Transregio
391

TRR 391 Working Paper #4

April 2025

(When) Do persistent predictors predict stock returns directionally?

Rouven Lindenau, Matei Demetrescu

Suggested citation:
Demetrescu, M., and Lindenau, L. (2025). "(When) Do persistent predictors predict stock returns directionally?". *TRR 391 Working Paper #4*. DOI: 10.17877/DE290R-25452.

Version 1.0, April 2025

Lindenau is a researcher at the University of Kiel. Demetrescu is a professor for econometrics and statistics at TU Dortmund University. The authors are grateful to Christian Leschinski for drawing their attention to the problem treated here, and Robin Kruse-Becher for many helpful comments. Demetrescu gratefully acknowledges financial support from the German Research Foundation (DFG) within TRR 391: Spatio-temporal Statistics for the Transition of Energy and Transport (520388526).

Correspondence to: matei.demetrescu@tu-dortmund.de

Abstract

This paper addresses the question of whether the direction of stock price movements can be predicted, even if only temporarily, by lagged financial predictors. We argue that, analogously to the usual problems faced by predictive regressions for stock returns, directional models like probit or logit type regressions tend to be severely oversized in the presence of highly persistent lagged predictors when employing standard inference. To robustify against uncertain persistence of the underlying time series we resort to estimation and testing of a specific auxiliary regression. Following Demetrescu et al. (2022, J. Econometrics 227, 85-113), we employ tailored instrumental variables for this purpose, deployed in a rolling window approach. Based on a fixed-regressor wild bootstrap implementation, we are able to pin down periods of highest predictability for each considered predictor irrespective of its degree of persistence, while controlling for the effect of repeated testing. Monte Carlo simulations for a variety of empirically relevant data generating processes are conducted to confirm the reliability of the testing scheme proposed here. For the sign of monthly returns of the S&P500 index, we find the inflation rate, the dividend yield, dividend payout ratio, net equity expansion, and term spread, to exhibit significant predictive power only within historical pockets, primarily between 1940 and 1980. Regarding daily returns of individual stocks of the DJIA, the (lagged) stock return itself, a moving average of closing prices, the divergence between price and moving average, and the stochastic oscillator exhibit periods of significant predictive power, though without a clear temporal pattern.

Key words: Sign predictability, Binary dependent variables, Predictive regression endogeneity, Uncertain persistence

1 Motivation

The evidence on the age-old yet empirically relevant puzzle regarding the existence of stock return predictability is still full of gaps and contrary findings. In principle, the existence of economically meaningful and exploitable predictability of returns has been largely ruled out—at least for longer periods of time—as it would lead to (rational) agents systematically receiving abnormal profits when employing appropriate investment decisions. Notwithstanding, empirical research has revealed the existence of predictive power of various financial ratios and economic fundamentals; see e.g. Campbell (1987), Campbell and Shiller (1988) or Fama and French (1988, 1989). Such studies set up regression models using lagged financial or macroeconomic regressors to discover statistical evidence for the predictability of stock returns. Further studies have since highlighted inconsistencies when considering out-of-sample forecasting (Welch and Goyal, 2008), pointed out that findings of predictability are largely rendered useless when taking transaction costs and the non-operationality of certain trading strategies into account (Bajgrowicz and Scaillet, 2012), and also discussed the issue of data snooping and short evaluation periods (Reschenhofer et al., 2020). Spurious statistical inference is particularly likely to occur in predictive time series regressions where unobservable properties of the predictor variables, specially their degree of persistence, lead to nonstandard inference. As shown in Stambaugh (1999), this becomes increasingly problematic the more the innovations of the predictor correlate with the innovations of the returns: when strong contemporaneous correlation of innovations is paired with highly persistent predictors, the issue of predictive regression endogeneity arises, leading to severe biases when conducting statistical inference based on standard critical values. The phenomenon of endogeneity is a common feature of price-scaled ratios, such as the dividend-price ratio or earnings-price ratio, which are popular indicators used to predict stock returns, e.g. in Rozeff (1984) or Campbell and Shiller (1988).

To robustify against such endogeneity, the literature proposed a number of statistical techniques. See e.g. the bias-correction approach of Amihud and Hurvich (2004), the Bonferroni procedure of Campbell and Yogo (2006), or more recently technically more involved approaches by Zhu et al. (2014) or Elliott et al. (2015). To the same end, Breitung and Demetrescu (2015) and Kostakis et al. (2015) study estimation and testing based on suitably generated instrumental variables. In fact, the IVX instruments technique of Kostakis et al. (2015) has become popular to recover asymptotically valid inference in linear regression models (see e.g. Rapach et al., 2016; Yu and Huang, 2023) but is also put to work in nonlinear predictive regressions models (Gonzalo and Pitarakis, 2012, 2017; Demetrescu and Hillmann, 2022). Once such robust techniques are employed the evidence of predictability is much weaker; see e.g. Goyal et al. (2024) or Breitung and Demetrescu (2015).

Due to the mixed empirical evidence of mean return predictability, a further body of literature suggests that it might be more likely to find predictability for the direction of price changes instead of predictability in the mean of returns, thereby attempting to forecast the sign of stock returns.[1] See, among others, Christoffersen and Diebold (2006), Nyberg (2011), Han et al. (2016), Becker and Leschinski (2018), Basak et al. (2019) or Campisi et al. (2024). Arguably, directional

[1]A related research question is whether specific quantiles instead of the mean of stock returns are predictable. See e.g. Lee (2016); Demetrescu et al. (2024), who use IV(X) methods, or Cai et al. (2023); Liu et al. (2024).

predictability is as important for market timing and asset allocation as predicting the level of returns. Leung et al. (2000) even conclude that the focus of practitioners should lie on predicting the direction of returns instead of the level as it yields a superior forecasting rate and thereby higher trading profits.

However, the literature on directional predictability largely builds on estimating models for binary dependent variables like probit or logit type regression, yet knowledge on the behavior of these models under realistic conditions, in particular uncertain predictor persistence, is scarce. Furthermore, to reinforce the potential uncovering of stock return predictability, treating the phenomenon as time-varying appears to be sensible. Among others, Paye and Timmermann (2006), Goyal and Welch (2003), Pettenuzzo and Timmermann (2011), Henkel et al. (2011) and Demetrescu et al. (2022) provide arguments supporting, and find evidence of, stock return predictability in certain times only. For predictive regression models, this implies that the coefficients determining predictability vary with time. Along these lines, predictability could be difficult to detect in some periods due to the low signal-to-noise ratio, while in other periods there might be stronger statistical evidence. Theoretical reasoning to consider a time-varying behavior is well-founded, in fact it is quite likely that predictability, if present, is arbitraged away after some time. And indeed, empirical evidence in favor of such structural breaks and model instabilities is extensive. In particular, Demetrescu et al. (2022) develop bootstrap implementations of tests of no predictability applied to various subsamples of the data. By suitable aggregation, these implementations control for the inherent multiple testing problem resulting from repeatedly applying predictability tests onto various subsets of the data, thus avoiding spurious rejections caused by repeated testing not properly accounted for. To the best of our knowledge, there are no comparable procedures to detect directional predictability that are able to cope with endogeneity, uncertain persistence and subsample implementations.

We therefore aim to uncover periods of directional predictability in stock returns using tailored, robust inferential techniques. We begin by confirming that classical predictability tests embedded in a binary dependent variable setup, concretely logit and probit approaches, are prone to the predictive regression endogeneity problem and may substantially overstate findings of directional predictability. To obtain robust inference, we take an LM approach to testing the null of no directional predictability, which may be implemented using linear regressions. Then, the IVX technique of Kostakis et al. (2015), extended by the 2SLS instrument variable combination test of Breitung and Demetrescu (2015) is used to robustify these regressions against uncertain predictor persistence. In particular, we adapt the subsample bootstrap implementation of the 2SLS approach proposed by Demetrescu et al. (2022) to the directional setup of this paper. The resulting directional 2SLS procedure is used to analyze predictability of the sign of daily and monthly returns in subsets of the data, where we resort to a rolling window approach to find the period of highest predictive power of various putative predictors and provide robust inference on predictive power by computing p-values based on a fixed-regressor wild bootstrap algorithm. Finally, we conduct two empirical studies of directional predictability. First, we use the updated data set of Welch and Goyal (2008) to check for periods of directional predictability in the S&P 500 on a monthly basis; second, we examine daily data and consider a total of 25 different stocks, which were part of the Dow Jones Industrial Average (DJIA) in December 2021. We find several rejections of the null of no predictability for monthly as well as daily data. Our analysis of monthly S&P 500 returns highlights that several predictors, such as the dividend yield, dividend payout ratio, net equity expansion, and term spread, exhibit significant predictability only within historical pockets, primarily between 1940 and 1980, with their significance fading in more recent decades, particularly after 1990. Notably, the inflation rate remained a significant predictor from 1980 to 2010, a pattern that would not have been uncovered using a full-sample approach. The analysis of daily returns for individual DJIA stocks reveals consistent evidence of predictability. Key predictors include the stock return itself, a moving average of closing prices, the divergence between price and moving average, and the stochastic oscillator, which showed the highest number of significant results. The identified periods of predictability, however, appear randomly distributed throughout the sampling period without a clear temporal pattern.

2 Detecting episodic directional predictability

2.1 Econometric setup

Let y_t denote the returns series of interest, and denote by $\boldsymbol{x}_{t-1}$ the K potential predictors of the direction of a price change, i.e. of the sign of y_t. To focus on the direction of a change of the relevant asset price, denote by s_t the sign of y_t, which we code as 1 if the return y_t is nonnegative and 0 otherwise:

$$s_t = \mathbb{I}\left(y_t \geq 0\right),$$

where $\mathbb{I}$ denotes the usual indicator function.[2] We do not treat the case of a zero return separately as returns for liquid stocks plausibly come from a continuous distribution.

In the empirical asset pricing literature, y_t is often assumed to be well described by models such as

$$y_t = f\left(\alpha + \boldsymbol{\beta}'\boldsymbol{x}_{t-1}\right) + u_t \tag{1}$$

with u_t zero-mean errors, where f is in most cases linear, like in Stambaugh (1986); Amihud and Hurvich (2004); Campbell and Yogo (2006); Welch and Goyal (2008); Breitung and Demetrescu (2015); Kostakis et al. (2015).[3]

The probability that y_t has a given sign then depends on the distribution of the latent error term u_t and on the predictor variables $\boldsymbol{x}_{t-1}$ via the regression function f. For instance, the classical logit or probit models are obtained for corresponding choice of the error term distribution u_t. And, indeed, logit and probit regression have been extensively used in studies of directional predictability; see e.g. Nyberg (2011), Nyberg and Pönkä (2016), Pönkä (2017) and Becker and Leschinski (2018). We note however that neither f nor the distribution of u_t are known a priori in practice.

Such knowledge is perhaps not even strictly necessary: since the focus is on directional predictability here, we make direct assumptions on the sign s_t of interest rather than on y_t from (1).

[2] In the forecasting literature the direction of change is often coded as $\{-1, 1\}$ rather than $\{0, 1\}$, which is of course just a matter of convention; we choose the latter as it is typical for the econometric literature on binary dependent variable models.

[3] See furthermore Demetrescu and Hillmann (2022) or Chen et al. (2023) for nonlinear approaches.

Assumption 1 *Let* $\mathcal{F}_{t-1} = \{\boldsymbol{x}_{t-1}, \boldsymbol{x}_{t-2}, \ldots, y_{t-1}, y_{t-2}, \ldots\}$ *and*

$$\mathrm{P}\left(y_t \geq 0 | \mathcal{F}_{t-1}\right) = \mathrm{P}\left(s_t = 1 | \mathcal{F}_{t-1}\right) = g\left(\alpha + \boldsymbol{\beta}'\boldsymbol{x}_{t-1}\right), \tag{2}$$

where g *is a smooth, strictly monotonic real function s.t.* $0 < g(u) < 1 \ \forall u \in \mathbb{R}$.

The "link" function g summarizes the dependence of the sign of y_t – rather than the levels y_t in usual predictive regression – on $\boldsymbol{x}_{t-1}$ and u_t. Of course, g depends in this model on f and the distribution of u_t as well, which are thus only dealt with indirectly. Importantly, the case of no directional predictability is still captured by zero slope coefficients $\boldsymbol{\beta}$ in this setup, and this is the formal null hypothesis we test: $H_0 : \boldsymbol{\beta} = \mathbf{0}$. We note that the assumption only imposes generic monotonicity and smoothness assumptions on the link function g. In contrast, logit regression, say, imposes a particular choice of g, the logistic function. We are able to side-step such strong assumptions by resorting to an LM type testing approach rather than to the perhaps more common Wald test building on parameter estimators in a fully specified parametric model; see Section 2.3. The only restriction we impose on g is strict monotonicity, which is however natural in this setup and in fact fulfilled by all standard binary dependent variable models such as probit or logit regression.

There is no need to assume any explicit properties of the disturbances u_t here. This is because we model the sign directly, such that the magnitude of price changes is irrelevant. In particular, u_t may be heteroskedastic, conditionally and unconditionally, and one does not even require existence of any moments of u_t.

The putative predictors are generated as

$$\boldsymbol{x}_t = \boldsymbol{\mu}_{\boldsymbol{x}} + \boldsymbol{\xi}_t \tag{3}$$

with $\boldsymbol{\mu}_{\boldsymbol{x}}$ a vector of intercepts, and we set the purely stochastic component to follow the autoregressive data generating process [DGP]

$$\boldsymbol{\xi}_t = \boldsymbol{\Gamma}\boldsymbol{\xi}_{t-1} + \boldsymbol{v}_t, \tag{4}$$

where $\boldsymbol{\Gamma}$ controls the persistence of the predictors and is unknown, and the increments $\boldsymbol{v}_t$ only exhibit weak serial dependence.

Assumption 2 *Let the predictor data be generated as in* (3) *and* (4), *where, with* $\mathbf{I}_K$ *denoting the* $K \times K$ *identity matrix, exactly one of the two following conditions holds true:*

1. ***Weakly persistent predictors:*** *The coefficient matrix* $\boldsymbol{\Gamma}$ *is fixed and has eigenvalues bounded away from unity,* $|\mathbf{I}_K - \lambda\boldsymbol{\Gamma}| = 0$ *implies* $|\lambda| < 1$.
2. ***Strongly persistent predictors:*** *The coefficient matrix* $\boldsymbol{\Gamma}$ *is local-to-unity with* $\boldsymbol{\Gamma} := \mathbf{I}_K - \frac{1}{T}\mathbf{C}$ *where* $\mathbf{C}$ *is a fixed* $K \times K$ *real matrix.*

We now get more precise on the weak serial dependence properties of the increments $\boldsymbol{v}_t$.

Assumption 3 *Let $\boldsymbol{v}_t = \mathbf{B}(L)\mathbf{H}(t/T)\boldsymbol{e}_t$ with $\boldsymbol{e}_t \sim$ White Noise$\,(\mathbf{0}, \mathbf{I}_K)$, where:*

1. *$\boldsymbol{e}_t$ is a strictly stationary and ergodic, L_4-bounded martingale difference sequence which is such that $\sup_t \mathrm{E}\left\|\mathrm{E}\left(\boldsymbol{e}_t\boldsymbol{e}_t' - \mathbf{I}_K|\mathcal{F}_{t-m}\right)\right\| \to 0$ as $m \to \infty$ and, with $\otimes$ denoting the Kronecker product, $\sup_{t\in\mathbb{Z}} \left\|\mathrm{E}\left(\left(\boldsymbol{e}_t\boldsymbol{e}_t' - \mathrm{E}\left(\boldsymbol{e}_t\boldsymbol{e}_t'\right)\right) \otimes \boldsymbol{e}_{t-j}\boldsymbol{e}_{t-k}'\right)\right\| \le C\,(jk)^{-1/2-\vartheta/2}$ for some $\vartheta > 0$;*
2. *$\mathbf{H}(\cdot)$ is a $K \times K$ matrix of piecewise Lipschitz-continuous bounded functions on $(-\infty, 1]$, which is of full rank at all but a finite number of points;*
3. *$\mathbf{B}\,(L)$, where L denotes the usual lag operator, is an invertible lag polynomial with $\mathbf{B}_0 = \mathbf{I}_K$ and 1-summable coefficients, $\sum_{j\ge0} j\,\|\mathbf{B}_j\| < \infty$, for which $\mathbf{\Omega} := \sum_{j\ge0}\mathbf{B}_j$ has full rank.*

In what concerns the putative predictors, this is essentially the DGP employed by Demetrescu et al. (2022) and we refer to their Section 2 and Section S.2.2 of the Supplementary Appendix for a detailed discussion of the importance of these technical requirements. In addition to uncertain persistence, the data may exhibit time-varying volatility, both slowly varying (as occuring e.g. during the Great Moderation) and fast-varying (as captured by standard GARCH-type models with volatility clustering). For simplicity of the presentation we focus on the case of a similar persistence for all regressors, but note that, according to Section S.2.2 of Demetrescu et al. (2022, Supplementary Appendix), we may allow for regressors of mixed persistence. Furthermore, the contemporaneous dependence of u_t and $\boldsymbol{v}_t$ is not restricted in any way such that we account for predictive regression endogeneity in this model.

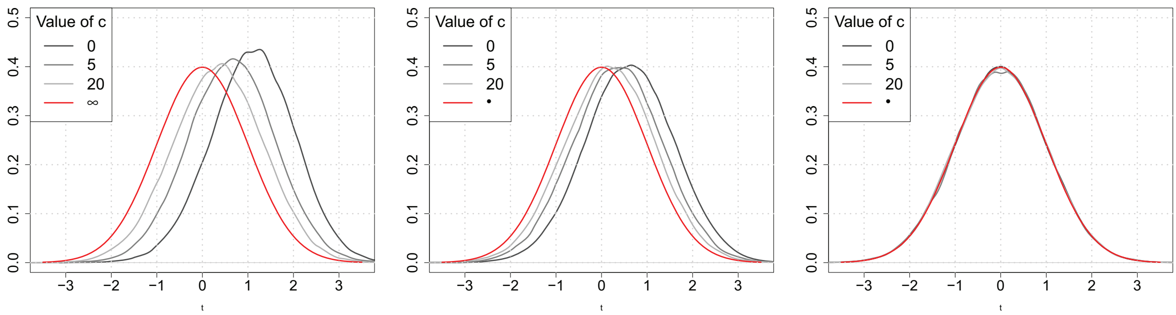
2.2 Extant methods of inference on sign predictability

For weakly dependent predictors, binary dependent variable regressions for known g are well-understood. With nonstationary regressors however, the distribution of corresponding estimators is nonstandard in general and the estimators often have reduced convergence rates. See, as a leading case, the analysis of Park and Phillips (2000) of the (conditional) ML estimator of $\hat{\boldsymbol{\beta}}$ with integrated regressors when $\boldsymbol{\beta} \neq \mathbf{0}$.

