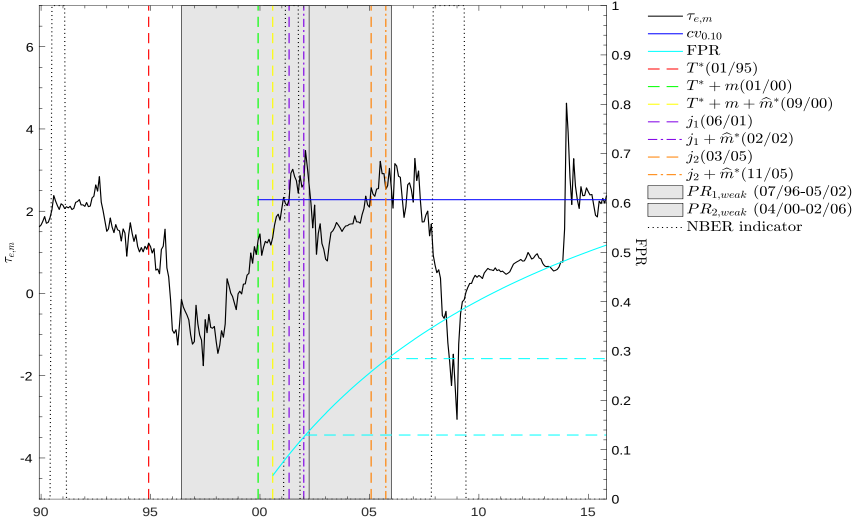
Empirical Application - Monitoring Results TAIs

- NB the TAI predictors are 0-1 dummy variables and often take the same value for several consecutive observations: the subsample $\tau_{e,m}$ values can therefore be undefined if the TAI does not change over the subsample. A large number of these in the training period could have a detrimental impact on the finite sample performance of the procedure. Effect likely to be greater the smaller is m. In practice, we recommend using $m \geq 60$ with these particular TAIs to minimize the number of undefined statistics over the training period.
- Alternatively, for a given m, reducing the value of l in the TAIs will result in fewer undefined 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.

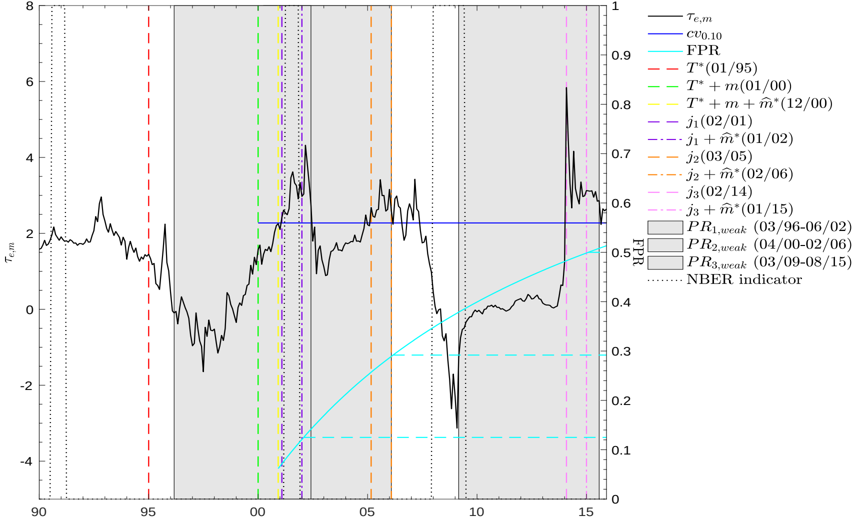
Empirical Application - Further Monitoring Results

- Next graphs for those cases where at least one PR is signalled for $\pi = 0.10$, displaying $\tau_{e,m}$ starting 5 years before the end of each training period. Indicated 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 PR j_i, the date at which the i-th PR is detected $j_i + \widehat{m}^*$, the FPR as a function of E, the weak set of PR dates and, where relevant, the strong set of PR dates.
- Neely *et al.* (2014) find that for both the MFVs and TAIs predictability is substantially higher over recessions than expansions. So we also include the NBER indicator of recessions to see if our procedure finds a similar pattern of support for predictability over the business cycle. There are two NBER 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).

