

ECONOMIC STUDIES
DEPARTMENT OF ECONOMICS
SCHOOL OF BUSINESS, ECONOMICS AND LAW
UNIVERSITY OF GOTHENBURG
240

Predictability in Equity Markets: Estimation and Inference

Tamás Kiss



UNIVERSITY OF
GOTHENBURG

ISBN 978-91-88199-39-3 (printed)
ISBN 978-91-88199-40-9 (pdf)
ISSN 1651-4289 (printed)
ISSN 1651-4297 (online)

Printed in Sweden, by BrandFactory AB
Gothenburg University 2019

Acknowledgements

Would you do it again? This is the question I received the most over the past five years from family and friends who followed my journey leading to this thesis. Admittedly, I was sometimes hesitant about the answer. One thing is unambiguous, though. Having great people who supported, encouraged, helped me during these years was the best part of this experience. This acknowledgement is just a vague attempt to describe how greatly indebted I am to all of you.

First of all, I would like to express my deepest gratitude to my supervisors Erik Hjalmarsson and Ádám Faragó for their support throughout these five years. Erik, I cannot be grateful enough for the countless hours you spent on reading my papers and discussing them with me down to the finest details. Your guidance was invaluable in both shaping research ideas and turning them into actual papers. But, most importantly, I would like to thank you for believing in me even in times of difficulty. Your continuous encouragement helped me tremendously to always reach further. Ádám, apart from the valuable insights and comments on the thesis and my academic work, I am grateful to you for standing by my side during the entire programme. I was very lucky to have both of you as supervisors and it was a great pleasure to get to know you.

I am immensely grateful to Randi Hjalmarsson for her continuous advice and feedback, and especially for all the invaluable help she gave throughout the job market process. A very special thanks to Marcin Zamojski both for his valuable comments during the final seminar and the friendly discussions in the corridor.

I benefited a lot from being part of the Centre for Finance. Many thanks to Martin Holmén for integrating me so well into the finance research environment and for an always welcoming attitude towards me. And thanks for Evert Carlsson, Victor Elliot, Dawei Fang, Alexander Herbertsson, Ted Lindblom, Taylan Mavruk, Conny Overland, Stefan Sjögren, and Jian Hua Zhang for always being open for discussions and giving input to

my research.

I am greatly indebted to all the colleagues at the Department of Economics for the open and collegial environment. Special thanks to Katarina Nordblom and Måns Söderbom for their great work as department heads and for always paying attention to our needs as doctoral students. Teaching was an important and enjoyable part of the work. Thanks for Åsa Löfgren for always listening to my preferences when setting up teaching plans. My appreciation goes to Åsa Adin, Elizabeth Földi, Katarina Forsberg, Selma Oliviera, Ann-Christin Räätäri Nyström, Maria Siirak and Marie Andersson for all the invaluable support with financial and administrative issues.

I was fortunate to share my PhD experience from the very beginning with great colleagues. Anna, thanks for being a great office mate and thanks for the fun discussions we had about Swedish politics and potato chips flavours. Simon, Melissa, Maks, Ida, Debbie, Sebastian, Eyoual, Teddy, Samson, Anh and Youngju, thanks for all the time spent together in and out of the office. I do believe we were a great team together, standing by each others side all the time. Thank you for that! I am also grateful for Andrea, Verena, Simon, Lisa, Laura, Reda and all the other PhD colleagues for their support and the helpful discussions we had.

I had the opportunity to spend two amazing months at the research division of the Swedish Central Bank (Riksbank). I am really grateful for Xin Zhang, for the great cooperation and his support that extended far beyond the walls of the Bank. Thanks for Jesper Lindé, Christoph Bertsch, Daria Finoracchio, Isaiah Hull, Thomas Jansson, Conny Olovsson, Karl Walentin, David Vestin and all the colleagues for a friendly and welcoming environment. And thanks for Melinda Siveg for being a great roommate and “intern-mate” there.

I am also grateful for all the friends in Sweden who made my time outside the office great. Thanks for the barbecues, board game nights, cinema visits, dinners, hikes at

Delsjön and all the great moments we shared. Jutka, I am greatly indebted for always having an open ear for me. Rita, Nándi, Laci and Peti, thanks for making my visit to Stockholm a pleasant time. And thanks for the friends in other parts of the world. Gábor and Jana, my greatest appreciation for all the encouragement you gave me. Mesi, Máté, András and Robi, it was very comforting to know you are always there for me, no matter how far we are from each other. And a very special thanks goes to the youngest friends: Flora, Santiago, Dániel, Bence and Péter. It was a privilege to meet you, play with you, look after you, and see you grow.