Under our null hypothesis of interest $\boldsymbol{\beta} = \mathbf{0}$, Guerre and Moon (2002) derive the convergence rate and limiting distribution of the logit regression estimator $\hat{\beta}$ in a model with a single regressor when g is known, x_t is integrated with shocks v_t independent of u_t, and there is no intercept. Superconsistency is given in this case, and the limiting distribution of $\hat{\beta}$ is mixed Gaussian, leading to a standard normal limiting null distribution of the t statistic for the null of no predictability $\beta = 0$. A multiple predictors extension of this finding is provided by Mao (2014), again under contemporaneous independence of the shocks u_t and v_t and no intercept.

But the assumption of contemporaneously independent errors is not plausible for stock return data and many of the candidate predictors available (such as dividend yields), and no intercept for model (2) implies that positive and negative signs are equally likely under the null (which is at odds with stylized facts of stock returns, broadly stating that upward price movements are somewhat more likely than downward price movements). If these conditions are dropped, essentially different limiting behavior of the ML estimators results: in general mixed Gaussianity is lost, and the Wald statistics based on the ML estimators $\hat{\boldsymbol{\beta}}$ have nonstandard distributions, depending on the relation between u_t and the properties of $\boldsymbol{x}_t$. It would in fact have been surprising if predictive regression endogeneity (i.e. contemporaneous dependence of u_t and $\boldsymbol{v}_t$) and regressor persistence had posed—unlike in a standard OLS predictive regression—no problems.

To gauge the extent of such endogeneity effects we now provide an illustrative Monte Carlo experiment. We focus on the case of a single predictor and generate 100000 replications of y_t, x_t according to (1), (3) and (4), $T = 1000$, under the null of no predictability, $\beta = 0$. The focus being on the regressor persistence, let furthermore $f(\alpha) = 0$. The shocks u_t, v_t are iid standard normal and correlate contemporaneously with correlation $\varphi \in \{-0.9, -0.5, 0\}$.[4] This constellation amounts to $\mathrm{P}(y_t \geq 0|\mathcal{F}_{t-1}) = \mathrm{P}(y_t < 0|\mathcal{F}_{t-1}) = 0.5$ in (2), but one may expect qualitatively similar findings for other sign probabilities. The putative predictor is generated as an AR(1) process with zero mean and near-unity coefficient $\rho = 1 - c/T$ for various choices of c. For each replication, we run a logit regression of $s_t = \mathbb{I}(y_t \geq 0)$ on x_{t-1} including an intercept and test the slope coefficient for significance using the usual t statistic based on the ML estimator $\hat{\beta}$ in the logit regression.[5] The finite-sample densities of the test statistic (obtained by kernel smoothing of the MC realizations) are depicted in Figure 1.



(a) $\varphi = -0.9$ (b) $\varphi = -0.5$ (c) $\varphi = 0$

Figure 1: Finite-sample distributions of t statistics testing $\beta = 0$ in a simple logit model relating $s_t = \mathbb{I}(y_t \geq 0)$; data generated under the null, $y_t = u_t$, $T = 1000$, the predictor $x_t = \rho x_{t-1} + v_t$ is near-integrated, $\rho = 1 - c/T$ for various choices of $c \in \{0, 5, 20\}$, and u_t, v_t are jointly standard normal with correlation -0.9 (left), -0.5 (middle) and 0 (right). The density of the standard normal distribution corresponds to $c = \infty$.

We note that the actual distributions of logit-based tests differ from the standard normal in spite of the relatively large sample size of $T = 1000$, in particular we notice noncentrality of the finite-sample distributions of the logit t statistic wheneven $\varphi \neq 0$. Furthermore, the distortions depend on the degree of persistence as captured by the mean reversion parameter c and decrease as mean reversion becomes stronger. The amount of endogeneity also plays an important role: the distortions decrease as the correlation decreases and vanishes completely in the limit for zero contemporaneous correlation. These problems are the binary dependent variable regression counterpart of the issues of OLS predictive regressions with persistent predictors (for overview articles, see Campbell, 2008; Phillips, 2015). In a nutshell, standard inferential procedures for binary predictive regression models can lead to quite misleading conclusions under regressor persistence and endogeneity.

[4]These are not uncommon figures for stock return predictive regressions; see e.g. Campbell and Yogo (2006).
[5]We alternatively ran probit regressions with virtually the same results.

But even if logit regression were reliable under endogeneity, the issue of controlling size when computing the test over several subsamples of the (potentially nonstationary) data remains. Moreover, Wald tests rely on specifying the correct link function g, with misspecification implying possibly unreliable inference. Therefore, we provide (simple) inferential methods that cope with all mentioned requirements in the following.

2.3 Subsample tests of no directional predictive power

We start with the full-sample directional predictability testing problem. To tackle the problem we resort to an LM approach. This, as we shall see, is straightforward to adapt for our purposes. The (normalized) log-likelihood of the model is given, up to an additive irrelevant term, by

$$\ell(\alpha, \boldsymbol{\beta}) = \frac{1}{T}\sum_{t=1}^{T}\left[(1-s_t)\log\left(1-g\left(\alpha+\boldsymbol{\beta}'\boldsymbol{x}_{t-1}\right)\right)+s_t\log\left(g\left(\alpha+\boldsymbol{\beta}'\boldsymbol{x}_{t-1}\right)\right)\right].$$

The score is then given by

$$\nabla\ell(\alpha, \boldsymbol{\beta}) = \frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix}1\\ \boldsymbol{x}_{t-1}\end{pmatrix}\left[\frac{s_t g'\left(\alpha+\boldsymbol{\beta}'\boldsymbol{x}_{t-1}\right)}{g\left(\alpha+\boldsymbol{\beta}'\boldsymbol{x}_{t-1}\right)}-\frac{(1-s_t)\, g'\left(\alpha+\boldsymbol{\beta}'\boldsymbol{x}_{t-1}\right)}{1-g\left(\alpha+\boldsymbol{\beta}'\boldsymbol{x}_{t-1}\right)}\right],$$

and the LM principle tests whether the score, evaluated at the null, is significantly different from $\mathbf{0}$. To derive an estimator of α under the null $\boldsymbol{\beta}=\mathbf{0}$, say $\hat{\alpha}_0$, we immediately obtain as f.o.c. that

$$\frac{1}{T}\sum_{t=1}^{T}\left[\frac{s_t g'(\alpha)}{g(\alpha)}-\frac{(1-s_t)\, g'(\alpha)}{1-g(\alpha)}\right]=0$$

or

$$g(\hat{\alpha}_0)=\frac{1}{T}\sum_{t=1}^{T}s_t=\bar{s}.$$

We therefore obtain as score under the null

$$\nabla\ell(\hat{\alpha}_0, \mathbf{0}) = \frac{1}{T}\frac{g'(\hat{\alpha}_0)}{\bar{s}(1-\bar{s})}\sum_{t=1}^{T}\begin{pmatrix}0\\ \boldsymbol{x}_{t-1}(s_t-\bar{s})\end{pmatrix}.$$

Since it follows from Assumption 1 that g' is nonzero, we may focus on the score component corresponding to $\boldsymbol{\beta}$,

$$\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{x}_{t-1}(s_t-\bar{s})=\frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{x}_{t-1}-\bar{\boldsymbol{x}})(s_t-\bar{s});$$

in particular, this sample covariance should not be significantly different from zero under the null of no directional predictability. This is however equivalent with testing the null that $\boldsymbol{\gamma}=\mathbf{0}$ in the auxiliary OLS regression

$$s_t=\gamma_0+\boldsymbol{\gamma}'\boldsymbol{x}_{t-1}+error. \tag{5}$$

The resulting test statistic does not depend on g (i.e. on the distribution of u_t and the shape of the conditional mean function f). This is convenient, since g is only identified under the alternative.

Nevertheless, the auxiliary regression is not the final answer to the inferential question posed here, since s_t correlates in general with the shocks $\boldsymbol{v}_t$ leading to nonpivotal distributions just like for the OLS predictive regression; see also Section 2.2. But the difference to the usual Wald statistic in logit regression models is that robustifying linear predictive regressions of the type (5) is well-understood; see, among others, Cavanagh et al. (1995); Campbell and Yogo (2006); Jansson and Moreira (2006); Maynard and Shimotsu (2009); Zhu et al. (2014); Kostakis et al. (2015); Breitung and Demetrescu (2015) or Elliott et al. (2015).

Furthermore, since we aim to detect *episodes* of sign predictability, we resort here to the overidentified 2SLS approach advocated Demetrescu et al. (2022) for such situations. The directional 2SLS approach then consists of testing for significance of $\hat{\boldsymbol{\gamma}}$ in (5) using two types of instruments for $\boldsymbol{x}_t$, computing a 2SLS based test of the null $\boldsymbol{\gamma} = \mathbf{0}$ over certain sequences of subsamples, and aggregating the subsample tests by taking the largest test outcome (in absolute value) over all considered subsamples.

For simplicity we will focus on the univariate case $K = 1$ from now on.[6] The two instruments we use are collected in the vector $\tilde{\boldsymbol{z}}_t = (z_{I,t}, z_{II,t})'$ and our choice of instruments is based on the discussion in Demetrescu et al. (2022). In particular we use the IVX instrument advocated by of Kostakis et al. (2015) as the type-I instrument, s.t.

$$z_{I,t} = \varrho z_{t-1} + \Delta x_t = \sum_{j=0}^{t-1} \varrho^j \Delta x_{t-j} \quad \text{with} \quad \varrho = 1 - \frac{a}{T^\eta} \tag{6}$$

with initial value $z_0 = 0$, user chosen IVX parameters $a > 0$ and $0 < \eta < 1$ and first difference operator Δ. For the type-II instrument we employ a sine function of time,

$$z_{II,t} = \sin\left(\frac{\pi t}{2T}\right). \tag{7}$$

Then, for a single subsample consisting of the observations $t = \lfloor \tau_1 T \rfloor + 1, \ldots, \lfloor \tau_2 T \rfloor$ for $0 \leq \tau_1 < \tau_2 \leq 1$ the resulting test statistic is given by

$$t_{(\tau_1,\tau_2)} = \frac{\boldsymbol{A}'_{(\tau_1,\tau_2)} \mathbf{B}^{-1}_{(\tau_1,\tau_2)} \boldsymbol{C}_{(\tau_1,\tau_2)}}{\sqrt{\boldsymbol{A}'_{(\tau_1,\tau_2)} \mathbf{B}^{-1}_{(\tau_1,\tau_2)} \mathbf{D}_{(\tau_1,\tau_2)} \mathbf{B}^{-1}_{(\tau_1,\tau_2)} \boldsymbol{A}_{(\tau_1,\tau_2)}}}, \tag{8}$$

where

$$\boldsymbol{A}_{(\tau_1,\tau_2)} = \textstyle\sum_{t=\lfloor \tau_1 T \rfloor+1}^{\lfloor \tau_2 T \rfloor} \tilde{x}_{t-1} \tilde{\boldsymbol{z}}_{t-1} \qquad \mathbf{B}_{(\tau_1,\tau_2)} = \textstyle\sum_{t=\lfloor \tau_1 T \rfloor+1}^{\lfloor \tau_2 T \rfloor} \tilde{\boldsymbol{z}}_{t-1} \tilde{\boldsymbol{z}}'_{t-1}$$

$$\boldsymbol{C}_{(\tau_1,\tau_2)} = \textstyle\sum_{t=\lfloor \tau_1 T \rfloor+1}^{\lfloor \tau_2 T \rfloor} \tilde{\boldsymbol{z}}_{t-1} \tilde{s}_t, \qquad \mathbf{D}_{(\tau_1,\tau_2)} = \textstyle\sum_{t=\lfloor \tau_1 T \rfloor+1}^{\lfloor \tau_2 T \rfloor} \tilde{\boldsymbol{z}}_{t-1} \tilde{\boldsymbol{z}}'_{t-1} \hat{u}_t^2,$$

using residuals under the null $\hat{u}_t = \tilde{s}_t$. The $\tilde{}$ denotes the subsample-specific demeaned version

[6] See Section S.2.2 in the Online Supplement of Demetrescu et al. (2022) for a thorough discussion of the multiple regression implementation.

of the variables, i.e. in general $\tilde{y}_t = y_t - \frac{1}{\lfloor \tau_2 T \rfloor - \lfloor \tau_1 T \rfloor - 1} \sum_{i=\lfloor \tau_1 T \rfloor + 1}^{\lfloor \tau_2 T \rfloor} y_i$ etc. Demeaning of the IVX instrument is not needed as discussed in Kostakis et al. (2015), so we let for notational convenience $\tilde{z}_{I,t} = z_{I,t}$.

The full sample test is nested for $\tau_1 = 0$ and $\tau_2 = 1$. Further, as pointed out in Breitung and Demetrescu (2015) and Demetrescu et al. (2022), the 2SLS based test, and therefore the directional 2SLS test in (8), must be used to test against two sided alternatives. We equivalently consider the squared version for which $t^2_{(\tau_1,\tau_2)} \xrightarrow{d} \chi^2(1)$ holds when employing the instruments of (6) and (7), regardless of the predictor persistence; see Demetrescu et al. (2022).

For practical applications, it is a priori unknown where periods of predictability might occur. It is then well-understood that one cannot just pick the first-best τ_1 and τ_2 without consequences. One tentative approach would be to compute the squared test-statistic in (8) for many different subsamples of the data and compare each with the $\chi^2(1)$ marginal critical values. Another approach would be to choose particular subsamples where predictability potentially exists based on theoretical arguments or where other literature has found significant relations previously. In either case we will potentially encounter spurious inference, in the former case because the probability of observing a (spurious) predictive relationship will tend to one as the number of subsamples considered increases which is simply a multiple testing issue, while, in the latter case, the test decisions will be incorrectly sized due to endogenously selected subsamples. To adjust for this problem, a commonly used strategy is to use various sequences of (2SLS-based) tests, and compute the maximum of those tests as an overall predictability test statistic. Since these sequences of statistics are agnostic of the data, the bias introduced by endogenously chosen subsamples is avoided. Focusing on the maximum of the subsample statistics furthermore deals with the issue of multiple testing in a natural manner. For the original 2SLS procedure, Demetrescu et al. (2022) further provide asymptotic distributions for the maximum of forward and backward recursive sequences as well as rolling and double-recursive sequences, which are defined as follows:

- The maximum of the forward recursive test statistic sequence is given by

$$\mathcal{T}^f = \max_{\tau_L \leq \tau \leq 1} t^2_{(0,\tau)}, \tag{9}$$

  with the parameter $\tau_L \in (0, 1)$ being specified by the user. It considers all subsamples from $t = 1, \ldots, \lfloor \tau T \rfloor$ where τ moves from τ_L to 1.

- The maximum of the backward recursive test statistic sequence is given by

$$\mathcal{T}^b = \max_{0 \leq \tau \leq \tau_U} t^2_{(\tau,1)}, \tag{10}$$

  with the parameter $\tau_U \in (0, 1)$ being specified by the user. It considers all subsamples from $t = \lfloor \tau T \rfloor + 1, \ldots, T$ where τ moves from 0 to τ_U.

- The maximum of the rolling window test statistic sequence given by

$$\mathcal{T}^r = \max_{0 \leq \tau \leq 1 - \Delta\tau} t^2_{(\tau,\tau+\Delta\tau)}, \tag{11}$$

with the parameter $\Delta\tau \in (0,1)$ being specified by the user to control the window fraction $\lfloor \Delta\tau T \rfloor$. It considers all subsamples from $t = \lfloor \tau T \rfloor + 1, \ldots \lfloor \tau T + \Delta\tau T \rfloor$, where τ moves from 0 to $1 - \Delta\tau$.

- The maximum of the double-recursive test statistic sequence is given by

$$\mathcal{T}^d = \max_{0 \leq \tau_1, \tau_2 \leq 1, \tau_2 - \tau_1 \geq \Delta\tau} t^2_{(\tau_1, \tau_2)}, \tag{12}$$

with the parameter $\tau_\Delta \in (0,1)$ being specified by the user. It entails the rolling window test statistic above if computed for all rolling window fractions which are between $\Delta\tau$ and, 1 as well as the backward and forward recursive test statistic across all $\tau_L \geq \Delta\tau$ and $\tau_U \leq \Delta\tau$.

These statistics exhibit nonstandard limiting distributions, in spite of statistics for any given subsample being indeed chi-squared distributed. In particular they depend on the volatility path of the data; see Demetrescu et al. (2022) again. Therefore we adapt the fixed-regressor wild bootstrap algorithm advocated there to the directional predictability testing setup as follows.

1. Build wild innovations $s_t^* = \tilde{s}_t R_t$ with $\tilde{s}_t$ being the demeaned version of the sample data of s_t and the bootstrap multipliers are $R_t \overset{iid}{\sim} N(0,1)$.

2. Replace the original data $\left\{s_t, x_{t-1}, \boldsymbol{z}'_{t-1}\right\}_{t=1,\ldots,T}$ by the fixed-regressor bootstrap data $\left\{s_t^*, x_{t-1}, \boldsymbol{z}'_{t-1}\right\}_{t=1,\ldots,T}$, so that the null is imposed. Construct the bootstrap analogue of $\mathcal{T}^s$ and denote it as $\mathcal{T}^{s*}$ for $s = f, b, d, r$.