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


$\tau_{e,m}$
$cv_{0.10}$
FPR
$T^*(01/95)$
$T^* + m(01/00)$
$T^* + m + \widehat{m}^*(09/00)$
$j_1(06/01)$
$j_1 + \widehat{m}^*(02/02)$
$j_2(03/05)$
$j_2 + \widehat{m}^*(11/05)$
$PR_{1,weak}$ (07/96-05/02)
$PR_{2,weak}$ (04/00-02/06)
NBER indicator
FPR
$\tau_{e,m}$
90
95
00
05
10
15
6
4
2
0
-2
-4
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0

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



8
6
4
2
0
-2
-4
$\tau_{e,m}$
90
95
00
05
10
15
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
FPR
$\tau_{e,m}$
$cv_{0.10}$
FPR
$T^*(01/95)$
$T^* + m(01/00)$
$T^* + m + \widehat{m}^*(12/00)$
$j_1(02/01)$
$j_1 + \widehat{m}^*(01/02)$
$j_2(03/05)$
$j_2 + \widehat{m}^*(02/06)$
$j_3(02/14)$
$j_3 + \widehat{m}^*(01/15)$
$PR_{1,weak}$ (03/96-06/02)
$PR_{2,weak}$ (04/00-02/06)
$PR_{3,weak}$ (03/09-08/15)
NBER indicator

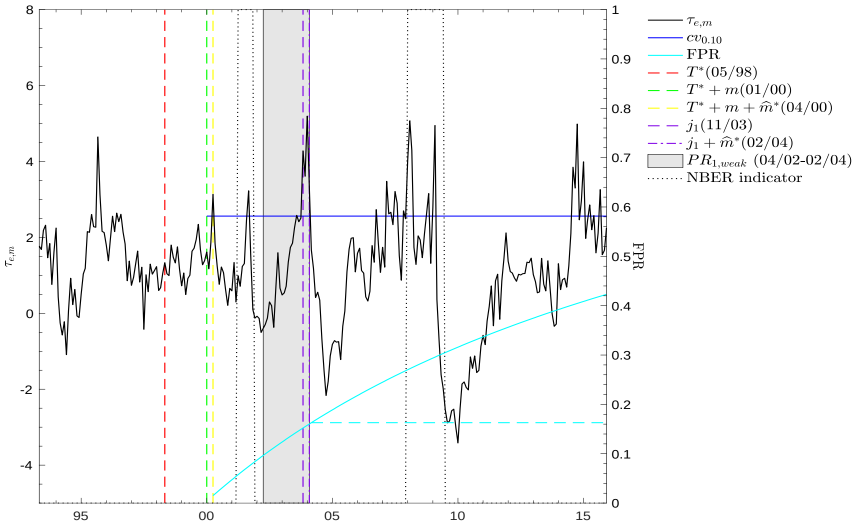
Empirical Application - Further Monitoring Results, MFVs

- For dy_{t-1} with $m = 60$, the $\tau_{e,m}$ test first rejects at 06/01 and a PR is signalled at 02/02 (FPR 0.12) with the weak dates 07/96-05/02. This PR is consistent with the period of the dot-com bubble, being detected towards the end of the boom with the associated PR ending shortly after the crash.
- A further PR is signalled in 11/05 (FPR 0.28), with weak dates 04/00-02/06 which overlap with the first PR dates.
- Similar results for dp_{t-1} with $m = 60$, except a third PR is detected in 01/15. According to the weak sets of dates for each PR, dp_{t-1} therefore had predictive power for the equity premium over the majority of the monitoring period.
- The breakdown of predictability during 03/06-02/09 correlates with a decline in the equity premium due to the global financial crisis. The weak dates 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 in early 2009.

Empirical Application - Further Monitoring Results, MFVs

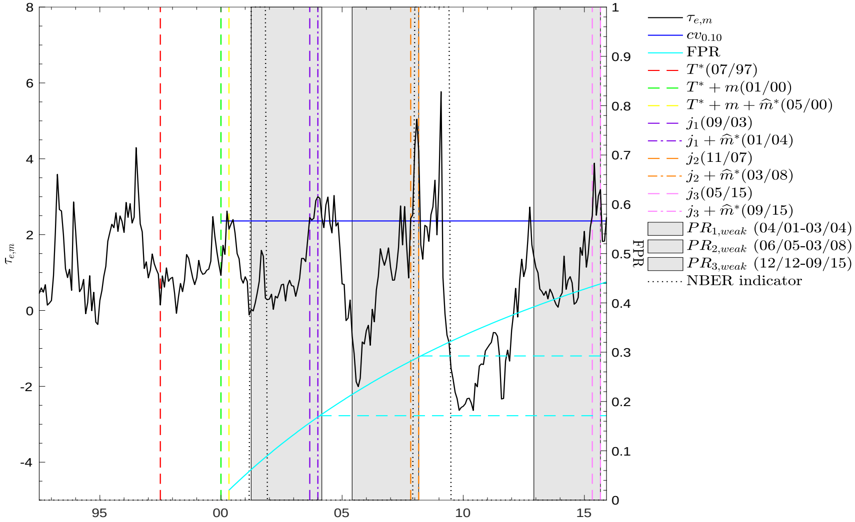
- For ep_{t-1} and $m = 20$ a single PR is signalled in 02/04 [FPR 0.16] with weak dates, 04/02-02/04. For $m = 30$ three PRs are signalled in 01/04, 03/08, and 09/15. On weak dates, the first PR ends two months after its detection, while the second and third end immediately after detection. For $m = 60$ a single PR is detected in 12/04 with weak dates 09/98-06/06.
- The results for bm_{t-1}, $m = 20$ and $m = 60$, are similar to those for dp_{t-1} in that the PR is detected early in the monitoring period around the time of the dot-com bubble, and ends one month later.
- For lt_{t-1} with $m = 20$ a single PR is detected in 09/03 with weak dates 04/01-03/04. For $m = 30$ similar results are seen. For $m = 60$ two PRs are detected in 08/05 and 11/12. Recall for this predictor, statistically significant evidence of predictability is detected during all three training periods.