Thanks for my family for putting up with my absence, the weird travelling schedules, and for helping with all the practical matters. And especially for my mother and my father for all the moral support and love you gave me.

Zoli, there would certainly be no thesis without you. Words cannot tell how greatly indebted I am for all the support and encouragement. Thank you!

Tamás Kiss

Göteborg, April 2019

Contents

- Introduction** **1**

- I Predictive Regressions in Predictive Systems** **11**
 - 1 Introduction 12
 - 2 Predictive regressions and predictive systems 16
 - 3 Inference under imperfect predictors 19
 - 3.1 Persistence adjusted predictive regression 19
 - 3.2 Comparison to Kalman filter 22
 - 4 Simulations 24
 - 4.1 Simulation setup 24
 - 4.2 In-sample results 25
 - 4.3 Out-of-sample results 26
 - 5 Empirical analysis 27
 - 6 Conclusions 31
 - References 33
 - Figures and Tables 36
 - Appendix 45

- II Testing Return Predictability with the Dividend-Growth Equation: An Anatomy of the Dog** **61**
 - 1 Introduction 62
 - 2 Testing return predictability 64
 - 2.1 Model formulation 64
 - 2.2 Standard OLS-based inference 66
 - 2.3 Cochrane’s simulation approach 67
 - 2.4 Altering the value of ϕ in the simulations 72
 - 2.5 Is the similarity with ML coincidental? 72
 - 3 Size of the test 73
 - 3.1 Lessons from the ML estimator 73
 - 3.2 Monte Carlo simulations 75
 - 3.3 Power 77
 - 4 Empirical results 78
 - 5 Conclusions 79
 - References 80

Figures	83
Appendix	85
III Vanishing Predictability and Non-Stationary Regressors	95
1 Introduction	96
2 The model	98
3 Local demeaning and subsample fixed effects	101
4 Simulations	103
5 Choice of subsample size	105
6 Empirical results	107
6.1 Vanishing predictability in the data	108
6.2 Applying subsample fixed effects	109
6.3 Vanishing or time-varying predictability	111
7 Conclusion	112
References	113
Figures and Tables	116
Appendix	126

Introduction

Predictability in equity markets is a central question in financial economics. Theoretical asset pricing models for time-varying expected returns suggest a relationship between expected returns and variables related to the aggregate risk in the economy such as valuation ratios (e.g. the dividend-price ratio or book-to-market value), term structure variables (e.g. the short rate or the term spread) or macroeconomic quantities (inflation, GDP growth).

Evaluating these relationships empirically is difficult because unexpected returns explain a large part of the return variation. Therefore, tests of return predictability are bound to lack power, which is also reflected by the inconclusiveness of the abundant empirical research. The weak evidence on predictability is exacerbated by a number of statistical difficulties one faces when conducting inference on equity returns. In particular, surveying the recent empirical literature, Kojien and Van Nieuwerburgh (2011) report three “disconcerting statistical features” of return predictability. First, the high persistence of the predictors renders standard testing procedures incorrect. Second, the relationship between returns and potential predictor variables exhibits significant instability over time. Third, the out-of-sample performance of predictive regressions is poor.

The aim of this thesis is to give a deeper understanding of the econometric properties of return predictions. More specifically, I analyse how the three statistical features proposed by Kojien and Van Nieuwerburgh (2011) interact with each other. In particular, how the persistence of the predictor variables affects estimation and inference, feeding into parameter instability and out-of-sample predictive power.

Since several prominent predictors are highly serially correlated, the literature on persistent regressor bias is abundant (Cavanagh et al., 1995; Stambaugh, 1999; Lewellen, 2004; Torous et al., 2004; Campbell and Yogo, 2006; Ang and Bekaert, 2007; Cochrane, 2008). The workhorse model in these papers assumes a linear relationship between the

forecasting variable and expected returns, which therefore inherit the persistence of the predictor. To reconcile this feature with the stylized fact that realized returns are nearly serially uncorrelated, expected returns are assumed to constitute a small fraction of the variation, and the unexpected returns dominate (c.f., Moon and Velasco, 2014). This observation plays a central role in the analysis of return predictability and serves as a common thread throughout the thesis.