3. Define the bootstrap p-value as $p^*(\mathcal{T}^s) = 1 - G^*(\mathcal{T}^s)$ where $G^*(\cdot)$ denotes the cumulative distribution function (cdf) of $\mathcal{T}^{s*}$. To obtain a good approximation to the unknown cdf $G^*(\cdot)$ steps 1 and 2 must be repeated for a large number of times. Denote the number of bootstrap replications as B. The true cdf $G^*(\cdot)$ can then be approximated by the empirical bootstrap cdf based on the B values of $\mathcal{T}^{s*}$ obtained through repeating steps 1 and 2 B times. Note that the variables $\{R_t\}_{t=1,\ldots,T}$ must be independent across all of the B bootstrap replications.

4. Obtain bootstrap critical value at significance level κ, $\mathcal{T}_\kappa^{s*}$, as $1-\kappa$ quantile of the respective bootstrap distribution. Reject the null hypothesis of no predictability in at nominal significance level κ if $p^*(\mathcal{T}^s) \leq \kappa$.

As we argue below, the bootstrap is able to correctly replicate the relevant limiting null distribution, leading to inference which is asymptotically robust to persistence.

Proposition 1 *For any K under assumptions 1–3 and the null hypothesis $\boldsymbol{\beta} = \mathbf{0}$, we have as $T \to \infty$ that $\mathrm{P}^*\left(\mathcal{T}^s \geq \mathcal{T}_\kappa^{s*}\right) \xrightarrow{p} \kappa$, for each of $s = f, b, d, r$, where P^* denotes probability conditional on the original sample $\{y_t, \boldsymbol{x}_{t-1}\}_{t=1,\ldots,T}$.*

To confirm the reliability of our proposed testing scheme we ran Monte Carlo simulations to check for size and power properties in the following section.

3 Finite sample behaviour

We now study the behaviour of the proposed procedure in samples of realistic size. To the best of our knowledge there is no alternative test procedure that deals with predictive regression endogeneity in a binary dependent variable setup, and even less so in subsamples, so we consider as benchmark a naïve subsample logit procedure. This naïve logit procedure is based on an off-the-shelf bootstrap implementation, where we replace $t_{(\tau_1,\tau_2)}$ by the standard logit test statistic of no predictability to compute $\mathcal{T}^r_{logit} = \max_{0\leq\tau\leq 1-\Delta\tau} t^2_{(\tau,\tau+\Delta\tau),logit}$. For the logit bootstrap, the wild bootstrap residuals are created as iid random Bernoulli variables using the sample mean of s_t as parameter to ensure that s_t are binary and unpredictable – unlike for the auxiliary regression. This does not result in robustness to the degreee of persistence of the putative predictor; at best, it may help with the issue of multiple testing, and we only consider it due to lack of a valid alternative inference method. For power comparisons we consider an unfeasible, size-corrected version of the subsample logit procedure, where the correct critical values for $\max_{0\leq\tau\leq 1-\Delta\tau} t^2_{(\tau,\tau+\Delta\tau),logit}$ are obtained via Monte Carlo simulations. Even if the infeasible procedure is not available in practice as the actual predictor persistence is uncertain, the resulting power comparison is nevertheless informative about the theoretical price paid for robust inference.

3.1 Simulation setup

Our simulation experiment is based on different samples sizes $T \in \{250, 500, 2000, 5000\}$ to gauge the behaviour in empirically relevant sample sizes when dealing with daily in addition to monthly stock data. All simulations for size and power are based on the squared directional 2SLS test from (8) using the previously outlined fixed regressor wild bootstrap algorithm. Further, we only consider the rolling window approach (11) as it offers simple interpretability and flexibility; see also the discussion in Demetrescu et al. (2022). The instruments are created as stated in (6) and (7) using $\eta = 0.95$ and $a = 1$ as IVX parameters for the first instrument following Demetrescu et al. (2022). Except – as discussed – for the type-I instrument $z_{I,t}$, all variables are demeaned. Further, we employ a finite sample correction factor in the fashion of Kostakis et al. (2015) as also used by Demetrescu et al. (2022). To this end, a weighted demeaning term is added to the matrix $\mathbf{D}$ in the denominator of (8) as follows. Let $\hat{\epsilon}_t$ denote the residuals of the linear regression of s_t on x_{t-1} and an intercept. Then, we replace the matrix $\mathbf{D}$ by the corrected version $\mathbf{D} - \boldsymbol{\Xi}$ where $\boldsymbol{\Xi} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} T\overline{z}^2_{I,t-1}(\hat{\sigma}^2_\epsilon - \hat{\sigma}^2_{\epsilon v}\hat{\sigma}^{-2}_v)$, with $\overline{z}^2_{I,t-1} = T^{-1}\sum_{t=1}^T z^2_{I,t-1}$ and $\hat{\sigma}_{\epsilon v}$ and $\hat{\sigma}^2_v$ representing estimates of the long run covariance between ϵ_t and v_t and of the long run variance of v_t, where the estimation of these terms in $\boldsymbol{\Xi}$ follows Kostakis et al. (2015). The correction is again implemented subsample-wise, i.e. in terms of $\mathbf{D}_{(\tau_1,\tau_2)} - \boldsymbol{\Xi}_{(\tau_1,\tau_2)}$. The nominal significance level is $\kappa = 5\%$ throughout all simulations and the number of bootstrap replications is set to $B = 399$. For all simulations regarding the test size the number of Monte Carlo simulations is $M = 2500$ and for the simulations regarding the test power it is set to $M = 1000$. The intercepts are set to zero, i.e. $\gamma_0 = \mu_x = 0$, w.l.o.g. Innovations are drawn from a joint normal distribution with correlation $\varphi = \mathrm{corr}(u_t, v_t)$. Predictors are modelled as AR(1) leading to the following DGP:

$$x_t = \rho x_{t-1} + v_t, \quad x_0 = 0, \quad \text{and} \quad \rho = 1 - \frac{c}{T} \tag{13}$$

and

$$(u_t \quad v_t)' \overset{iid}{\sim} N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\sigma_{u,t}^2 & \varphi\sigma_{u,t}\sigma_{v,t}\\ \varphi\sigma_{u,t}\sigma_{v,t} & \sigma_{v,t}^2\end{pmatrix}\right). \tag{14}$$

3.2 Size properties

For the empirical rejection rates under the null hypothesis, we impose no predictability in (1) by setting w.l.o.g. $f(\alpha) = 0$, such that the direction of returns are drawn as

$$s_t = \mathbb{I}\left(y_t \geq 0\right) = \mathbb{I}\left(u_t \geq 0\right). \tag{15}$$

For the predictors we employ different degrees of persistence by varying the mean reversion $c \in \{-5, 0, 5, 20, 0.5T\}$. A value of $c = -5$ corresponds to a mildly explosive, yet local-to-unity process. The value of $c = 0$ represents an exactly integrated predictor, while $c = 5$ and $c = 20$ model an near integrated predictor, where $c = 5$ yields higher persistence than $c = 20$. Finally, a value of $c = 0.5T$ always corresponds to a fixed autoregressive coefficient of $\rho = 0.5$ regardless of the sample size and therefore corresponds to a stationary predictor. The results are reported for a value of $\varphi = 0$, which induces no endogeneity and would lead to standard inference with a plain t-test, and for $\varphi = -0.9$, as well as $\varphi = -0.95$ which are typical values in the literature of predictive regression models for stock returns; see e.g. Campbell and Yogo (2006). Such contemporaneous correlation induces size distortions when combined with highly persistent predictors run on plain t-test with standard critical values. Further, three different DGPs are considered to model the empirically relevant phenomenon of heteroskedasticity:

$$\text{DGP1:} \quad \sigma_{u,t}^2 = \sigma_{v,t}^2 = 1 \tag{16}$$

$$\text{DGP2:} \quad \sigma_{u,t}^2 = \sigma_{v,t}^2 = \mathbb{I}(t \leq \lfloor 0.5T \rfloor) + 4\mathbb{I}(t > \lfloor 0.5T \rfloor) \tag{17}$$

$$\text{DGP3:} \quad \sigma_{i,t}^2 = (1 - \theta_1 - \theta_2) + \theta_1 i_{t-1}^2 + \theta_2 \sigma_{i,t-1}^2 \quad \text{for} \quad i = u, v. \tag{18}$$

DGP1 represents the plain homoskedastic case while DGP2 models a simultaneous upward change in the second half of the time series in the innovation variance reflecting a simple form of unconditional volatility change. The factor of four for the upward shift in the variance is chosen to be the same as in Demetrescu et al. (2022). To model the stylized fact of volatility clustering DGP3 generates GARCH(1,1) innovations, initialized with the unconditional variance, where we choose $\theta_1 = 0.1$ and $\theta_2 = 0.85$. Finally, we consider three different values for the window fraction $\Delta\tau \in \{0.125, 0.25, 0.35\}$. Even shorter window fractions have shown to deliver unsatisfactory results with regards to size and power in our simulation setup; we do not report the figures here but advise against using shorter windows. Since the naïve logit does not control size we report the corresponding results in the appendix only.

The results are displayed in Table 1 for the homoskedastic case and in Tables 2 and 3 for the heteroskedastic cases. First, we note that there is no substantial difference between the results for the homoskedastic and the heteroskedastic cases. Therefore, the test is as expected robust with respect to the introduced heteroskedasticity in our setup. Generally, relatively strong undersizedness is observed in the case $c = 0$ and $c = 5$ in the small sample $T = 250$ when the window fraction is also small with $\Delta\tau = 0.125$. A similar phenomenon is found by Demetrescu et al. (2022), where the level instead of the direction of y_t is considered in a similar simulation setup. The conservativeness is however disappearing with growing sample size T and window fraction $\Delta\tau$, and is virtually not observable anymore for $T = 5000$. A slight exceedance of of the nominal level is observed only for some cases of $\Delta\tau = 0.125$ when $c = -5$ or $c = 0.5T$. All remaining size rates are close to the nominal 5% level and therefore deliver satisfactory rejection rates under the null. Further, we find that with a larger sample T the size rates are converging to the nominal level in the sense that for $T = 5000$ there are only smaller deviations from $\kappa = 5\%$ compared to the smaller values of T. When $T \leq 2000$ the user chosen parameter $\Delta\tau$ to control the window fraction appears to deliver more stable results that are closer to the nominal level when choosing a larger value, i.e. $\Delta\tau = 0.35$ then yields the most reliable results. Most importantly, our finite sample results reveal that the issue of regression endogeneity is resolved as the empirical size is robust with respect to the correlation of the innovations. For all values of φ there is good size control such that every type of predictor regardless of its persistence c can be employed without running into danger of endogeneity. In summary, in case of directional predictability our regression setup and testing scheme leads to properly sized tests yielding a reasonable base for a valid inference.

To emphasize the need for size-controlled testing, examine the rejection rates for the naïve logit counterpart located in the Appendix. There, it can be seen that, for $c \in \{-5, 0, 5, 20\}$ and $\delta \in \{-0.9, -0.95\}$, severe overrejections (ranging from 15% to 55%) occur.

Table 1: Empirical size (nominal: 5%) of subsample-wise directional 2SLS under DGP1: $\sigma^2_{u,t} = \sigma^2_{v,t} = 1$

c		$\varphi = 0$			$\varphi = -0.9$			$\varphi = -0.95$		
	$\Delta\tau$	0.125	0.25	0.35	0.125	0.25	0.35	0.125	0.25	0.35
					$T = 250$					
-5		0.0612	0.0704	0.0736	0.0608	0.0656	0.0728	0.0500	0.0624	0.0612
0		0.0092	0.0084	0.0276	0.0004	0.0108	0.0236	0.0004	0.0124	0.0348
5		0.0056	0.0128	0.0356	0.0000	0.0144	0.0384	0.0000	0.0224	0.0500
20		0.0032	0.0360	0.0532	0.0024	0.0344	0.0564	0.0000	0.0448	0.0536
$0.5T$		0.0736	0.0692	0.0668	0.0668	0.0676	0.0612	0.0548	0.0668	0.0624
					$T = 500$					
-5		0.0492	0.0484	0.0508	0.0428	0.0556	0.0640	0.0464	0.0520	0.0612
0		0.0016	0.0156	0.0284	0.0000	0.0212	0.0384	0.0000	0.0204	0.0444
5		0.0008	0.0176	0.0396	0.0004	0.0248	0.0488	0.0004	0.0292	0.0488
20		0.0024	0.0572	0.0548	0.0044	0.0532	0.0592	0.0008	0.0596	0.0692
$0.5T$		0.0748	0.0616	0.0572	0.0704	0.0648	0.0560	0.0732	0.0592	0.0620
					$T = 2000$					
-5		0.0444	0.0500	0.0672	0.0412	0.0660	0.0720	0.0552	0.0464	0.0588
0		0.0012	0.0248	0.0532	0.0036	0.0328	0.0514	0.0164	0.0444	0.0492
5		0.0048	0.0408	0.0480	0.0128	0.0416	0.0572	0.0144	0.0472	0.0616
20		0.0364	0.0592	0.0508	0.0328	0.0608	0.0532	0.0404	0.0644	0.0660
$0.5T$		0.0600	0.0468	0.0512	0.0628	0.0684	0.0572	0.0572	0.0508	0.0508
					$T = 5000$					
-5		0.0554	0.0480	0.0536	0.0276	0.0552	0.0596	0.0496	0.0500	0.0580
0		0.0068	0.0444	0.0508	0.0092	0.0572	0.0596	0.0104	0.0504	0.0600
5		0.0224	0.0428	0.0444	0.0260	0.0628	0.0620	0.0248	0.0700	0.0532
20		0.0480	0.0604	0.0500	0.0512	0.0696	0.0556	0.0516	0.0660	0.0632
$0.5T$		0.0700	0.0512	0.0596	0.0504	0.0680	0.0576	0.0588	0.0600	0.0532

Table 2: Empirical size (nominal: 5%) of subsample-wise directional 2SLS under DGP2: $\sigma^2_{u,t} = \sigma^2_{v,t} = \mathbb{I}(t \leq \lfloor 0.5T \rfloor) + 4\mathbb{I}(t > \lfloor 0.5T \rfloor)$

c		$\varphi = 0$			$\varphi = -0.9$			$\varphi = -0.95$		
	$\Delta\tau$	0.125	0.25	0.35	0.125	0.25	0.35	0.125	0.25	0.35
					$T = 250$					
-5		0.0608	0.0532	0.0532	0.0472	0.0624	0.0624	0.0452	0.0640	0.0496
0		0.0060	0.0116	0.0212	0.0000	0.0104	0.0388	0.0004	0.0104	0.0284
5		0.0044	0.0156	0.0320	0.0004	0.0152	0.0416	0.0000	0.0232	0.0548
20		0.0044	0.0388	0.0576	0.0024	0.0396	0.0560	0.0004	0.0524	0.0448
$0.5T$		0.0604	0.0632	0.0576	0.0588	0.0732	0.0648	0.0680	0.0676	0.0552
					$T = 500$					
-5		0.0476	0.0500	0.0484	0.0436	0.0548	0.0652	0.0380	0.0652	0.0656
0		0.0008	0.0164	0.0332	0.0008	0.0196	0.0444	0.0000	0.0212	0.0496
5		0.0004	0.0244	0.0472	0.0012	0.0284	0.0480	0.0004	0.0296	0.0596
20		0.0080	0.0404	0.0544	0.0068	0.0524	0.0592	0.0008	0.0644	0.0592
$0.5T$		0.0732	0.0560	0.0572	0.0736	0.0584	0.0628	0.0760	0.0592	0.0692
					$T = 2000$					
-5		0.0504	0.0600	0.0612	0.0304	0.0528	0.0608	0.0348	0.0672	0.0628
0		0.0008	0.0372	0.0524	0.0056	0.0324	0.0504	0.0128	0.0404	0.0524
5		0.0060	0.0444	0.0580	0.0100	0.0512	0.0540	0.0104	0.0468	0.0512
20		0.0356	0.0604	0.0488	0.0332	0.0520	0.0624	0.0396	0.0652	0.0452
$0.5T$		0.0704	0.0484	0.0524	0.0592	0.0540	0.0544	0.0596	0.0476	0.0600
					$T = 5000$					
-5		0.0556	0.0536	0.0516	0.0376	0.0600	0.0564	0.0364	0.0644	0.0608
0		0.0152	0.0456	0.0500	0.0188	0.0660	0.0560	0.0312	0.0552	0.0516
5		0.0324	0.0552	0.0552	0.0304	0.0532	0.0536	0.0256	0.0500	0.0556
20		0.0496	0.0608	0.0532	0.0696	0.0552	0.0564	0.0496	0.0624	0.0500
$0.5T$		0.0604	0.0516	0.0524	0.0604	0.0580	0.0576	0.0640	0.0512	0.0564