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



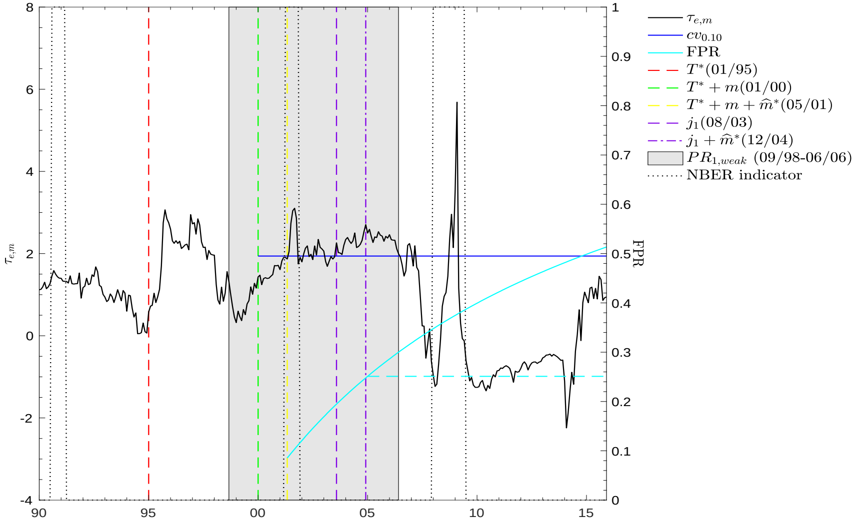
8
6
4
2
0
-2
-4
$\tau_{e,m}$
95
00
05
10
15
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
FPR
$\tau_{e,m}$
$cv_{0.10}$
FPR
$T^*(05/98)$
$T^* + m(01/00)$
$T^* + m + \widehat{m}^*(04/00)$
$j_1(11/03)$
$j_1 + \widehat{m}^*(02/04)$
$PR_{1,weak}$ (04/02-02/04)
NBER indicator

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



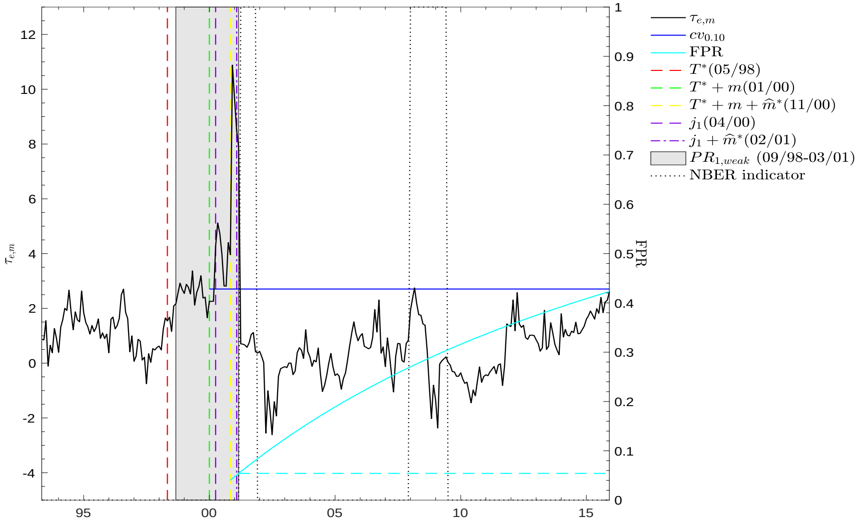
8
6
4
2
0
-2
-4
$\tau_{e,m}$
95
00
05
10
15
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
FPR
$\tau_{e,m}$
$cv_{0.10}$
FPR
$T^*(07/97)$
$T^* + m(01/00)$
$T^* + m + \widehat{m}^*(05/00)$
$j_1(09/03)$
$j_1 + \widehat{m}^*(01/04)$
$j_2(11/07)$
$j_2 + \widehat{m}^*(03/08)$
$j_3(05/15)$
$j_3 + \widehat{m}^*(09/15)$
$PR_{1,weak}$ (04/01-03/04)
$PR_{2,weak}$ (06/05-03/08)
$PR_{3,weak}$ (12/12-09/15)
NBER indicator