In the first chapter of the dissertation, Predictive Regressions in Predictive Systems, I analyse inference on return predictability under the assumption that the predictor variables are imperfect proxies of the expected returns. I show that if there are differences in the dynamic properties of the expected returns and the predictor(s), the predictive regression uses the predictive information inefficiently. This effect is especially strong if the predictors and the expected returns are highly, but not equally, persistent.

As a solution, I propose a persistence adjusted predictive regression. The resulting estimator is a two-stage method, where the expected return process and the predictor process are modelled separately, allowing for the two to have distinct dynamic properties. For instance, the procedure formally allows for highly persistent expected returns to be explained by less persistent term structure variables, a feature not possible in a standard predictive regression formulation. Simulations, as well as empirical results, show that the method leads to both better in-sample fit and real-time forecasting performance.

The second chapter of my dissertation, Testing Return Predictability with the Dividend-Growth Equation: An Anatomy of the Dog, is a joint work with Erik Hjalmarsson. We analyse the dividend-growth based test of return predictability proposed by Cochrane (2008). In his study, Cochrane finds that testing for the absence of dividend growth is a more powerful test of return predictability than a direct test using returns. The key insight is that under the Campbell and Shiller (1988) decomposition either dividend growth or returns must be predictable. Our aim is to better understand the power

gains in the dividend-growth based test of return predictability.

Our main finding is that Cochrane's dividend-growth based test is very similar to a test based on the full information maximum likelihood estimator of the return predictive regression, where the autoregressive (AR) parameter in the dividend-price ratio is treated as known. The power gain is achieved because the dividend-growth based test makes strong use of the postulated value of the autoregressive coefficient. We show that using the same information one could use a maximum likelihood procedure for the return equation that dominates the dividend-growth based test. That is, if one compares testing approaches based on the same information set, there are no power gains from using the dividend-growth regression in testing for return predictability.

The maximum likelihood test is very sensitive to the choice of the autoregressive coefficient, which implies a similar sensitivity in Cochrane's procedure. Moreover, we show that if one uses the OLS estimate of the autoregressive parameter (which is downward biased, e.g. Kendall, 1954), then the dividend-growth based test results in severe size distortion. From an empirical perspective, our findings imply that there are no apparent gains from using the dividend-growth equation when testing for return predictability and that one's prior belief on the persistence of the predictor can substantially affect the outcome of the tests.

In the third chapter of my thesis, Vanishing Predictability and Non-Stationary Regressors, I propose a framework in which predictor persistence and parameter instability are closely connected. I assume that expected returns are stationary and potentially predictable by highly persistent variables. Analogous to the work on noisy predictors (Torous et al., 2004), the information in the predictor is confounded by an uninformative, non-stationary component. This implies that in large samples the persistent but uninformative part becomes dominant. Therefore the predictive power weakens, and eventually vanishes as the number of observations increases. This is consistent with a specific form

of parameter instability, namely that predictors appear to lose power, and the evidence of predictability weakens over time (Ferson et al., 2003; Goyal and Welch, 2008).

I also propose a simple and flexible estimation framework, subsample fixed effects (SFE), that accounts for the presence of a non-stationary non-informative component in the predictor. It builds on the idea that the bias in the ordinary least squares estimation increases with the sample size because the non-stationary component becomes dominant in larger samples. Therefore estimating the parameters on shorter subsamples and pooling them via a fixed effects estimator mitigates the problem. Applying this method to well-known predictors of stock market returns shows an overall increase in the significance of these predictors, supporting the empirical relevance of the proposed model.