Table 3: Empirical size (nominal: 5%) of subsample-wise directional 2SLS under DGP3: $\sigma^2_{i,t} = 0.05 + 0.1 i^2_{t-1} + 0.85\sigma^2_{i,t-1}$ for $i = u, v$

c		$\varphi = 0$			$\varphi = -0.9$			$\varphi = -0.95$		
	$\Delta\tau$	0.125	0.25	0.35	0.125	0.25	0.35	0.125	0.25	0.35
					$T = 250$					
-5		0.0512	0.0648	0.0688	0.0464	0.0640	0.0644	0.0464	0.0668	0.0728
0		0.0064	0.0100	0.0264	0.0004	0.0140	0.0340	0.0004	0.0164	0.0280
5		0.0044	0.0120	0.0296	0.0008	0.0148	0.0372	0.0008	0.0204	0.0520
20		0.0032	0.0392	0.0524	0.0004	0.0524	0.0564	0.0012	0.0408	0.0512
$0.5T$		0.0708	0.0680	0.0596	0.0664	0.0704	0.0608	0.0680	0.0712	0.0672
					$T = 500$					
-5		0.0456	0.0504	0.0520	0.0488	0.0552	0.0692	0.0512	0.0496	0.0600
0		0.0016	0.0144	0.0250	0.0008	0.0184	0.0416	0.0004	0.0240	0.0408
5		0.0008	0.0232	0.0492	0.0008	0.0328	0.0524	0.0016	0.0364	0.0552
20		0.0072	0.0432	0.0536	0.0072	0.0544	0.0616	0.0032	0.0520	0.0600
$0.5T$		0.0796	0.0652	0.0588	0.0748	0.0664	0.0556	0.0776	0.0684	0.0564
					$T = 2000$					
-5		0.0380	0.0488	0.0524	0.0396	0.0528	0.0648	0.0440	0.0596	0.0594
0		0.0016	0.0208	0.0388	0.0008	0.0344	0.0624	0.0040	0.0376	0.0548
5		0.0008	0.0412	0.0480	0.0032	0.0500	0.0556	0.0144	0.0544	0.0696
20		0.0424	0.0528	0.0576	0.0416	0.0556	0.0508	0.0388	0.0612	0.0632
$0.5T$		0.0696	0.0568	0.0500	0.0676	0.0508	0.0480	0.0680	0.0672	0.0528
					$T = 5000$					
-5		0.0560	0.0508	0.0532	0.0396	0.0564	0.0604	0.0488	0.0520	0.0540
0		0.0104	0.0520	0.0488	0.0108	0.0452	0.0496	0.0196	0.0412	0.0572
5		0.0224	0.0428	0.0448	0.0380	0.0492	0.0472	0.0392	0.0596	0.0632
20		0.0380	0.0536	0.0512	0.0476	0.0612	0.0564	0.0500	0.0552	0.0556
$0.5T$		0.0604	0.0552	0.0524	0.0532	0.0608	0.0572	0.0532	0.0612	0.0512

3.3 Power

Before examining the findings pertaining to power, let us recall that predictive regressions on stock returns usually exhibit very low signal-to-noise ratios. Therefore, good local power is of importance to detect significant relations; see Phillips and Lee (2013). The DGP is now such that a local alternative holds with an appropriate Pitman drift rate. Setting w.l.o.g. $\alpha = 0$ and noting that, in narrow neighbourhoods of the null, f is approximately linear, the DGP we employ under the alternative is given by

$$s_t = \mathbb{I}\left(\frac{b_{1,t}}{\sqrt{T}} x_{t-1} + u_t \geq 0\right) \quad \text{for } c = 0.5T, \tag{19}$$

and

$$s_t = \mathbb{I}\left(\frac{b_{1,t}}{T} x_{t-1} + u_t \geq 0\right) \quad \text{for } c \in \{0, 20\}, \tag{20}$$

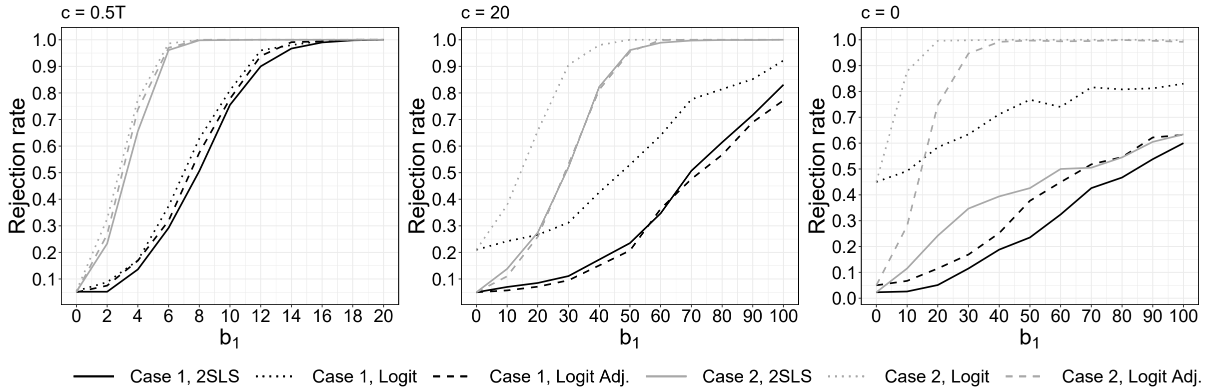
with $b_{1,t} \in \{10, 20, \ldots, 100\}$ when $c \in \{0, 20\}$ and $b_{1,t} \in \{2, 4, , 20\}$ when $c = 0.5T$. Therefore, predictability is now modeled by setting our parameter to $\beta_1 = b_{1,t}/T$ when the predictor is highly persistent and $\beta_1 = b_{1,t}/\sqrt{T}$ when the predictor is weakly persistent. Since the focus is on detecting pockets of predictability, the DGP is such that the slope coefficient is non-zero only for some subsample and zero else. In particular, the power simulations are based on two symmetric cases.

$$\text{Case 1} = \begin{cases} b_{1,t} > 0 & \text{for } 1 \leq t \leq \lfloor 0.2T \rfloor \\ b_{1,t} = 0 & \text{else} \end{cases} \tag{21}$$

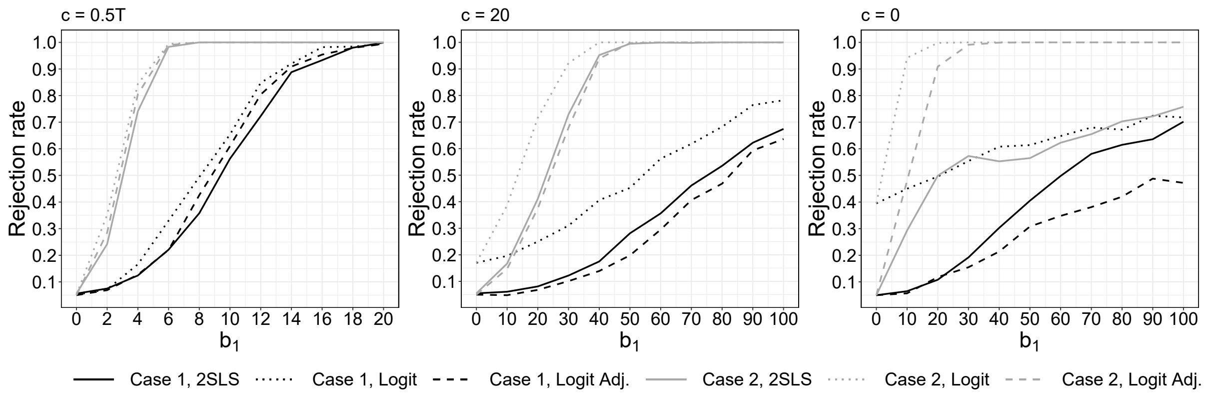
$$\text{Case 2} = \begin{cases} b_{1,t} > 0 & \text{for } \lfloor 0.8T \rfloor \leq t \leq T \\ b_{1,t} = 0 & \text{else} \end{cases} \tag{22}$$

This means predictability is only present in the first or last fifth of the time series and the slope coefficient is zero else. This corresponds to a true window fraction of $\Delta\tau^* = 0.20$, located right at the beginning or end of the time series. Our preliminary simulations showed relatively low discriminatory power for $\Delta\tau = 0.125$, which was expected due to the undersizing occuring in finite samples, especially when $c \in \{0, 20\}$. Therefore, we only report results for a user chosen window fraction $\Delta\tau$ of 0.25 and 0.35 in the power simulations on which we will also build our empirical analysis. To keep the results managable results are only reported for the homoskedastic DGP1 with $\varphi = -0.9$ and a sample size of $T = 1000$.

To adequately compare the power of our directional 2SLS test statistic to the benchmark we have to ensure that they possess the same (asymptotic) size. While size control is given for the directional 2SLS procedure, this is not the case for logit-based subsample tests under persistence. We therefore generate critical values for the logit-based subsample test by means of simulations. Of course, this results in an infeasible procedure, as it requires knowledge of the exact DGP. We ran $M^* = 10000$ replications to draw $\mathcal{T}^r_{logit} = \max_{0 \leq \tau \leq 1-\Delta\tau} (t_{(\tau,\tau+\Delta\tau)})^2_{,logit}$ based on the logit test statistic for each combination of $\Delta\tau$ and c for both cases considered in the power analysis. We then used the $(1-\kappa)\%$ quantile of our test statistic as the true (approximated) critical value under the null to ensure a properly sized test. We display the raw rejection rates as well as the size-corrected power of the logit approach in our plots below.



(a) $\Delta\tau = 0.25$



(b) $\Delta\tau = 0.35$

Figure 2: Local power curves for directional 2SLS (solid), logit (dotted) and size adjusted logit (dashed) with error correlation $\varphi = -0.9$ for various degrees of persistence $c \in \{0, 20, 0.5T\}$.

The results are displayed in Figure 2. First, expectedly, the logit without size adjustment is over-rejecting. It is most noticeable that power is strictly smaller in case 1 compared to case 2 despite the period of predictability being of the same length in both cases. Therefore, the test appears to pick up predictability occuring at the end of the sample better than at the beginning of the sample. This pattern is also found in Demetrescu et al. (2022) who provide possible technical explanations for this behavior: in a nutshell, highly persistent variables, being nonstationary, often have higher variabililty at the end of the sample. Also, in general, it appears that power is larger when $c = 20$ compared to $c = 0$. Only when $\Delta\tau = 0.35$ we find that when $c = 0$ the power is higher than when $c = 20$ under case 1. This perhaps surprising behavior (as the random walk with $c = 0$ has the strongest signal) is due to the fact that the regressor enters the model via an indicator function. This nonlinear transformation has a dampening effect which is apparently stronger for higher degrees of persistence.

Since, in our setup, we choose in both cases the window of predictability to be exactly of size $0.2T$, it is unclear which value of user chosen parameter $\Delta\tau$ will yield the largest power. Theoretically, a window fraction that is smaller will lead to a greater amount of statistics considered in the rolling window sequence over which the maximum is taken. This in turn will lead to a larger critical value to maintain a properly sized test and thereby decrease power. Should the chosen window fraction be larger we will have critical values which are not as strict but at the same time possibly encounter subsamples that contain more observations where no predictability is present and therefore yield a lower value of the corresponding test statistic. Surprisingly, power is slightly larger when $\Delta\tau = 0.35$ than when employing $\Delta\tau = 0.25$ for $c = 0$ and the other way around for $c = 20$, despite the shorter window being closer to the true window fraction. This is likely a finite sample result which emerges from the undersizing that we observed particularly in smaller samples paired with a shorter window fraction. As expected, in the stationary case $c = 0.5T$ we observe a slightly higher power for $\Delta\tau = 0.25$ compared to 0.35 in case 1, yet a similar power in case 2.

To examine a possible cost of our robust estimation in terms of power, we may compare it to the size adjusted power curve of the logit approach. We can see that under stationary predictors with $c = 0.5T$ logit is properly sized and therefore its power coincides with the size adjusted version. However, also the directional 2SLS approach has virtually the same power, regardless of whether predictability occurs at the end or beginning of the time series, captured by employing Case 1 or 2, and for both chosen values of $\Delta\tau$. This means there is no significant loss of power in using our robust approach in such standard scenarios. Let us now turn our attention to the persistent case with $c \in \{0, 20\}$. For $c = 20$ we can see that logit already suffers from severe size distortions with rejection rates of around 20% when $b_1 = 0$ in all cases. Our properly size test still remains to have the same power as the (infeasible) size adjusted logit. In the random walk case of $c = 0$ we observe that the overrejection of the logit approach amplifies even further yielding virtually nonsense inference as rejection rates under the null are over 40%. Here we see for the first time that the logit approach slightly outperforms the directional 2SLS approach in Case 1 and perhaps more so in Case 2 when $\Delta\tau = 0.25$. For $\Delta\tau = 0.35$, however, this only occurs for Case 2 while in Case 1 the directional 2SLS surprisingly dominates the adjusted logit. However, for $c = 0$ the directional 2SLS test showed some undersizing with $\Delta\tau = 0.25$ for smaller T, therefore this comparison is in favor of the logit approach as it is adjusted to have a size of precisely $\kappa = 5\%$. As T grows and the undersizing of our test vanishes we can expect the local power to increase as well. The slight excess of (unattainable) power can also be interpreted as a price to robustify against regression endogeneity, which turns out to be very small and even non existent considering the power for Case 1.

4 Episodes of directional predictability in stock returns

We now analyze real stock data to see if there have been periods of directional predictability of the returns in the past. We consider both monthly-level returns and higher-frequency, daily returns. To examine stock returns at a monthly frequency we use the updated version of the well known data set from Welch and Goyal (2008) looking at returns of the S&P 500. For daily data we consider individual returns from 25 stocks that were part of the DJIA in December 2021. The computation of p-values is based on the subsample version of the squared directional 2SLS test from (8) using the previously outlined fixed regressor wild bootstrap algorithm and rolling window approach similar to Demetrescu et al. (2022). The instruments are created as in (6) and

(7) with $\eta = 0.95$ and $a = 1$. All variables entering the test statistics are demeaned except for the first instrument. We also include the finite-sample correction term introduced in Section 3. The number of bootstrap replication is boosted to $B = 999$ to refine the results. Consequently, p-values can then change in increments of 0.001. The results are reported for the window fractions that showed satisfactory power in the Monte Carlo simulations, $\Delta\tau \in \{0.25, 0.35\}$.

4.1 Monthly directional predictability

First we look at predictability at a monthly frequency where regression endogeneity often is a crucial feature. We use the updated data set of Welch and Goyal (2008) which contains returns of the S&P 500 and several putative predictor variables. Data are available from 1926:12 to 2023:12 yielding $T = 1164$. As dependent variable we look at the sign of the equity premium, i.e. the log returns on the value weighted stock market index including dividends minus the log return of the risk free treasury bill. Single-variable regressions for each of the following predictors, employed in lagged form, are considered. The data is publicly available on the website of Amit Goyal.[7] Table 4 outlines a list of the applied predictors.

Table 4: Predictors for monthly data

1.	dp_{t-1}	log dividend price ratio
2.	dy_{t-1}	log dividend yield
3.	e/p_{t-1}	log earnings price ratio
4.	de_{t-1}	log dividend payout ratio
5.	$svar_{t-1}$	stock variance
6.	bm_{t-1}	book to market ratio
7.	$ntis_{t-1}$	net equity expansion
8.	tbl_{t-1}	treasury bill rate
9.	lty_{t-1}	long-term government bond yield
10.	ltr_{t-1}	long-term government bond rate of return
11.	tms_{t-1}	term spread
12.	dfy_{t-1}	default yield spread
13.	dfr_{t-1}	default return spread
14.	$infl_{t-1}$	inflation

A detailed description of the data can be found in Welch and Goyal (2008).

Table 5 shows the corresponding p-values as well as the period in which the maximum of the subsample test statistics occured for both window fractions $\Delta\tau = 0.35$ and $\Delta\tau = 0.25$. To highlight the advantage of the rolling window approach we also report the p-values for the full sample when $\Delta\tau = 1$ is chosen to serve as a comparison. Additionally, we report the estimated AR(1) coefficient $\hat{\rho}$ of the predictor as well as the estimated correlation between the regressions residuals $\hat{\varphi}$ based on the full sample. Many of the considered predictors exhibit strong persistence with autoregressive coefficents close to unity paired with non-zero correlation of the innovation. In particular the dividend price ratio, the earnings price ratio and the book to market ratio, which are often considered in the literature of predictive regression are expected to suffer from regression endogeneity when employing non-robust estimation techniques. Indeed, a naïve logit

[7] We thank Amit Goyal for making the data available on his website https://sites.google.com/view/agoyal145

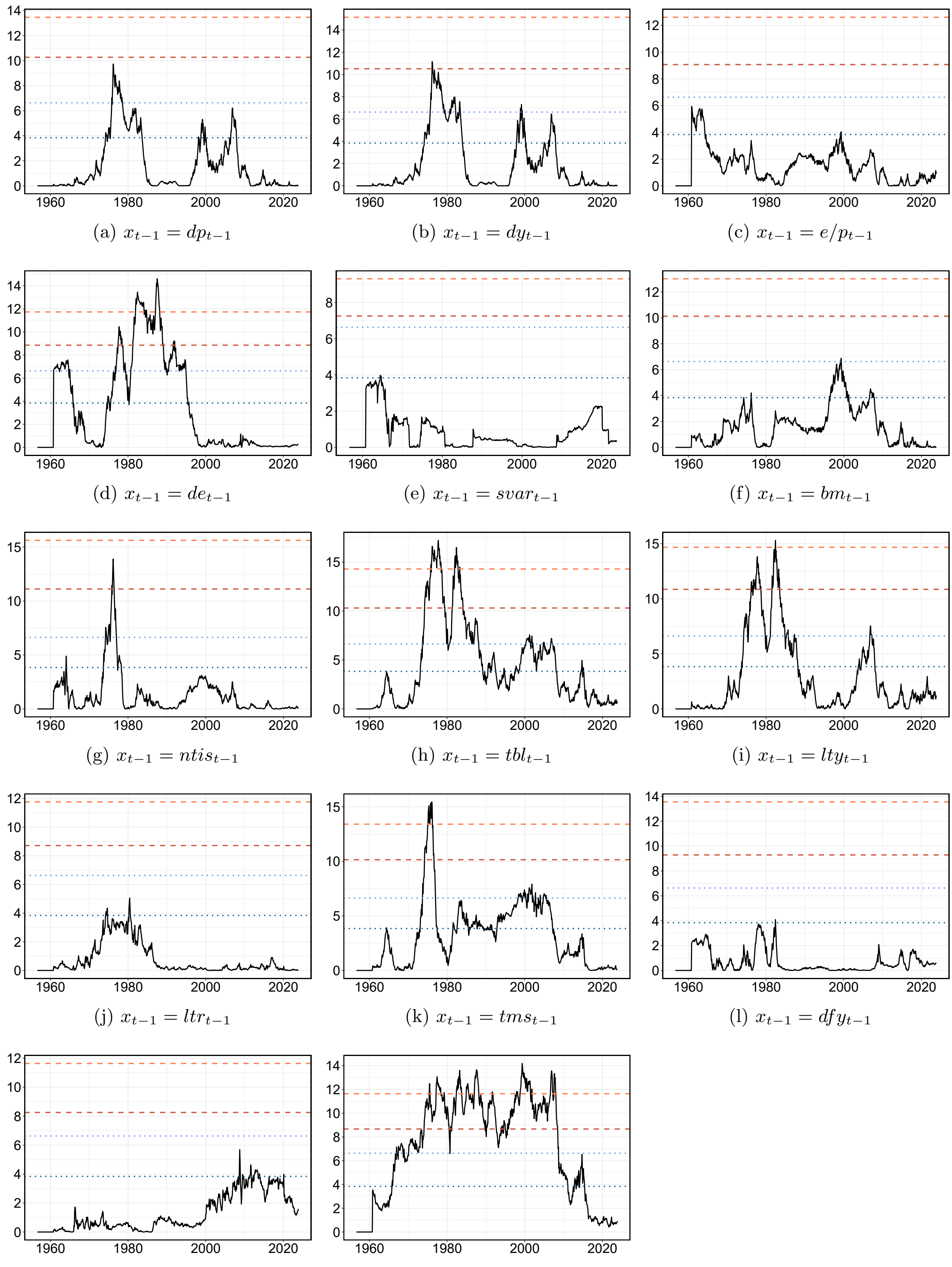
approach as explained in Section 3 yields similar results but generally overstates the significance of the findings, see table 15 in the Appendix.