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



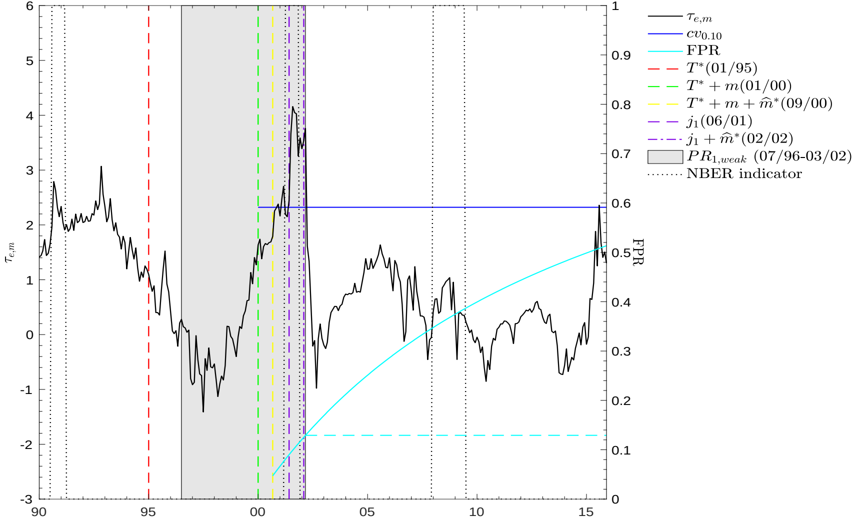
$\tau_{e,m}$
$cv_{0.10}$
FPR
$T^*(01/95)$
$T^* + m(01/00)$
$T^* + m + \widehat{m}^*(05/01)$
$j_1(08/03)$
$j_1 + \widehat{m}^*(12/04)$
$PR_{1,weak}$ (09/98-06/06)
NBER indicator
$\tau_{e,m}$
FPR
8
6
4
2
0
-2
-4
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
90
95
00
05
10
15

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



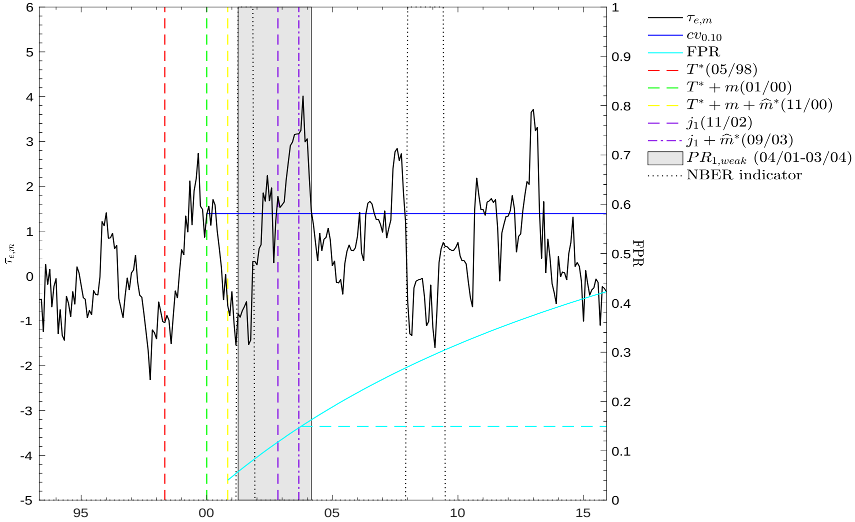
$\tau_{e,m}$
$cv_{0.10}$
FPR
$T^*(05/98)$
$T^* + m(01/00)$
$T^* + m + \widehat{m}^*(11/00)$
$j_1(04/00)$
$j_1 + \widehat{m}^*(02/01)$
$PR_{1,weak}$ (09/98-03/01)
NBER indicator
95
00
05
10
15

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



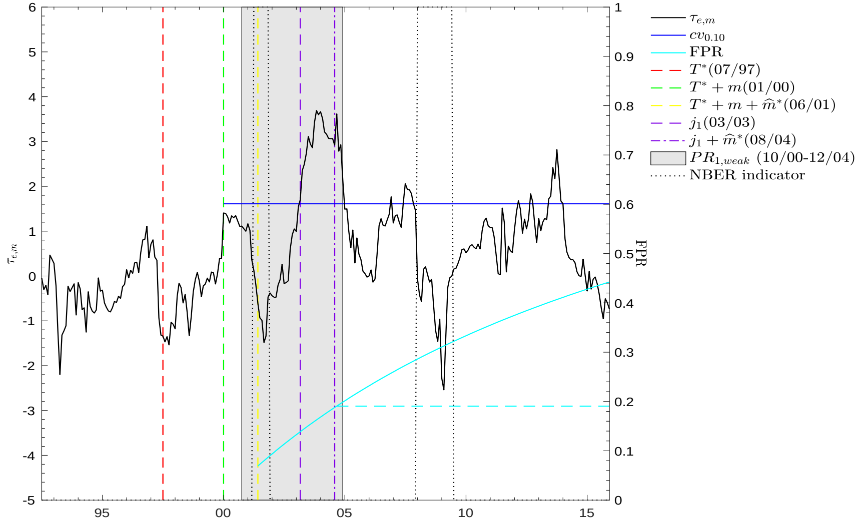
$\tau_{e,m}$
$cv_{0.10}$
FPR
$T^*(01/95)$
$T^* + m(01/00)$
$T^* + m + \widehat{m}^*(09/00)$
$j_1(06/01)$
$j_1 + \widehat{m}^*(02/02)$
$PR_{1,weak}$ (07/96-03/02)
NBER indicator
$\tau_{e,m}$
FPR
6
5
4
3
2
1
0
-1
-2
-3
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
90
95
00
05
10
15

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



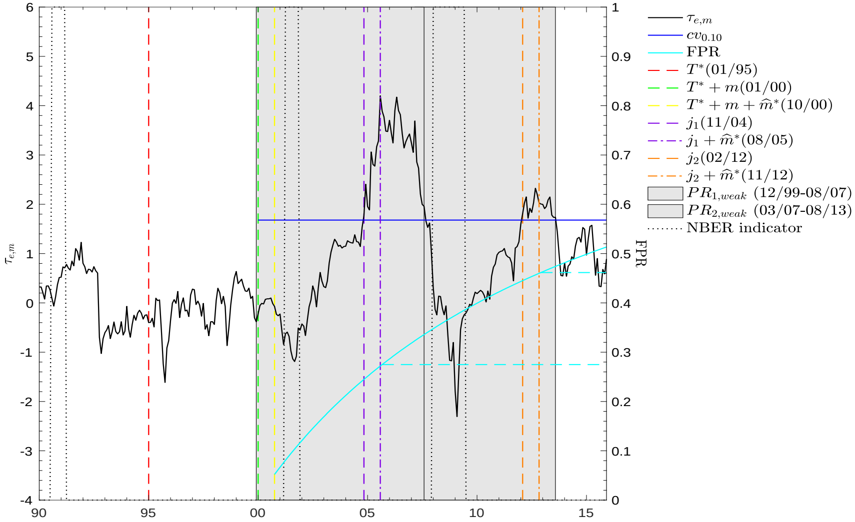
6
5
4
3
2
1
0
-1
-2
-3
-4
-5
$\tau_{e,m}$
95
00
05
10
15
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
FPR
$\tau_{e,m}$
$cv_{0.10}$
FPR
$T^*(05/98)$
$T^* + m(01/00)$
$T^* + m + \widehat{m}^*(11/00)$
$j_1(11/02)$
$j_1 + \widehat{m}^*(09/03)$
$PR_{1,weak}$ (04/01-03/04)
NBER indicator

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



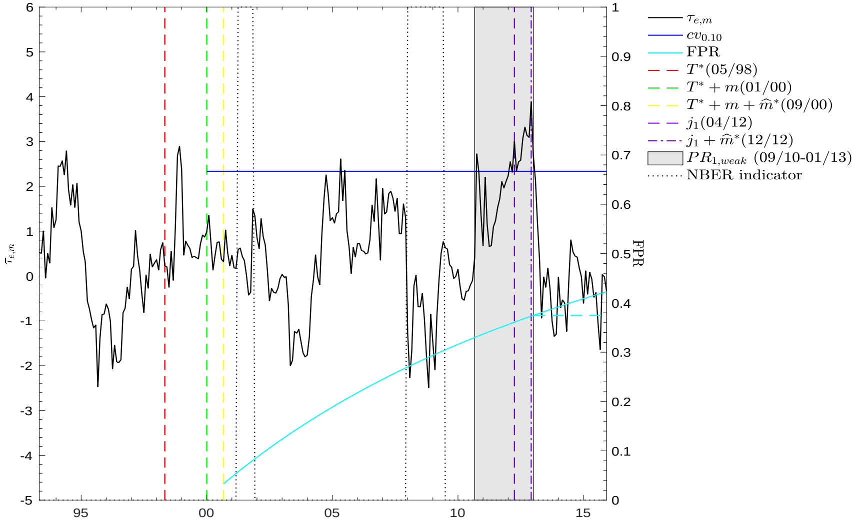
$\tau_{e,m}$
$cv_{0.10}$
FPR
$T^*(07/97)$
$T^* + m(01/00)$
$T^* + m + \widehat{m}^*(06/01)$
$j_1(03/03)$
$j_1 + \widehat{m}^*(08/04)$
$PR_{1,weak}$ (10/00-12/04)
NBER indicator
$\tau_{e,m}$
FPR
95
00
05
10
15