Table 5: Empirical bootstrapped p-values and period with largest test statistic for updated monthly data from Welch and Goyal (2008), 1926:12 to 2023:12.

x_{t-1}	$\Delta\tau = 0.25$		$\Delta\tau = 0.35$		$\Delta\tau = 1$		$\hat{\rho}$	$\hat{\varphi}$
	p-values	Period	p-values	Period	p-values	Period		
dp_{t-1}	0.070	1973:12 1998:04	0.056	1942:04 1976:04	0.681	1926:12 2023:12	0.994	-0.974
dy_{t-1}	**0.036**	1973:12 1998:04	**0.034**	1942:04 1976:04	0.737	1926:12 2023:12	0.994	-0.067
e/p_{t-1}	0.119	1930:03 1954:07	0.252	1926:12 1960:12	0.901	1926:12 2023:12	0.988	-0.757
de_{t-1}	**0.001**	1953:06 1977:10	**0.002**	1953:08 1987:08	0.731	1926:12 2023:12	0.992	-0.072
$svar_{t-1}$	0.718	1930:10 1955:02	0.485	1930:07 1964:07	0.114	1926:12 2023:12	0.578	-0.095
bm_{t-1}	0.274	1973:12 1998:04	0.198	1965:06 1999:06	0.342	1926:12 2023:12	0.988	-0.810
$ntis_{t-1}$	0.203	1942:04 1966:08	**0.021**	1942:04 1976:04	0.136	1926:12 2023:12	0.982	-0.032
tbl_{t-1}	**0.004**	1953:06 1977:10	**0.004**	1943:11 1977:11	**0.005**	1926:12 2023:12	0.993	-0.052
lty_{t-1}	**0.023**	1953:06 1977:10	**0.009**	1948:07 1982:07	**0.012**	1926:12 2023:12	0.996	-0.093
ltr_{t-1}	0.472	1943:01 1967:05	0.296	1946:07 1980:07	0.733	1926:12 2023:12	0.052	-0.002
tms_{t-1}	**0.002**	1951:02 1975:06	**0.002**	1942:03 1976:03	0.295	1926:12 2023:12	0.964	-0.013
dfy_{t-1}	0.683	1950:08 1974:12	0.556	1948:07 1982:07	0.162	1926:12 2023:12	0.975	-0.248
dfr_{t-1}	0.346	1973:06 1997:10	0.225	1974:11 2008:11	0.218	1926:12 2023:12	-0.105	-0.006
$infl_{t-1}$	**0.001**	1953:10 1978:02	**0.003**	1965:06 1999:06	**0.002**	1926:12 2023:12	0.483	0.026

Notes: Bold entries denote significance at the 5% level. The error correlation $\hat{\varphi}$ and autocorrelation coefficient $\hat{\rho}$ are based on the full sample.

Similar to Demetrescu et al. (2022) we find that when considering the full sample using $\Delta\tau = 1$ the treasury bill rate (tbl_{t-1}), long-term government bond yield (lty_{t-1}) and inflation ($infl_{t-1}$) are significant predictors at the 5% level. However, this only reflects the average relationship over the sample period and is unable to detect putative temporary significance. When using the rolling window approach to analyze subsamples we find several periods where further predictors are statistically significant. For both window fractions $\Delta\tau = 0.35$ and $\Delta\tau = 0.25$ now additionally the dividend yield (dy_{t-1}) rejects the null at the 5% level and the dividend payout ratio (de_{t-1}), net equity expansion ($ntis_{t-1}$), term spread (tms_{t-1}) reject the null even at the 1% level, except for the bond yield which rejects for $\Delta\tau = 0.25$ only at the 5% level. Also tbl_{t-1}, lty_{t-1} as well as $infl_{t-1}$ which were significant at the full sample still remain significant predictors for both applied values of $\Delta\tau$.



(a) $x_{t-1} = dp_{t-1}$ (b) $x_{t-1} = dy_{t-1}$ (c) $x_{t-1} = e/p_{t-1}$

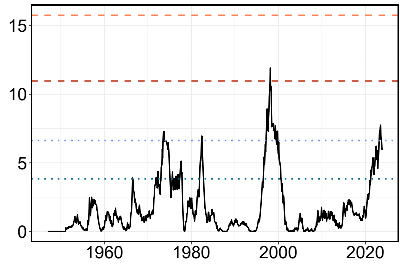
(d) $x_{t-1} = de_{t-1}$ (e) $x_{t-1} = svar_{t-1}$ (f) $x_{t-1} = bm_{t-1}$

(g) $x_{t-1} = ntis_{t-1}$ (h) $x_{t-1} = tbl_{t-1}$ (i) $x_{t-1} = lty_{t-1}$

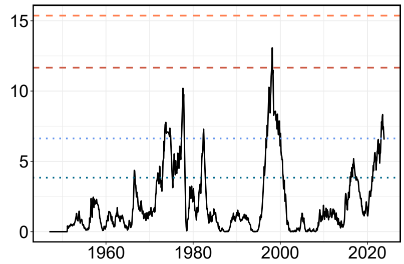
(j) $x_{t-1} = ltr_{t-1}$ (k) $x_{t-1} = tms_{t-1}$ (l) $x_{t-1} = dfy_{t-1}$

(m) $x_{t-1} = dfr_{t-1}$ (n) $x_{t-1} = infl_{t-1}$

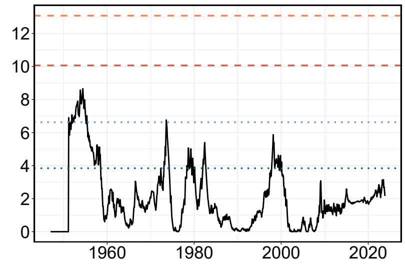
Figure 3: Rolling test-statistic $\mathcal{T}^r$ for $\Delta\tau = 0.35$ at the end point of the corresponding subsample considered. Horizontal lines are 99% and 95% marginal $\chi^2(1)$ (blue, dotted) and robust bootstrap (red, dashed) critical values.



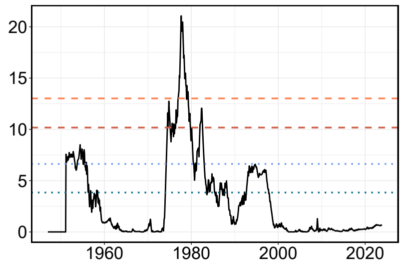
(a) $x_{t-1} = dp_{t-1}$



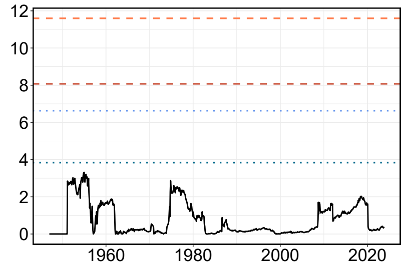
(b) $x_{t-1} = dy_{t-1}$



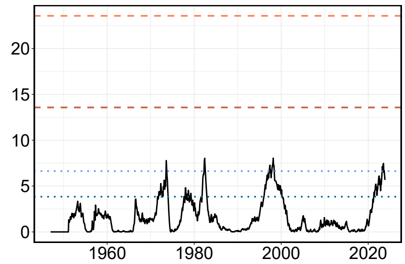
(c) $x_{t-1} = e/p_{t-1}$



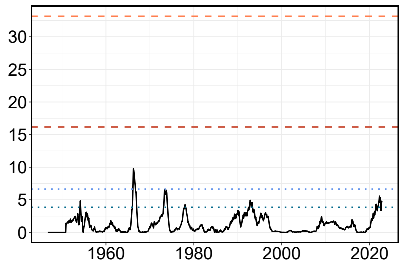
(d) $x_{t-1} = de_{t-1}$



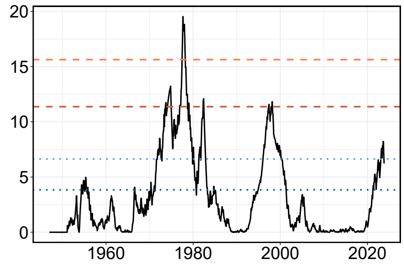
(e) $x_{t-1} = svar_{t-1}$



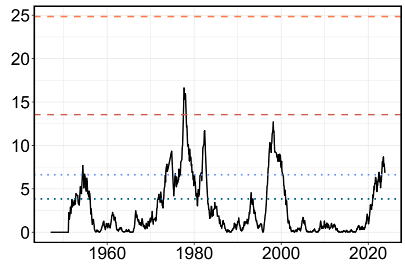
(f) $x_{t-1} = bm_{t-1}$



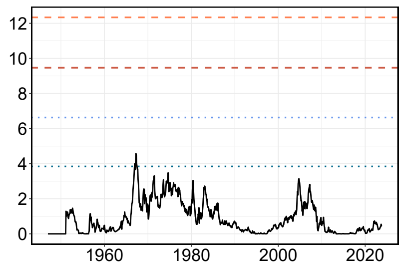
(g) $x_{t-1} = ntis_{t-1}$



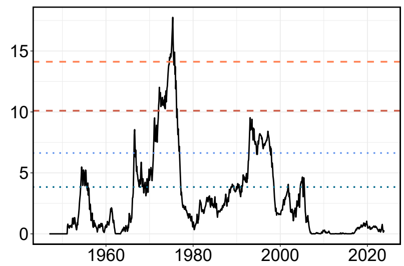
(h) $x_{t-1} = tbl_{t-1}$



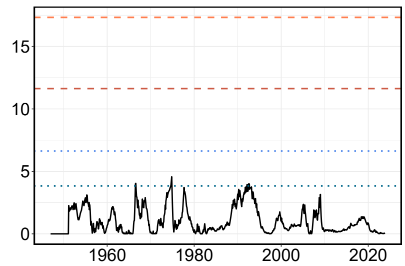
(i) $x_{t-1} = lty_{t-1}$



(j) $x_{t-1} = ltr_{t-1}$

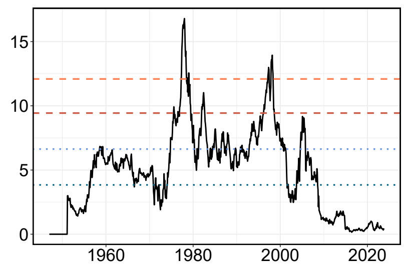


(k) $x_{t-1} = tms_{t-1}$



(l) $x_{t-1} = dfy_{t-1}$

(m) $x_{t-1} = dfr_{t-1}$



(n) $x_{t-1} = infl_{t-1}$

Figure 4: Rolling test-statistic $\mathcal{T}^r$ for $\Delta\tau = 0.25$ at the end point of the corresponding subsample considered. Horizontal lines are 99% and 95% marginal $\chi^2(1)$ (blue, dotted) and robust bootstrap (red, dashed) critical values.

To further explore the time-varying behavior of the relationship between the predictors and the sign of the equity premium, Figures 3 and 4 display the observed test statistics for each considered subsample together with the marginal $\chi^2(1)$ and robust bootstrapped critical values at the 5% and 1% level for both values of $\Delta\tau$. This allows us to illustrate the substantial variation in magnitude occurring over time highlighting the benefit of the rolling window approach. In the plots we mapped the endpoint of the corresponding subsample to its realized test statistic.

First we notice that the bootstrapped critical values which correct for the multiple testing issue are considerably larger than their marginal $\chi^2(1)$ counterparts which would coincide only when $\Delta\tau = 1$. While there would have been rejections of the null for several subsamples when tested individually employing the $\chi^2(1)$ critical values, we see that for many predictors there are only a few pockets of time where the null can still be rejected based on the bootstrap critical values. Looking at Figure 3 we see for example the dividend yield, net equity expansion and term spread can reject the null only during a short period of time which is mostly between 1940 and 1980 but not before and after that.

Generally, we observe that for $\Delta\tau = 0.35$ no significant predictability is found anymore after 1990 except for the inflation rate for which the test statistics only drastically drop just after observations past 2010 enter the subsamples. Also, we notice that in recent times there seems to be very little evidence for predictability as for all predictors the test statistic seem to decrease and remain well below the critical values when including data from 2010 and onward up until the very end of our considered sample period. Summing up, for $\Delta\tau = 0.35$ we find evidence against the null of no predictability for numerous predictors, however most of it being present only in pockets of time which occurred in the past and appeared to have vanished again.

Next, we take a look at what happens when $\Delta\tau = 0.25$ is employed to see how robust these results are. Because of the smaller window fraction the length of the considered subsample shrinks by roughly ten years so that a noticeable change in the corresponding plots is to be expected. However, all predictors which were significant at the 5% level with $\Delta\tau = 0.35$ remain significant except for $ntis_{t-1}$. At the same time for de_{t-1}, tbl_{t-1}, lty_{t-1} and tms_{t-1} the period with highest predictability for $\Delta\tau = 0.25$ is a subperiod of the period with highest predictability when $\Delta\tau = 0.35$ meaning we can pin down the pocket of highest predictability further. For the remaining predictors dy_{t-1} and $infl_{t-1}$ the periods have changed substantially. Especially in the noisy periods of no predictability the magnitude of the test statistics differs compared to the larger window yielding quite a different picture. For many predictors, however, the peaks of the curves seem to remain at a similar pocket in time. We also observe a general pattern of multiple smaller peaks followed by periods of rather low values of the test statistic. The smaller subsamples appear to display more volatility over time which, however, is a general feature of a smaller window as single observations have a larger relative impact on the rolling test statistic. Yet, Figure 4 further reveals that for dp_{t-1}, dy_{t-1}, bm_{t-1}, tbl_{t-1} and lty_{t-1} there seems to be a potential rise of the test statistic, already exceeding the marginal $\chi^2(1)$ critical value which is unlike the case for which $\Delta\tau = 0.35$. This suggests a potential development of predictability in the aftermath of the inflation surge in 2022.

4.2 Daily directional predictability

Similar to Becker and Leschinski (2018) and Gülmez (2023) we look at stocks which were part of the DJIA in December 2021. We choose the evaluation period from 1987:01 to 2021:12 due to data availability, which ensures that we have the same sample size of $T \approx 8822$ for all stocks. Then values of $\Delta\tau = 0.35$ and $\Delta\tau = 0.25$ yield a subsample size of 3088 and 2206 trading days which translates to roughly 12.3 and 8.6 years, respectively. Because data was not available yet at 1987:01 we left out Cisco Systems Inc., Dow Inc., Goldman Sachs Group, Salesforce.com Inc. and Visa Inc. Class A from our analysis. Therefore, only 25 out of the 30 stocks which were part of the DJIA are included. Table 6 lists the analyzed stocks with their Refinitiv Instrument Code (RIC). All data is taken from the Refinitiv Datastream database.

Table 6: Analyzed stocks from the DJIA in December 2021

MMM	3M Company	**JPM**	JPMorgan Chase & Co.
AXP	American Express Co.	**MCD**	McDonald's Corp.
AMGN	Amgen Inc.	**MRK**	Merck & Co. Inc.
AAPL	Apple Inc.	**MSFT**	Microsoft Corp.
BA	Boeing Co.	**NKE**	NIKE Inc. Class B
CAT	Caterpillar Inc.	**PG**	Procter & Gamble Co.
CVX	Chevron Corp.	**TRV**	Travelers Companies Inc.
KO	Coca-Cola Co.	**UNH**	UnitedHealth Group Inc.
HD	Home Depot Inc.	**VZ**	Verizon Communications Inc.
HON	Honeywell International	**WBA**	Walgreens Boots Alliance Inc.
INTC	Intel Corp.	**WMT**	Walmart Inc.
IBM	International Business Machines	**DIS**	Walt Disney Co.
JNJ	Johnson & Johnson		

When dealing with daily data, the existent literature provides less guidance on putative predictors. The typical predictors such as valuation ratios like the dividend price, price earnings or book to market ratio are not suited for higher frequencies as those variables either exhibit no meaningful variation anymore or are simply not available unless one interpolates between the quarters in which those are usually changing. The same holds true for other macroeconomic predictors such as the treasury bill rate and inflation rate. Also other potential predictors like credit ratings, consumer sentiment or analyst recommendations, which are often available only monthly, quarterly, annually or sometimes just sporadically are not immediately suitable on a daily frequency. Therefore, predictors need to be chosen more carefully. They are required to have meaningful variation on a daily basis and to preferably be easily accessible. Based on that, Table 7 shows the ten predictors we use in the empirical analysis, where they are employed in lagged form. To understand the computation of the predictors let C_t,H_t and L_t denote the close, high and low price of a certain stock at time t, respectively. We then denote $H_{t,n}$ as the highest and $L_{t,n}$ as the lowest price a certain stock has reached within the last n trading days including time t. Finally, V_t is the trade volume of the stock at time t. Based on that we can define the predictors using simple formulas. The average number of rising periods in the last five days given by $nrp_t = \frac{1}{5}\sum_{i=1}^{5} \mathbb{I}\left(s_{t-i+1} = 1\right)$. The five periods stock level moving average is $ma_t = \sum_{i=1}^{5} \frac{C_{t-i+1}}{5}$. Divergence between the current price and moving average is computed as $dcp_t = \frac{C_t - MA_t}{MA_t}$. The on balance volume is calculated as $obv_t = obv_{t-1} + \mathbb{I}(C_t \leq C_{t-1})V_t - \mathbb{I}(C_t > C_{t-1})V_t$ and initial value $obv_0 = V_0$. Realized volatility based on high-low variance is $hlv_t = (\log(H_t) - \log(L_t))^2$. The stochastic oscillator is given by $sto_t = \frac{C_t - L_{t,5}}{H_{t,5} - L_{t,5}}$. In some rare cases where $L_{t,5} = H_{t,5}$ the corresponding value of the stochastic oscillator is set to zero to avoid division by zero. The employed predictors can be classified as naïve (y_{t-1} and s_{t-1}), momentum-based (nrp_{t-1}, ma_{t-1}, dcp_{t-1} and sto_{t-1}), volume-based (obv_{t-1} and vc_{t-1}) as well as volatility-based (y^2_{t-1} and hlv_{t-1}) predictors. Many of the above predictors were also considered in Qiu and Song (2016) and Becker and Leschinski (2018), who also analyze directional predictability of daily stock returns and reported significant predictive relations. We summarize their properties in Tables 14 and 15 in the Appendix. We note that, although some of them are quite persistent, endogeneity is reduced at a daily level compared to the monthly level.

Table 7: Predictors for daily data

1.	y_{t-1}	level of stock return
2.	s_{t-1}	sign of stock return
3.	nrp_{t-1}	average number of rising periods in last five days
4.	ma_{t-1}	five period stock level moving average
5.	dcp_{t-1}	divergence current price and moving average
6.	obv_{t-1}	on balance volume
7.	vc_{t-1}	volume change
8.	y^2_{t-1}	volatility based on squared daily return
9.	hlv_{t-1}	volatility based on high-low variance
10.	sto_{t-1}	stochastic oscillator

Table 8 and 9 indicate the significance of the above predictors, concretely the corresponding p-values as well as the period in which the maximum of the subsample test statistics occurred for each window fraction $\Delta\tau = 0.35$ and $\Delta\tau = 0.25$. Again, we find rejections of the null in several cases. We have a total 250 p-values for $\Delta\tau = 0.35$ and $\Delta\tau = 0.25$ each, of which 79 and 64, respectively, are at the 5% level or below. Those cases are bold faced and are considered as significant rejections of the null for the remaining of this paper. Looking at the predictors, particularly many rejections are found for the level of the stock return (y_{t-1}), the five period stock level moving average (ma_{t-1}), the divergence of current price and moving average (dcp_{t-1}) and the stochastic oscillator (sto_{t-1}) yielding 12,12, 17 and 16 cases, respectively, for $\Delta\tau = 0.35$ and 8,8, 17 and 17 cases for $\Delta\tau = 0.25$. For the sign of the stock return (s_{t-1}), the average number of rising periods in last five days (nrp_{t-1}) and on balance volume (obv_{t-1}) there are between 3 and 7 rejections of the null for both windows considered. For the realized volatility based on high-low variance (hlv_{t-1}) we only find one rejection at the 5% level for both window fractions. Finally, for the volume change (vc_{t-1}) we only find one rejection when $\Delta\tau = 0.35$ and the realized volatility based on squared daily returns (y^2_{t-1}) yields no significant results. In general, we observe generally slightly more rejections for $\Delta\tau = 0.35$ compared to $\Delta\tau = 0.25$ for each predictor. However, the amount of stocks for which a predictor is significant for $\Delta\tau = 0.35$ is either the same as for $\Delta\tau = 0.25$ including dcp_{t-1}, hlv_{t-1} and y^2_{t-1} or only marginally different by at most four stocks so that both windows yield similar results in terms of the number of rejections. Studying table 8 and 9 further reveals that 56 out of the 64 significant cases for $\Delta\tau = 0.25$ are also significant for $\Delta\tau = 0.35$ strengthening the robustness of the results.

Table 8: Empirical bootstrapped p-values for daily stock data with $\Delta\tau = 0.35$

	y_{t-1}	s_{t-1}	nrp_{t-1}	ma_{t-1}	dcp_{t-1}	obv_{t-1}	vc_{t-1}	y^2_{t-1}	hlv_{t-1}	sto_{t-1}
MMM	**0.021**	0.117	0.248	0.125	**0.001**	0.613	0.819	0.494	0.059	**0.001**
Start	30.06.92	18.02.09	20.03.91	16.03.94	03.09.87	07.06.91	08.01.88	04.11.05	14.10.88	07.03.91
End	10.06.05	02.02.22	27.02.04	02.03.07	02.08.00	17.05.04	05.12.00	23.10.18	20.09.01	13.02.04
AXP	**0.000**	**0.000**	**0.001**	**0.011**	**0.000**	**0.015**	0.695	0.767	0.327	**0.001**
Start	18.12.01	06.09.01	05.07.91	15.10.92	16.10.87	15.10.92	14.06.91	13.02.89	12.10.95	19.02.97
End	04.12.14	28.08.14	15.06.04	27.09.05	14.09.00	27.09.05	24.05.04	18.01.02	29.09.08	08.02.10
AMGN	0.279	0.095	**0.009**	**0.028**	**0.000**	0.274	0.142	0.249	0.069	**0.003**
Start	18.03.08	09.06.00	09.10.92	21.04.98	22.03.94	14.01.98	11.05.07	28.08.98	31.05.95	09.12.93
End	04.03.21	03.06.13	21.09.05	07.04.11	08.03.07	03.01.11	28.04.20	17.08.11	15.05.08	21.11.06
AAPL	0.290	0.883	0.930	0.607	0.085	0.622	0.259	0.853	0.972	0.084
Start	22.12.97	28.07.98	06.12.90	02.02.93	08.05.90	18.12.03	23.06.05	28.10.10	29.10.08	16.05.94
End	09.12.10	14.07.11	12.11.03	12.01.06	14.04.03	02.12.16	08.06.18	16.10.23	14.10.21	30.04.07
BA	0.153	0.170	0.274	0.223	0.319	0.297	**0.039**	0.701	0.563	0.084
Start	23.11.01	22.10.99	12.02.93	01.12.92	16.02.94	10.12.10	10.11.05	08.09.92	03.04.07	23.11.01
End	10.11.14	10.10.12	26.01.06	10.11.05	02.02.07	29.11.23	29.10.18	18.08.05	19.03.20	10.11.14
CAT	0.809	0.087	0.961	0.500	0.193	0.659	0.555	0.496	0.206	0.510
Start	26.11.08	21.04.87	19.04.94	14.05.04	31.01.94	19.06.09	20.07.09	26.03.09	13.03.09	19.10.92
End	12.11.21	17.03.00	04.04.07	02.05.17	17.01.07	06.06.22	06.07.22	11.03.22	28.02.22	29.09.05
CVX	**0.005**	0.190	**0.013**	0.105	**0.003**	0.308	0.828	0.419	0.683	**0.005**
Start	06.11.07	06.11.07	29.03.94	03.10.88	02.01.96	23.02.87	21.05.08	15.10.87	22.12.10	24.01.95
End	22.10.20	22.10.20	15.03.07	31.08.01	16.12.08	20.01.00	07.05.21	13.09.00	11.12.23	09.01.08
KO	0.219	0.326	0.881	**0.003**	**0.029**	0.075	0.110	0.081	0.330	0.060
Start	09.10.08	20.11.01	20.11.09	21.07.88	09.10.08	21.02.90	03.12.91	07.06.07	24.01.00	04.09.87
End	27.09.21	06.11.14	08.11.22	20.06.01	27.09.21	29.01.03	11.11.04	22.05.20	14.01.13	03.08.00
HD	0.989	0.575	0.375	**0.017**	0.550	0.179	0.395	0.260	0.962	0.607
Start	26.01.06	08.12.92	20.07.92	27.10.89	30.07.92	30.01.90	22.06.92	13.01.87	26.03.07	04.12.92
End	14.01.19	17.11.05	29.06.05	04.10.02	12.07.05	07.01.03	02.06.05	09.12.99	11.03.20	15.11.05
HON	0.269	0.887	0.920	0.558	0.080	0.629	0.277	0.840	0.973	0.105
Start	22.12.97	28.07.98	06.12.90	02.02.93	08.05.90	18.12.03	23.06.05	28.10.10	29.10.08	16.05.94
End	09.12.10	14.07.11	12.11.03	12.01.06	14.04.03	02.12.16	08.06.18	16.10.23	14.10.21	30.04.07
INTC	**0.037**	0.358	0.372	**0.037**	0.278	0.059	0.792	0.292	0.403	0.142
Start	03.09.08	02.01.87	04.10.96	23.11.88	03.06.97	28.08.00	24.12.97	17.11.09	20.08.08	14.09.98
End	19.08.21	30.11.99	23.09.09	30.10.01	21.05.10	20.08.13	14.12.10	03.11.22	06.08.21	31.08.11
IBM	0.104	0.395	0.554	0.108	0.303	**0.001**	0.353	0.470	0.078	0.134
Start	02.06.93	13.10.08	07.01.08	14.04.94	16.06.97	22.07.96	14.05.98	20.12.95	03.10.95	16.06.97
End	15.05.06	29.09.21	21.12.20	30.03.07	04.06.10	09.07.09	03.05.11	05.12.08	18.09.08	04.06.10
JNJ	0.407	**0.034**	0.382	**0.003**	**0.004**	**0.000**	0.558	0.246	0.296	**0.029**
Start	01.04.02	24.04.07	01.05.92	29.03.94	18.03.93	07.10.93	01.09.10	29.03.94	11.04.94	17.08.92
End	17.03.15	08.04.20	13.04.05	15.03.07	01.03.06	20.09.06	21.08.23	15.03.07	27.03.07	28.07.05
JPM	**0.001**	**0.001**	0.076	**0.001**	**0.021**	0.407	0.438	0.531	0.651	**0.031**
Start	09.04.09	02.08.05	03.04.98	14.10.92	17.03.09	15.05.07	06.03.02	08.01.88	18.11.87	14.04.05
End	25.03.22	19.07.18	23.03.11	26.09.05	02.03.22	30.04.20	20.02.15	05.12.00	17.10.00	02.04.18
MCD	**0.038**	0.115	**0.040**	0.122	**0.000**	0.160	0.544	0.629	0.644	**0.010**
Start	10.10.08	29.12.09	03.09.87	01.11.89	16.10.87	01.11.89	05.01.11	27.02.87	27.02.87	08.01.87
End	28.09.21	14.12.22	02.08.00	09.10.02	14.09.00	09.10.02	22.12.23	26.01.00	26.01.00	06.12.99
MRK	**0.018**	0.524	0.307	0.154	**0.003**	0.303	0.835	0.768	0.985	0.062
Start	25.03.08	25.04.08	07.06.93	04.08.94	25.04.08	07.01.92	27.03.07	10.10.08	27.10.89	26.03.08
End	10.03.21	13.04.21	18.05.06	20.07.07	13.04.21	15.12.04	12.03.20	28.09.21	04.10.02	11.03.21
MSFT	**0.008**	0.063	0.545	**0.037**	**0.008**	**0.026**	0.063	0.656	**0.008**	**0.023**
Start	16.09.08	18.04.08	01.10.10	01.11.88	09.10.08	01.11.88	09.11.94	23.10.87	26.10.87	29.12.10
End	01.09.21	06.04.21	20.09.23	08.10.01	27.09.21	08.10.01	25.10.07	21.09.00	22.09.00	15.12.23
NKE	0.054	0.491	0.125	0.234	**0.001**	0.503	0.250	0.929	0.521	**0.003**
Start	06.01.09	27.09.07	24.02.06	03.12.87	19.05.10	16.10.90	06.04.09	19.06.09	11.01.88	13.09.07
End	21.12.21	14.09.20	12.02.19	31.10.00	05.05.23	24.09.03	22.03.22	06.06.22	06.12.00	28.08.20
PG	0.094	**0.036**	0.263	0.395	**0.002**	0.180	0.775	0.142	0.346	**0.000**
Start	02.04.02	23.03.01	09.07.08	27.07.88	22.03.01	13.09.04	20.10.92	22.10.08	12.08.88	03.07.97
End	18.03.15	17.03.14	24.06.21	26.06.01	14.03.14	28.08.17	30.09.05	08.10.21	13.07.01	23.06.10
TRV	**0.000**	**0.001**	0.759	**0.001**	**0.010**	**0.003**	0.249	0.587	0.498	**0.017**
Start	04.06.07	08.05.07	23.06.09	21.04.88	01.06.07	21.04.88	09.05.91	30.01.95	28.09.95	03.12.03
End	19.05.20	23.04.20	08.06.22	21.03.01	18.05.20	21.03.01	19.04.04	15.01.08	15.09.08	17.11.16
UNH	**0.033**	0.234	**0.000**	**0.000**	**0.027**	**0.001**	0.189	0.902	0.921	**0.009**
Start	06.10.09	12.12.08	29.04.87	21.01.87	15.12.09	03.02.87	19.01.87	19.11.08	20.01.87	14.07.10
End	22.09.22	30.11.21	27.03.00	17.12.99	01.12.22	31.12.99	15.12.99	05.11.21	16.12.99	30.06.23
VZ	**0.013**	0.230	0.294	**0.001**	**0.000**	0.076	0.700	0.591	0.457	**0.003**
Start	23.02.95	08.06.94	18.04.91	17.03.92	18.01.91	04.11.92	04.11.92	21.02.90	13.05.91	01.12.93
End	08.02.08	23.05.07	26.03.04	25.02.05	26.12.03	17.10.05	17.10.05	29.01.03	21.04.04	13.11.06
WBA	0.077	0.193	0.269	0.081	0.072	0.336	0.758	0.505	0.387	**0.020**
Start	06.11.92	26.06.95	08.02.91	16.10.90	14.07.05	26.11.93	12.10.94	20.01.87	14.10.88	12.02.97
End	19.10.05	11.06.08	20.01.04	24.09.03	29.06.18	08.11.06	27.09.07	16.12.99	20.09.01	02.02.10
WMT	**0.020**	**0.012**	**0.028**	**0.002**	**0.000**	0.294	0.436	0.470	0.455	**0.002**
Start	09.05.00	14.06.00	23.04.92	05.10.94	09.06.00	06.10.94	10.07.97	10.12.99	05.09.08	20.03.00
End	30.04.13	05.06.13	06.04.05	21.09.07	31.05.13	24.09.07	28.06.10	29.11.12	20.08.21	11.03.13
DIS	0.373	0.711	**0.021**	0.152	**0.031**	**0.039**	0.675	0.617	0.660	**0.041**
Start	27.05.08	10.09.10	27.10.10	09.12.10	09.10.08	08.03.93	12.02.87	09.08.95	09.08.95	03.07.06
End	12.05.21	29.08.23	16.10.23	28.11.23	27.09.21	16.02.06	11.01.00	25.07.08	25.07.08	20.06.19

Table 9: Empirical bootstrapped p-values for daily stock data with $\Delta\tau = 0.25$