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



6
5
4
3
2
1
0
-1
-2
-3
-4
$\tau_{e,m}$
90
95
00
05
10
15
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
FPR
$\tau_{e,m}$
$cv_{0.10}$
FPR
$T^*(01/95)$
$T^* + m(01/00)$
$T^* + m + \widehat{m}^*(10/00)$
$j_1(11/04)$
$j_1 + \widehat{m}^*(08/05)$
$j_2(02/12)$
$j_2 + \widehat{m}^*(11/12)$
$PR_{1,weak}$ (12/99-08/07)
$PR_{2,weak}$ (03/07-08/13)
NBER indicator

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



6
5
4
3
2
1
0
-1
-2
-3
-4
-5
$\tau_{e,m}$
95
00
05
10
15
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
FPR
$\tau_{e,m}$
$cv_{0.10}$
FPR
$T^*(05/98)$
$T^* + m(01/00)$
$T^* + m + \widehat{m}^*(09/00)$
$j_1(04/12)$
$j_1 + \widehat{m}^*(12/12)$
$PR_{1,weak}$ (09/10-01/13)
NBER indicator

Empirical Application - Further Monitoring Results, MFVs

- For dsp_{t-1} with $m = 20$ a single PR is detected in 12/12 with weak dates 09/10-01/13. Interestingly, this period is one in which the Federal Reserve was operating quantitative easing. The results suggest that this may have influenced the predictive relationship between the yield spread and the equity premium at this time.
- For nearly all the MFVs results, the contiguous run of $\tau_{e,m}$ rejections associated with each PR ends shortly after detection. Consequently the strong set of dates for each predictor is empty. From a practical perspective this suggests that although investors using our procedure in real-time would have been able to detect predictability, there would have been very little time after the point of detection to exploit the predictability before it no longer existed. Consistent with the view of Paye and Timmermann (2006) and Timmermann (2008) 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.

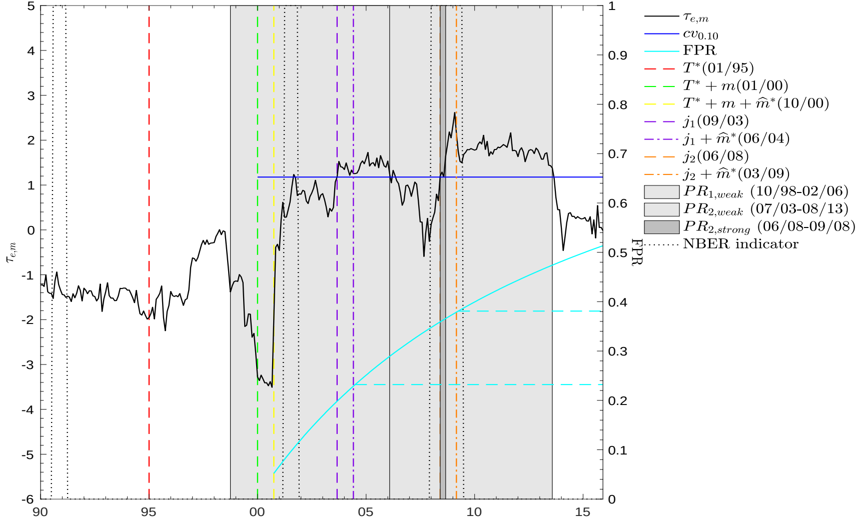
Empirical Application - Further Monitoring Results, MFVs

- Consider now the sequential $\tau_{e,m}$ values in these graphs relative to the NBER indicator of recessions and expansions. Interestingly, 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.
- Eg for ep_{t-1} [figures (c)-(e)], $\tau_{e,m}$ peaks during both the 2001 and 2008-2009 recessions for all three values of m. However for lt_{t-1} [figures (h)-(j)] the statistics fall during or at the start of both recessions. For dy_{t-1} and dp_{t-1} [figures (a) and (b)] $\tau_{e,m}$ peaks during the first recession but reaches a minimum during the 2008-2009 recession.

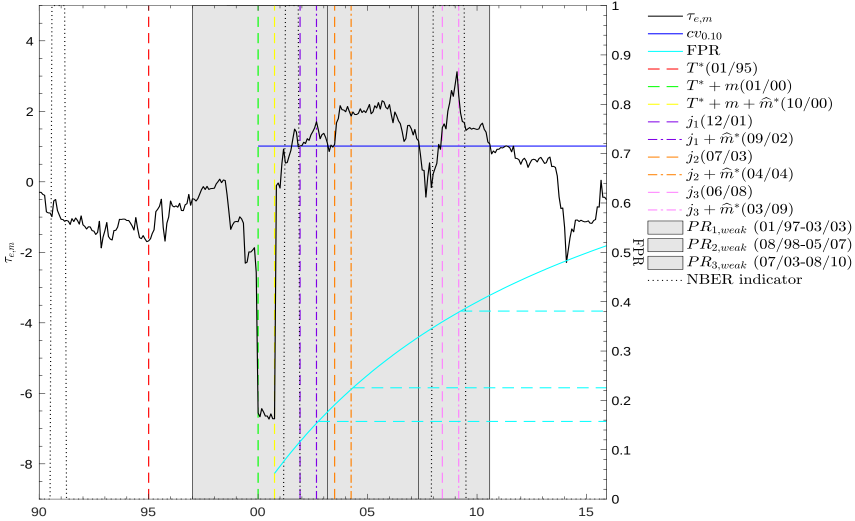
Empirical Application - Further Monitoring Results, TAIs

- For brevity, TAI results just for $m = 60$, $s = 1$.
- For $MAI_{1,9,t-1}$, a first PR is signalled in 06/04 [FPR 0.23] and a second in 03/09 [FPR 0.38]. The first is correlated with the dot-com bubble (although not detected until after the crash) and the second with the global financial crisis. The weak set of dates for the PRs are 10/98-02/06 and 07/03-08/13. For the second PR the strong set of dates is non-empty, 06/08-09/08.
- For both PRs the contiguous rejections continue for much longer after detection than for the MFVs. For the first (second) PR they continue for 20 months (over 4 years) after detection. If these results 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.
- For $MAI_{1,12,t-1}$, similarly to $MAI_{1,9,t-1}$, two PRs are signalled in 04/04 and 03/09. An earlier short PR is also detected at 09/02, again apparently correlated with the dot-com bubble and crash.

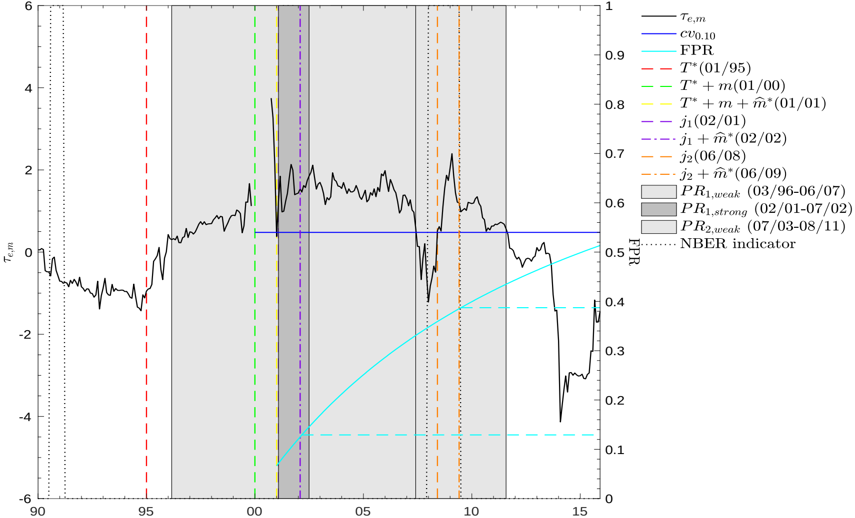
(a) $MAI_{1,9,t-1}$, $m = 60$



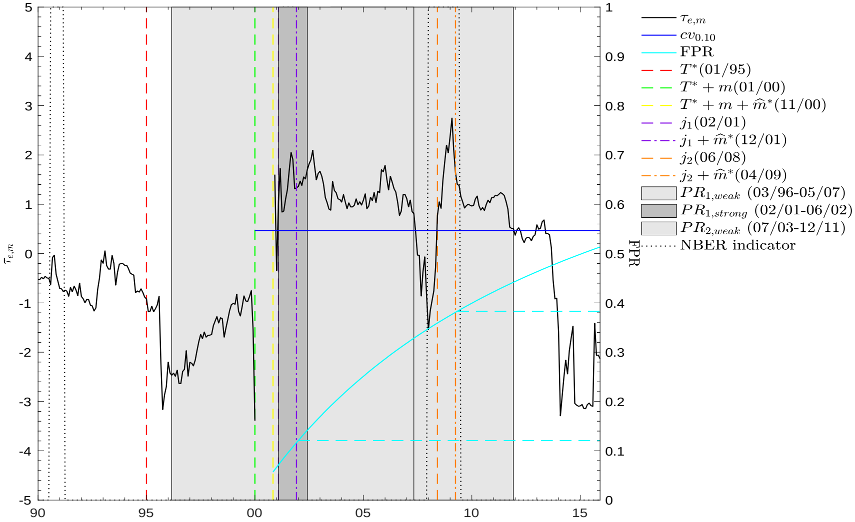
(b) $MAI_{1,12,t-1}$, $m = 60$



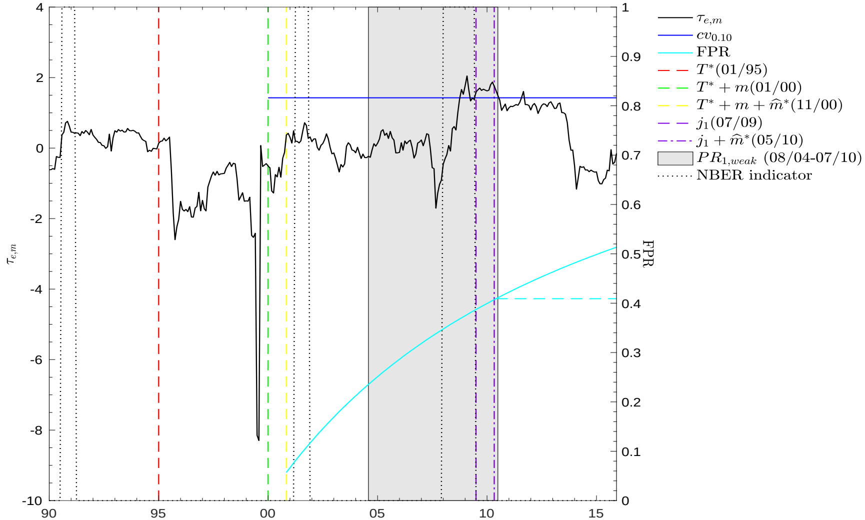
(c) $MOI_{1,9,t-1}$, $m = 60$



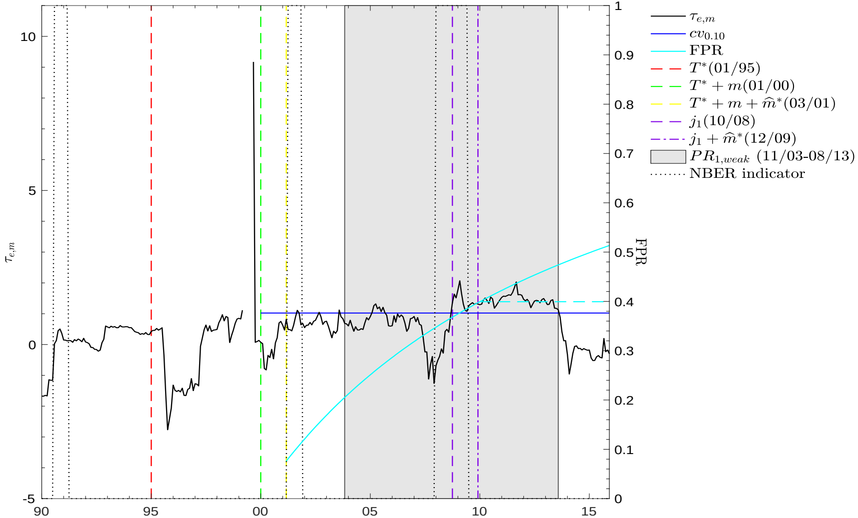
(d) $MOI_{1,12,t-1}$, $m = 60$



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



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



Empirical Application - Further Monitoring Results, TAIs

- The results for $MOI_{9,t-1}$ and $MOI_{12,t-1}$ are similar to the results for the MAIs, in the sense that the first PR 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 PRs are first detected than we found for the MFVs.
- For both of these predictors the strong set of dates for the first PR is not empty. Because the first PR 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 (for both of the MOIs the test statistic is undefined because the indicator has a fixed value for the period 2000-2001).