	y_{t-1}	s_{t-1}	nrp_{t-1}	ma_{t-1}	dcp_{t-1}	obv_{t-1}	vc_{t-1}	y^2_{t-1}	hlv_{t-1}	sto_{t-1}
MMM	0.073	0.119	0.274	**0.046**	**0.000**	0.566	0.876	0.649	0.137	**0.000**
Start	05.04.88	24.04.14	10.03.93	18.04.94	12.08.87	23.11.12	18.07.12	16.03.06	23.06.92	17.06.87
End	23.06.97	28.07.23	10.06.02	21.07.03	29.10.96	28.02.22	21.10.21	19.06.15	20.09.01	04.09.96
AXP	**0.000**	**0.000**	**0.007**	0.130	**0.001**	0.144	0.634	0.649	0.527	**0.000**
Start	28.03.03	24.06.05	16.06.92	25.05.94	03.06.88	25.05.94	18.06.92	23.07.91	23.06.99	06.04.00
End	28.06.12	29.09.14	07.09.01	26.08.03	21.08.97	26.08.03	17.09.01	11.10.00	29.09.08	16.07.09
AMGN	0.602	0.338	**0.048**	0.112	**0.001**	0.054	0.096	0.177	0.074	**0.002**
Start	07.01.11	06.06.00	13.07.01	29.10.98	25.01.96	21.04.98	22.04.14	30.10.98	23.10.87	28.02.96
End	15.04.20	14.09.09	20.10.10	07.02.08	29.04.05	27.07.07	26.07.23	08.02.08	14.01.97	02.06.05
AAPL	0.209	0.412	0.957	0.605	0.114	0.282	0.396	0.556	0.999	0.090
Start	17.09.99	26.04.99	10.11.05	09.10.96	10.10.00	10.11.88	23.06.05	30.11.10	19.10.87	26.04.99
End	22.12.08	30.07.08	17.02.15	12.01.06	19.01.10	03.02.98	25.09.14	05.03.20	08.01.97	30.07.08
BA	0.067	0.087	0.757	0.086	0.340	0.148	0.144	0.515	0.934	**0.046**
Start	04.11.99	06.12.01	25.04.02	28.01.13	13.01.99	22.08.12	01.07.05	26.10.87	13.12.10	03.01.05
End	12.02.09	11.03.11	27.07.11	29.04.22	22.04.08	26.11.21	06.10.14	14.01.97	19.03.20	08.04.14
CAT	0.290	0.075	0.891	0.183	0.235	0.168	0.285	0.396	0.316	0.622
Start	05.03.87	24.08.87	19.04.94	01.07.93	19.04.94	07.07.93	20.10.06	11.03.99	18.01.13	06.11.96
End	22.05.96	08.11.96	22.07.03	01.10.02	22.07.03	04.10.02	27.01.16	17.06.08	22.04.22	13.02.06
CVX	**0.010**	0.358	**0.026**	0.152	**0.029**	0.453	0.804	0.512	0.684	**0.030**
Start	22.07.02	07.05.12	23.12.94	05.11.92	30.01.95	05.11.92	30.07.91	14.10.87	02.09.14	13.03.95
End	20.10.11	11.08.21	29.03.04	06.02.02	03.05.04	06.02.02	18.10.00	02.01.97	05.12.23	15.06.04
KO	0.691	0.725	0.947	**0.000**	**0.004**	0.375	0.303	0.268	0.700	0.064
Start	16.01.87	26.07.11	31.08.11	21.01.94	19.02.87	04.08.92	10.08.95	01.09.00	24.01.00	11.06.87
End	04.04.96	28.10.20	04.12.20	25.04.03	08.05.96	31.10.01	11.11.04	10.12.09	01.05.09	28.08.96
HD	0.998	0.524	0.594	0.136	0.676	0.954	0.363	0.217	0.968	0.578
Start	17.04.08	22.10.96	04.06.93	05.02.07	11.02.93	20.01.87	24.01.92	25.08.88	04.12.08	01.09.05
End	20.07.17	27.01.06	04.09.02	09.05.16	14.05.02	09.04.96	18.04.01	12.11.97	12.03.18	05.12.14
HON	0.219	0.378	0.956	0.626	0.124	0.308	0.391	0.586	0.999	0.075
Start	17.09.99	26.04.99	10.11.05	09.10.96	10.10.00	10.11.88	23.06.05	30.11.10	19.10.87	26.04.99
End	22.12.08	30.07.08	17.02.15	12.01.06	19.01.10	03.02.98	25.09.14	05.03.20	08.01.97	30.07.08
INTC	**0.009**	0.066	0.629	0.319	0.575	0.141	0.809	0.509	0.391	0.471
Start	21.03.88	11.08.88	24.01.06	16.12.94	10.06.02	30.08.00	25.06.14	19.11.12	12.08.08	09.01.96
End	09.06.97	29.10.97	29.04.15	22.03.04	09.09.11	08.12.09	28.09.23	23.02.22	13.11.17	13.04.05
IBM	**0.050**	0.104	0.499	0.193	0.304	**0.011**	0.756	0.586	0.273	0.290
Start	19.07.93	06.02.13	16.07.12	05.05.94	22.04.93	22.07.96	19.03.99	22.06.95	22.06.95	18.12.92
End	16.10.02	10.05.22	19.10.21	06.08.03	23.07.02	24.10.05	25.06.08	24.09.04	24.09.04	21.03.02
JNJ	0.356	0.083	0.598	0.185	**0.043**	**0.011**	0.536	0.218	**0.025**	0.081
Start	19.07.11	30.04.07	01.05.92	07.10.93	06.08.96	07.10.93	14.09.09	06.03.97	18.12.97	18.07.96
End	21.10.20	01.08.16	25.07.01	08.01.03	08.11.05	08.01.03	17.12.18	12.06.06	29.03.07	20.10.05
JPM	**0.004**	**0.000**	0.120	**0.007**	**0.010**	0.299	0.613	0.388	0.595	**0.045**
Start	16.04.09	07.03.07	14.09.06	04.06.93	17.03.09	03.05.91	22.05.06	25.06.91	08.01.88	06.03.09
End	19.07.18	08.06.16	17.12.15	04.09.02	19.06.18	25.07.00	25.08.15	14.09.00	27.03.97	08.06.18
MCD	0.192	0.132	**0.024**	**0.037**	**0.000**	0.065	0.310	0.589	0.526	**0.003**
Start	11.10.13	14.10.13	30.07.87	06.05.94	16.10.87	12.07.93	01.04.14	03.04.09	20.10.87	12.08.87
End	17.01.23	18.01.23	16.10.96	07.08.03	06.01.97	09.10.02	06.07.23	09.07.18	08.01.97	29.10.96
MRK	0.066	0.392	0.177	0.118	**0.001**	0.177	0.750	0.855	0.983	**0.012**
Start	29.11.11	17.07.13	21.01.14	26.07.94	03.08.11	30.12.91	15.04.97	12.03.09	15.09.92	03.08.11
End	08.03.21	18.10.22	25.04.23	24.10.03	05.11.20	22.03.01	20.07.06	14.06.18	12.12.01	05.11.20
MSFT	**0.020**	0.052	0.912	0.076	**0.026**	0.242	0.354	0.628	0.095	**0.022**
Start	31.01.13	07.12.11	03.12.12	04.11.11	17.05.13	19.04.10	07.03.89	23.10.87	23.10.87	24.05.12
End	04.05.22	16.03.21	08.03.22	11.02.21	19.08.22	23.07.19	27.05.98	13.01.97	13.01.97	30.08.21
NKE	0.099	0.669	0.224	0.226	**0.003**	0.512	0.367	0.497	0.267	**0.007**
Start	25.03.14	06.06.11	11.08.05	27.06.94	30.09.11	27.06.94	25.03.13	18.09.09	18.12.90	25.09.06
End	28.06.23	09.09.20	13.11.14	26.09.03	06.01.21	26.09.03	27.06.22	21.12.18	10.03.00	29.12.15
PG	0.128	0.062	0.581	0.765	**0.003**	0.110	0.938	0.221	0.467	**0.000**
Start	22.03.01	15.03.01	11.07.01	07.08.13	22.03.01	22.06.09	13.12.90	24.10.90	17.12.90	15.03.01
End	30.06.10	23.06.10	18.10.10	08.11.22	30.06.10	24.09.18	07.03.00	14.01.00	09.03.00	23.06.10
TRV	**0.001**	**0.000**	0.887	**0.048**	**0.008**	0.137	0.095	0.584	0.734	**0.004**
Start	29.05.07	17.07.07	04.03.09	30.03.94	06.07.07	16.11.93	12.03.93	08.10.98	11.05.98	06.07.07
End	29.08.16	17.10.16	06.06.18	02.07.03	06.10.16	19.02.03	12.06.02	16.01.08	16.08.07	06.10.16
UNH	0.124	0.148	**0.001**	**0.000**	**0.049**	**0.000**	0.093	0.954	0.892	**0.005**
Start	03.05.12	12.12.08	29.04.87	05.05.87	03.05.12	20.01.87	15.05.89	19.11.08	29.04.87	01.06.12
End	09.08.21	20.03.18	17.07.96	23.07.96	09.08.21	09.04.96	04.08.98	26.02.18	17.07.96	07.09.21
VZ	**0.008**	0.233	0.097	0.214	**0.000**	0.467	0.543	0.890	0.318	**0.000**
Start	20.05.94	03.06.94	04.01.94	01.11.91	13.10.93	17.07.96	15.07.96	27.10.93	17.05.91	03.12.93
End	21.08.03	04.09.03	07.04.03	25.01.01	14.01.03	19.10.05	17.10.05	29.01.03	08.08.00	07.03.03
WBA	0.171	0.267	0.481	**0.047**	0.178	0.084	0.724	0.756	0.299	**0.027**
Start	08.01.02	24.02.97	21.10.93	21.07.94	20.12.01	27.09.93	12.03.90	02.01.87	11.04.88	14.06.00
End	11.04.11	31.05.06	23.01.03	21.10.03	25.03.11	26.12.02	02.06.99	21.03.96	27.06.97	22.09.09
WMT	0.054	0.110	**0.048**	**0.015**	**0.000**	**0.000**	0.264	0.627	0.337	**0.002**
Start	01.11.00	02.07.07	08.03.95	23.01.96	01.11.00	05.10.94	12.12.95	08.08.91	17.10.08	10.03.95
End	10.02.10	30.09.16	10.06.04	28.04.05	10.02.10	08.01.04	18.03.05	30.10.00	22.01.18	15.06.04
DIS	0.791	0.398	0.061	0.163	**0.049**	0.107	0.786	0.417	0.729	**0.038**
Start	07.02.12	01.03.13	17.09.13	03.12.93	08.09.11	13.12.89	01.04.87	19.10.89	01.07.10	08.09.11
End	13.05.21	02.06.22	19.12.22	07.03.03	11.12.20	08.03.99	19.06.96	11.01.99	04.10.19	11.12.20

Looking at the stocks we observe that for 21 out of the 25 stocks considered the number of significant predictors is either the same or differs by only one for the two considered values of $\Delta\tau$. For the remaining five stocks the difference is at most two, except for MSFT where it is three. For $\Delta\tau = 0.35$ there are eleven stocks including AAPL, BA, CAT, KO, HD, HON, INTC, IBM, MRK, NKE and WBA for which at most two rejections occur, nine stocks including MMM, AMGN, CVX, JNJ, JPM, MCD, PG, VZ and DIS for which three to five rejections occur and five stocks including AXP, MSFT, TRV, UNH and WMT for which six to seven rejections occur. Similar results are observed for $\Delta\tau = 0.25$.

When looking at the years corresponding to the periods of highest predictability for the 56 cases that are significant for both values of $\Delta\tau$, we find that in 43 cases the periods for $\Delta\tau = 0.25$ are strict subperiods of the periods when $\Delta\tau = 0.35$ while in 11 cases the periods are partly overlapping and in only two cases there is no overlap at all. This suggests that at least for the most part both window fractions pick the same pocket in time where predictability has occurred. In total this suggests fairly robust findings with regard to a change in $\Delta\tau$. The observed periods of predictability, however, seem to exhibit no clear pattern but are more or less uniformly distributed among the considered sampling period.

5 Concluding remarks

This paper explores the possibility of directional predictability of stock returns from an empirical perspective. To this end, we developed an LM type test of no directional predictability which is robust to empirical features of the data such as predictive regression endogeneity, persistent regressors, and time-varying volatility. To the best of our knowledge, there is no statistical procedure that properly deals with binary dependent variables in this setup. Acknowledging the possibility of episodic predictability, we implement our test for subsamples using a bootstrap procedure. This approach gives insight in the time varying-behavior of the relationships and helps uncovering periods of significant predictive relationships. Monte Carlo simulations confirm the reliability of our testing scheme and highlight its effectiveness when standard inference fails.

On a monthly basis, we investigated returns of the S&P 500 where several predictors, such as the dividend yield, dividend payout ratio, net equity expansion and term spread, show significant predictability only in historical pockets which mostly occur between 1940-1980 while their significance diminishes in more recent periods, particularly after 1990. Only the inflation rate has been a significant predictor from 1980 up until 2010. Using a full sample approach these relationships would not have been uncovered.

The application to daily returns of individual stocks of the DJIA reveals robust patterns of significant rejections of the null of no directional predictability. Key predictors are the stock return itself, a moving average of closing prices, the divergence of price and moving average and the stochastic oscillator yielding the most rejections. The periods of predictability show however no discernible pattern and seem to be evenly spread across the sampling period.

Appendix

Proof of Proposition 1

The result follows from Corollary 3 of Demetrescu et al. (2022); to invoke it we essentially check the conditions of their Lemma 1. To this end, we note that our Assumption 2 is the multivariate counterpart of Assumption 1 of Demetrescu et al. (2022); see also Section S.2.2 of their Supplementary Appendix. Moreover, s_t is a Bernoulli variable with $\mathrm{E}\left(s_t | s_{t-1}, \boldsymbol{x}_{t-1}, s_{t-2}, \boldsymbol{x}_{t-2}, \ldots\right) = g\left(\alpha\right)$ under the null, such that $s_t - g\left(\alpha\right)$ is a martingale difference sequence w.r.t. the filtration $\mathcal{F}_{t-1}$. For the auxiliary regression, we may therefore write as data generating process under the null as

$$s_t = \alpha_s + \boldsymbol{\beta}' \boldsymbol{x}_{t-1} + \zeta_t \qquad \text{with} \quad \boldsymbol{\beta} = \mathbf{0},$$

where $\alpha_s = g\left(\alpha\right)$, $\zeta_t = s_t - g\left(\alpha\right)$, and ζ_t follows a bounded distribution. Since $\mathbf{H}(\cdot)$ is a $K \times K$ matrix of bounded functions on $(-\infty, 1]$, we may conclude that $\left(\zeta_t, \boldsymbol{v}_t'\right)'$ satisfies a multivariate counterpart of Assumption 3 of Demetrescu et al. (2022). Finally, our condition on $\sup_{t \in \mathbb{Z}} \left\| \mathrm{E}\left(\left(\boldsymbol{e}_t \boldsymbol{e}_t' - \mathrm{E}\left(\boldsymbol{e}_t \boldsymbol{e}_t'\right)\right) \otimes \boldsymbol{e}_{t-j} \boldsymbol{e}_{t-k}'\right) \right\|$ is the exact multivariate version of the additional condition stated in Lemma 1 of Demetrescu et al. (2022). The instruments used here being the ones dealt with in Lemma 1 of Demetrescu et al. (2022), we may conclude that their Propositions 1 and 2, and implicitly Corollary 3, apply. The result follows.

Additional simulation results

Table 10: Logit size rates under DGP1: $\sigma^2_{u,t} = \sigma^2_{v,t} = 1$

c		$\varphi = 0$			$\varphi = -0.9$			$\varphi = -0.95$		
	$\Delta\tau$	0.125	0.25	0.35	0.125	0.25	0.35	0.125	0.25	0.35
					$T = 250$					
-5		0.0564	0.0496	0.0460	0.1280	0.1980	0.1824	0.1732	0.2004	0.1744
0		0.0484	0.0508	0.0544	0.2652	0.3512	0.3516	0.2436	0.3996	0.4044
5		0.0528	0.0508	0.0484	0.2392	0.2920	0.2780	0.2224	0.2760	0.2996
20		0.0600	0.0560	0.0552	0.1760	0.1880	0.1608	0.1752	0.1724	0.1588
$0.5T$		0.0580	0.0488	0.0576	0.0860	0.0776	0.0876	0.0680	0.0748	0.0820
					$T = 500$					
-5		0.0504	0.0432	0.0588	0.1764	0.2120	0.1824	0.1932	0.2264	0.1868
0		0.0428	0.0616	0.0472	0.3912	0.4312	0.3744	0.3736	0.4400	0.4632
5		0.0440	0.0544	0.0460	0.3232	0.3284	0.2720	0.3584	0.3268	0.3344
20		0.0500	0.0500	0.0524	0.2368	0.2080	0.1600	0.2500	0.2488	0.1744
$0.5T$		0.0455	0.0512	0.0452	0.0860	0.0772	0.0708	0.0824	0.0780	0.0800
					$T = 2000$					
-5		0.0368	0.0548	0.0552	0.2544	0.2116	0.1928	0.2948	0.2620	0.1824
0		0.0540	0.0608	0.0596	0.5504	0.4888	0.4244	0.5344	0.5024	0.4696
5		0.0628	0.0648	0.0458	0.4352	0.3780	0.3008	0.4740	0.4044	0.3348
20		0.0484	0.0512	0.0504	0.3324	0.2280	0.1844	0.3128	0.2600	0.2152
$0.5T$		0.0512	0.0504	0.0588	0.0848	0.0508	0.0504	0.0660	0.0488	0.0628
					$T = 5000$					
-5		0.0456	0.0480	0.0580	0.2660	0.2264	0.1832	0.3352	0.2600	0.2048
0		0.0532	0.0512	0.0588	0.5536	0.4428	0.4408	0.6144	0.5156	0.4856
5		0.0452	0.0604	0.0492	0.5124	0.3844	0.2980	0.5264	0.4316	0.3400
20		0.0552	0.0512	0.0600	0.3308	0.2408	0.1844	0.3580	0.2556	0.2092
$0.5T$		0.0452	0.0544	0.0556	0.0800	0.0612	0.0532	0.0700	0.0608	0.0548

Note: the nominal size is 5%

Table 11: Logit size rates under DGP2: $\sigma_{u,t}^2 = \sigma_{v,t}^2 = \mathbb{I}(t \leq \lfloor 0.5T \rfloor) + 4\mathbb{I}(t > \lfloor 0.5T \rfloor)$

c		$\varphi = 0$			$\varphi = -0.9$			$\varphi = -0.95$		
	$\Delta\tau$	0.125	0.25	0.35	0.125	0.25	0.35	0.125	0.25	0.35
					$T = 250$					
-5		0.0476	0.0544	0.0496	0.1604	0.2000	0.2004	0.1526	0.2224	0.2288
0		0.0500	0.0488	0.0588	0.2608	0.3320	0.3544	0.2588	0.3860	0.3996
5		0.0560	0.0472	0.0520	0.2212	0.2688	0.2400	0.2488	0.2968	0.2844
20		0.0452	0.0624	0.0532	0.1736	0.1620	0.1456	0.1924	0.2004	0.1732
$0.5T$		0.0576	0.0544	0.0400	0.0780	0.0924	0.0856	0.0788	0.0808	0.0652
					$T = 500$					
-5		0.0480	0.0444	0.0560	0.2332	0.1952	0.2084	0.2636	0.2500	0.2080
0		0.0640	0.0548	0.0496	0.3132	0.3888	0.3368	0.4012	0.4260	0.4232
5		0.0312	0.0600	0.0508	0.3316	0.3280	0.2516	0.3636	0.3212	0.2484
20		0.0568	0.0512	0.0532	0.2188	0.1856	0.1588	0.2352	0.2124	0.1888
$0.5T$		0.0500	0.0504	0.0492	0.0932	0.0630	0.0600	0.1100	0.0604	0.0584
					$T = 2000$					
-5		0.0512	0.0452	0.0532	0.2720	0.2224	0.2224	0.3320	0.2460	0.1888
0		0.0544	0.0604	0.0496	0.5004	0.4456	0.4180	0.5284	0.5000	0.4384
5		0.0480	0.0512	0.0556	0.4528	0.3996	0.2688	0.4844	0.3444	0.3060
20		0.0596	0.0556	0.0604	0.3496	0.2004	0.1828	0.3664	0.2728	0.1740
$0.5T$		0.0604	0.0524	0.0536	0.0888	0.0752	0.0524	0.0588	0.0520	0.0608
					$T = 5000$					
-5		0.0592	0.0524	0.0596	0.2904	0.2288	0.2068	0.3284	0.2304	0.2080
0		0.0496	0.0496	0.0572	0.5480	0.4820	0.5264	0.5844	0.5260	0.4344
5		0.0484	0.0556	0.0540	0.4564	0.3536	0.3052	0.4900	0.3636	0.3068
20		0.0536	0.0520	0.0592	0.3272	0.2288	0.1856	0.3432	0.2472	0.1896
$0.5T$		0.0624	0.0476	0.0504	0.0832	0.0788	0.0608	0.0768	0.0566	0.0604