- For the volume-based indicator $OBV_{1,9,t-1}$ a single PR is detected in 05/10 and similar to the other TAIs the contiguous rejections continue for several years after the initial detection date. For $OBV_{1,12,t-1}$, a PR is detected six-months earlier in 12/09 and the contiguous rejections continue through to 08/13 ($\tau_{e,m}$ is undefined between 1999-2000).

Empirical Application - Further Monitoring Results, TAIs

- 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.
- There is evidence supporting this argument in our results. Consider, eg, $MAI_{1,12,t-1}$. A first PR is detected at 09/02 shortly after the 2001 recession and the second 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 PR because the contiguous run of rejections does not exceed the $\widehat{m}^*$ threshold (this does not happen until 09/02). For both of the MOIs, PRs are detected during or shortly after the 2001 and 2008-2009 recessions, and for the volume-based indicators PRs are detected shortly after the 2008-2009 recession.

Conclusions

- We have developed a new real-time monitoring procedure for detecting the emergence of predictive regimes based on the sequential application of standard heteroskedasticity-robust (predictive) regression t-statistics for predictability to end-of-sample data.
- Critical values are estimated using subsampling methods over a training period. A predictive regime is signalled once a certain number m^* (set by the practitioner) of consecutive t-statistics in the sequence exceed this critical value. We suggest a data-based procedure for choosing m^*, based on the longest run of exceedances in the training period, such that the false positive rate can be controlled, for a given monitoring period.
- Application to predictability of U.S. equity premium at the one-month horizon by macroeconomic and financial variables, and by binary technical analysis indicators. Results suggest the one-month ahead equity premium has displayed episodes of temporary predictability which could have been detected in real-time using our methodology.