Note: the nominal size is 5%

Table 12: Logit size rates under DGP3: $\sigma_{i,t}^2 = 0.05 + 0.1i_{t-1}^2 + 0.85\sigma_{i,t-1}^2$ for $i = u, v$

c		$\varphi = 0$			$\varphi = -0.9$			$\varphi = -0.95$		
	$\Delta\tau$	0.125	0.25	0.35	0.125	0.25	0.35	0.125	0.25	0.35
					$T = 250$					
-5		0.0484	0.0524	0.0466	0.1712	0.2044	0.2124	0.1566	0.2352	0.2312
0		0.0512	0.0496	0.0548	0.2688	0.3500	0.3680	0.2672	0.3996	0.4064
5		0.0544	0.0480	0.0552	0.2352	0.2732	0.2508	0.2360	0.3068	0.2940
20		0.0452	0.0624	0.0532	0.1736	0.1620	0.1456	0.1924	0.2004	0.1732
$0.5T$		0.0512	0.0532	0.0472	0.0744	0.0896	0.0888	0.0728	0.0760	0.0672
					$T = 500$					
-5		0.0500	0.0484	0.0524	0.2440	0.1996	0.2000	0.2532	0.2688	0.2152
0		0.0604	0.0552	0.0476	0.3008	0.3792	0.3412	0.4196	0.4292	0.4356
5		0.0488	0.0520	0.0528	0.3420	0.3320	0.2672	0.3512	0.3364	0.2504
20		0.0536	0.0544	0.0600	0.2200	0.1912	0.1720	0.2460	0.2392	0.1904
$0.5T$		0.0508	0.0520	0.0532	0.0948	0.0672	0.0652	0.1108	0.0652	0.0600
					$T = 2000$					
-5		0.0504	0.0488	0.0516	0.2840	0.2204	0.2580	0.3440	0.2564	0.1904
0		0.0520	0.0600	0.0596	0.5216	0.4460	0.4212	0.5396	0.5104	0.4304
5		0.0488	0.0532	0.0536	0.4500	0.4176	0.2728	0.4904	0.3548	0.3000
20		0.0524	0.0560	0.0524	0.3532	0.2104	0.1796	0.3768	0.2708	0.1996
$0.5T$		0.0612	0.0504	0.0556	0.0808	0.0792	0.0584	0.0604	0.0540	0.0612
					$T = 5000$					
-5		0.0576	0.0504	0.0536	0.3004	0.2348	0.2008	0.3308	0.2444	0.2112
0		0.0492	0.0492	0.0560	0.5500	0.4728	0.5124	0.5940	0.5320	0.4444
5		0.0464	0.0532	0.0544	0.4552	0.3676	0.3160	0.4996	0.3728	0.3292
20		0.0516	0.0500	0.0588	0.3332	0.2208	0.2004	0.3808	0.2312	0.1988
$0.5T$		0.0644	0.0496	0.0500	0.0796	0.0780	0.0708	0.0776	0.0524	0.0600

Note: the nominal size is 5%

Additional empirical results

Table 13: Empirical bootstrapped p-values for naïve logit and period with largest test statistic for updated monthly data from Welch and Goyal (2008), 1926:12 to 2023:12.

x_{t-1}	$\Delta\tau = 0.25$		$\Delta\tau = 0.35$		$\Delta\tau = 1$		$\hat{\rho}$	$\hat{\varphi}$
	p-values	Period	p-values	Period	p-values	Period		
dp_{t-1}	**0.018**	1994:11 2019:03	**0.044**	1942:04 1976:04	0.875	1926:12 2023:12	0.994	−0.974
dy_{t-1}	**0.006**	1994:11 2019:03	**0.027**	1942:04 1976:04	0.954	1926:12 2023:12	0.994	−0.067
e/p_{t-1}	0.089	1930:03 1954:07	0.133	1926:12 1960:12	0.877	1926:12 2023:12	0.988	−0.757
de_{t-1}	**0.000**	1953:06 1977:10	**0.003**	1953:08 1987:08	0.674	1926:12 2023:12	0.992	−0.072
$svar_{t-1}$	0.719	1927:01 1951:05	0.403	1930:10 1964:10	0.090	1926:12 2023:12	0.578	−0.095
bm_{t-1}	0.203	1937:02 1961:06	0.151	1937:02 1971:02	0.316	1926:12 2023:12	0.988	−0.810
$ntis_{t-1}$	0.286	1946:02 1970:06	0.141	1939:11 1973:11	0.065	1926:12 2023:12	0.982	−0.032
tbl_{t-1}	**0.000**	1953:06 1977:10	**0.001**	1943:11 1977:11	**0.002**	1926:12 2023:12	0.993	−0.052
lty_{t-1}	**0.001**	1953:06 1977:10	**0.001**	1943:10 1977:10	**0.008**	1926:12 2023:12	0.996	−0.093
ltr_{t-1}	0.405	1943:01 1967:05	0.329	1946:07 1980:07	0.686	1926:12 2023:12	0.052	−0.002
tms_{t-1}	**0.002**	1951:02 1975:06	**0.000**	1941:06 1975:06	0.143	1926:12 2023:12	0.964	−0.013
dfy_{t-1}	0.137	1942:04 1966:08	0.842	1930:10 1964:10	0.213	1926:12 2023:12	0.975	−0.248
dfr_{t-1}	0.386	1984:07 2008:11	0.214	1974:11 2008:11	0.171	1926:12 2023:12	−0.105	−0.006
$infl_{t-1}$	**0.002**	1953:10 1978:02	**0.000**	1941:08 1975:08	**0.000**	1926:12 2023:12	0.483	0.026

Notes: Bold entries denote significance at the 5% level. The error correlation $\hat{\varphi}$ and auto-correlation coefficient $\hat{\rho}$ are based on the full sample.

Table 14: Full sample estimated AR(1) coefficient $\hat{\rho}$ for daily data

	y_{t-1}	s_{t-1}	nrp_{t-1}	ma_{t-1}	dcp_{t-1}	obv_{t-1}	vc_{t-1}	y^2_{t-1}	hlv_{t-1}	sto_{t-1}
MMM	-0.042	-0.031	0.787	1.000	0.631	0.999	-0.339	0.140	0.399	0.609
AXP	-0.052	-0.042	0.785	1.000	0.628	0.999	-0.333	0.221	0.576	0.598
AMGN	0.006	-0.006	0.787	1.000	0.646	0.999	-0.360	0.284	0.044	0.606
AAPL	-0.009	-0.005	0.795	1.001	0.661	1.000	-0.306	0.035	0.355	0.626
BA	0.026	0.001	0.797	1.000	0.682	0.998	-0.309	0.301	0.531	0.612
CAT	0.024	0.026	0.805	1.001	0.676	0.999	-0.334	0.139	0.455	0.640
CVX	-0.073	-0.021	0.787	1.000	0.622	1.000	-0.384	0.174	0.367	0.600
KO	-0.020	-0.014	0.797	1.000	0.634	1.000	-0.374	0.379	0.520	0.605
HD	0.015	-0.003	0.793	1.000	0.662	0.999	-0.321	0.091	0.441	0.627
HON	-0.009	-0.005	0.795	1.001	0.661	1.000	-0.306	0.035	0.355	0.626
INTC	-0.022	0.011	0.790	1.000	0.645	0.998	-0.339	0.181	0.466	0.616
IBM	-0.040	-0.019	0.792	1.000	0.650	0.999	-0.278	0.217	0.354	0.618
JNJ	-0.009	-0.014	0.789	1.000	0.634	0.999	-0.389	0.227	0.389	0.613
JPM	-0.044	-0.008	0.792	1.000	0.636	1.000	-0.321	0.260	0.449	0.624
MCD	-0.032	-0.020	0.787	1.000	0.631	1.000	-0.318	0.192	0.348	0.607
MRK	-0.008	-0.001	0.787	1.000	0.652	0.999	-0.318	0.048	0.414	0.618
MSFT	-0.024	-0.021	0.787	1.001	0.634	0.999	-0.360	0.130	0.419	0.605
NKE	0.014	0.001	0.794	1.000	0.657	0.999	-0.322	0.194	0.488	0.625
PG	-0.055	-0.021	0.784	1.000	0.621	0.999	-0.341	0.161	0.119	0.597
TRV	-0.072	-0.015	0.798	1.000	0.632	1.000	-0.375	0.332	0.590	0.613
UNH	-0.010	-0.003	0.799	1.000	0.660	0.999	-0.311	0.213	0.382	0.616
VZ	-0.040	-0.019	0.790	1.000	0.627	1.000	-0.412	0.290	0.432	0.613
WBA	-0.018	-0.018	0.788	1.000	0.649	1.000	-0.340	0.180	0.425	0.607
WMT	-0.019	-0.030	0.784	1.000	0.633	0.999	-0.339	0.205	0.536	0.599
DIS	-0.025	-0.006	0.786	1.000	0.640	0.999	-0.290	0.131	0.582	0.614

Table 15: Full sample estimated error correlation $\hat{\varphi}$ for daily data

	y_{t-1}	s_{t-1}	nrp_{t-1}	ma_{t-1}	dcp_{t-1}	obv_{t-1}	vc_{t-1}	y^2_{t-1}	hlv_{t-1}	sto_{t-1}
MMM	0.042	0.021	-0.096	0.338	-0.376	-0.023	-0.019	0.028	-0.003	-0.285
AXP	0.052	0.028	-0.094	0.295	-0.376	-0.037	-0.006	0.006	-0.030	-0.278
AMGN	-0.006	0.004	-0.082	0.252	-0.368	0.045	-0.010	0.002	0.001	-0.280
AAPL	0.009	0.003	-0.107	0.134	-0.379	0.023	-0.010	0.009	-0.009	-0.288
BA	-0.026	0.000	-0.077	0.361	-0.392	0.052	0.001	0.029	0.008	-0.283
CAT	-0.024	-0.019	-0.081	0.321	-0.386	0.039	-0.019	0.014	0.020	-0.300
CVX	0.073	0.015	-0.100	0.348	-0.381	-0.037	-0.032	0.016	-0.016	-0.290
KO	0.020	0.009	-0.088	0.365	-0.370	0.001	-0.016	0.030	-0.074	-0.279
HD	-0.015	0.002	-0.073	0.250	-0.377	0.035	-0.024	0.017	-0.007	-0.291
HON	0.009	0.003	-0.107	0.134	-0.379	0.023	-0.010	0.009	-0.009	-0.288
INTC	0.022	-0.008	-0.072	0.321	-0.379	0.037	-0.025	0.022	-0.008	-0.298
IBM	0.040	0.013	-0.091	0.370	-0.389	-0.001	-0.016	0.017	0.036	-0.288
JNJ	0.009	0.010	-0.081	0.309	-0.365	0.019	-0.028	0.022	-0.031	-0.288
JPM	0.044	0.005	-0.086	0.318	-0.380	0.004	-0.014	0.001	-0.010	-0.286
MCD	0.032	0.014	-0.081	0.268	-0.373	0.000	-0.011	0.010	-0.016	-0.288
MRK	0.008	0.001	-0.090	0.380	-0.378	0.004	-0.034	0.009	0.023	-0.290
MSFT	0.024	0.014	-0.086	0.203	-0.370	-0.006	-0.022	0.016	0.036	-0.289
NKE	-0.014	-0.001	-0.077	0.228	-0.376	0.030	0.002	-0.003	-0.011	-0.284
PG	0.055	0.014	-0.076	0.323	-0.372	0.001	-0.036	0.047	0.012	-0.269
TRV	0.072	0.010	-0.084	0.321	-0.389	-0.033	-0.026	0.011	0.024	-0.282
UNH	0.010	0.002	-0.075	0.165	-0.381	0.039	-0.042	0.033	0.040	-0.266
VZ	0.040	0.013	-0.094	0.387	-0.373	0.006	-0.018	-0.003	-0.024	-0.294
WBA	0.018	0.013	-0.086	0.356	-0.383	0.021	-0.020	0.012	0.015	-0.299
WMT	0.019	0.021	-0.097	0.321	-0.370	0.007	-0.016	-0.006	-0.016	-0.290
DIS	0.025	0.004	-0.091	0.300	-0.376	0.015	0.002	0.016	-0.016	-0.286

References

Amihud, Y. and C. M. Hurvich (2004). Predictive regressions: A reduced-bias estimation method. *Journal of Financial and Quantitative Analysis 39*(4), 813–841.

Bajgrowicz, P. and O. Scaillet (2012). Technical trading revisited: False discoveries, persistence tests, and transaction costs. *Journal of Financial Economics 106*(3), 473–491.

Basak, S., S. Kar, S. Saha, L. Khaidem, and S. R. Dey (2019). Predicting the direction of stock market prices using tree-based classifiers. *The North American Journal of Economics and Finance 47*, 552–567.

Becker, J. and C. Leschinski (2018). Directional predictability of daily stock returns. Technical Report 624, University of Hannover, Hannover Economic Papers.

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

Cai, Z., H. Chen, and X. Liao (2023). A new robust inference for predictive quantile regression. *Journal of Econometrics 234*(1), 227–250.

Campbell, J. Y. (1987). Stock returns and the term structure. *Journal of Financial Economics 18*(2), 373–400.

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

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

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

Campisi, G., S. Muzzioli, and B. De Baets (2024). A comparison of machine learning methods for predicting the direction of the US stock market on the basis of volatility indices. *International Journal of Forecasting 40*(3), 869–880.

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

Chen, L., M. Pelger, and J. Zhu (2023). Deep learning in asset pricing. *Management Science 70*(2), 714–750.

Christoffersen, P. and F. Diebold (2006). Financial asset returns, direction-of-change forecasting, and volatility dynamics. *Management Science 52*(8), 1273–1287.

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

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

Demetrescu, M., P. M. M. Rodrigues, and A. M. R. Taylor (2024). Predictive quantile regressions with persistent and heteroskedastic predictors: A powerful 2SLS testing approach. *Available at SSRN 4767275*.

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

Fama, E. F. and K. R. French (1988). Dividend yields and expected stock returns. *Journal of Financial Economics 22*(1), 3–25.

Fama, E. F. and K. R. French (1989). Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics 25*(1), 23–49.

Gonzalo, J. and J.-Y. Pitarakis (2012). Regime-specific predictability in predictive regressions. *Journal of Business & Economic Statistics 30*(2), 229–241.

Gonzalo, J. and J.-Y. Pitarakis (2017). Inferring the predictability induced by a persistent regressor in a predictive threshold model. *Journal of Business & Economic Statistics 35*(2), 202–217.

Goyal, A. and I. Welch (2003). Predicting the equity premium with dividend ratios. *Management Science 49*(5), 639–654.

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

Guerre, E. and H. R. Moon (2002). A note on the nonstationary binary choice logit model. *Economics Letters 76*(2), 267–271.

Gülmez, B. (2023). Stock price prediction with optimized deep lstm network with artificial rabbits optimization algorithm. *Expert Systems with Applications 227*, 120346.

Han, H., O. Linton, T. Oka, and Y. Whang (2016). The cross-quantilogram: Measuring quantile dependence and testing directional predictability between time series. *Journal of Econometrics 193*(1), 251–270.

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

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

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

Lee, J. H. (2016). Predictive quantile regression with persistent covariates: IVX-QR approach. *Journal of Econometrics 192*(1), 105–118.

Leung, M., H. Daouk, and A. Chen (2000). Forecasting stock indices: A comparison of classification and level estimation models. *International Journal of Forecasting 16*(2), 173–190.

Liu, X., W. Long, L. Peng, and B. Yang (2024). A unified inference for predictive quantile regression. *Journal of the American Statistical Association 119*(546), 1526–1540.

Mao, G. (2014). Testing for joint significance in nonstationary binary choice model. *Economics Letters 122*(2), 311–313.

Maynard, A. and K. Shimotsu (2009). Covariance-based orthogonality tests for regressors with unknown persistence. *Econometric Theory 25*(1), 63–116.

Nyberg, H. (2011). Forecasting the direction of the US stock market with dynamic binary probit models. *International Journal of Forecasting 27*(2), 561–578.

Nyberg, H. and H. Pönkä (2016). International sign predictability of stock returns: The role of the United States. *Economic Modelling 58*, 323–338.

Park, J. Y. and P. C. B. Phillips (2000). Nonstationary binary choice. *Econometrica 68*(5), 1249–1280.

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

Pettenuzzo, D. and A. Timmermann (2011). Predictability of stock returns and asset allocation under structural breaks. *Journal of Econometrics 164*(1), 60–78.

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

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

Pönkä, H. (2017). Predicting the direction of US stock markets using industry returns. *Empirical Economics 52*(4), 1451–1480.

Qiu, M. and Y. Song (2016). Predicting the direction of stock market index movement using an optimized artificial neural network model. *PloS ONE 11*(5).

Rapach, D. E., M. C. Ringgenberg, and G. Zhou (2016). Short interest and aggregate stock returns. *Journal of Financial Economics 121*(1), 46–65.

Reschenhofer, E., M. K. Mangat, C. Zwatz, and S. Guzmics (2020). Evaluation of current research on stock return predictability. *Journal of Forecasting 39*(2), 334–351.

Rozeff, M. S. (1984). Dividend yields are equity risk premiums. *Journal of Portfolio Management 11*(1), 68–75.

Stambaugh, R. F. (1986). Bias in regressions with lagged stochastic regressors. *CRSP Working Paper No 156, University of Chicago*.

Stambaugh, R. F. (1999). Predictive regressions. *Journal of Financial Economics 54*(3), 375–421.

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

Yu, D. and D. Huang (2023). Cross-sectional uncertainty and expected stock returns. *Journal of Empirical Finance 72*, 321–340.

Zhu, F., Z. Cai, and L. Peng (2014). Predictive regressions for macroeconomic data. *The Annals of Applied Statistics 8*(1), 577–594.