References

- Ang, A. and Bekaert, G. (2007). Stock return predictability: Is it there? *Review of Financial Studies*, 20(3):651–707.
- Campbell, J. Y. and Shiller, R. J. (1988). The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies*, 1(3):195–228.
- Campbell, J. Y. and Yogo, M. (2006). Efficient tests of stock return predictability. *Journal of Financial Economics*, 81(1):27–60.
- Cavanagh, C. L., Elliott, G., and Stock, J. H. (1995). Inference in models with nearly integrated regressors. *Econometric Theory*, 11(05):1131–1147.
- Cochrane, J. H. (2008). The dog that did not bark: A defense of return predictability. *Review of Financial Studies*, 21(4):1533–1575.
- Ferson, W. E., Sarkissian, S., and Simin, T. T. (2003). Spurious regressions in financial economics? *Journal of Finance*, 58(4):1393–1414.
- Goyal, A. and Welch, I. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 21(4):1455–1508.
- Kendall, M. G. (1954). Note on bias in the estimation of autocorrelation. *Biometrika*, 41(3-4):403–404.
- Koijen, R. S. and Van Nieuwerburgh, S. (2011). Predictability of returns and cash flows. *Annual Review of Financial Economics*, 3:467–491.
- Lewellen, J. (2004). Predicting returns with financial ratios. *Journal of Financial Economics*, 74(2):209–235.

Moon, S. and Velasco, C. (2014). On the properties of regression tests of stock return predictability using dividend-price ratios. *Journal of Financial Econometrics*, 12(1):151–173.

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

Torous, W., Valkanov, R., and Yan, S. (2004). On predicting stock returns with nearly integrated explanatory variables. *The Journal of Business*, 77(4):937–966.

Chapter I

Predictive Regressions in Predictive Systems*

Tamás Kiss[†]

Abstract

This paper analyses predictive regressions in a predictive system framework, where the predictor is an imperfect proxy for the expected returns. I show that when there are differences between the dynamic structure of the expected returns and the predictor, the predictive regression uses predictive information inefficiently. The effect is especially strong if the predictor and the expected returns are highly, but not equally, persistent. As a solution, I propose a persistence adjustment for the predictive regression. The resulting estimator is a two-stage method, where the expected return and predictor processes are modelled separately, allowing for each to have distinct dynamic properties. Simulations, as well as empirical results, show that the method leads to both better in-sample fit and real-time forecasting performance. The empirical results highlight that the proposed method is especially useful in the case of multiple predictors.

Keywords: Persistence adjustment; Predictive system; Return predictability;

JEL classification: C22, G1.

*I am grateful for the comments by Erik Hjalmarsson, Adam Farago, Hossein Ashgarian, Joakim Westerlund, Emre Aylar, as well as the participants of the 5th Annual PhD Workshop (2018) at the University of Gothenburg, and the joint KWC-CFF workshop (2018) in Varberg.

[†]Department of Economics, Centre for Finance, University of Gothenburg; Email: tamas.kiss@cff.gu.se.

1 Introduction

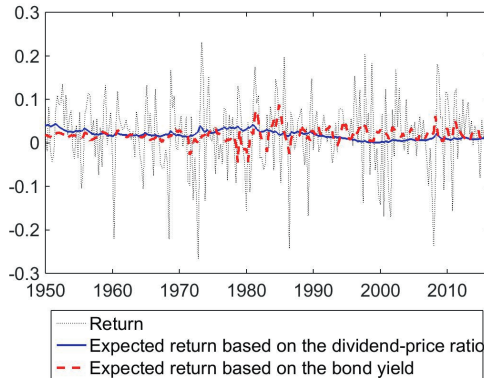
Since the seminal contribution of Campbell and Shiller (1988), several studies have argued for the existence of time-varying expected returns (Lettau and Ludvigson, 2001; Ang and Bekaert, 2007; Cochrane, 2008, 2011). The consensus in the financial literature has subsequently converged toward accepting the existence of return predictability, and the focus has shifted towards understanding how potential predictors contribute to predictability. A significant body of empirical literature has found that the evidence on predictability using predictive regressions is subject to statistical problems (Goyal and Welch, 2008; Kojien and Van Nieuwerburgh, 2011), which has spurred the development of sophisticated inference techniques for testing the null of no return predictability. The proposed tests primarily deal with correcting for the persistent regressor bias to conduct valid tests on whether returns are predictable (Cavanagh et al., 1995; Stambaugh, 1999; Lewellen, 2004; Torous et al., 2004; Campbell and Yogo, 2006; Jansson and Moreira, 2006; Kostakis et al., 2015).

The inferential problem changes, however, when the aim is to assess which variables are useful predictors, rather than explicitly test a null of no predictability. In this case, it is critical to understand how a certain predictor is related to future expected returns, and how this relationship can best be estimated. Indicatively, all predictive regressions cannot *simultaneously* be the true data-generating processes for the expected returns. For instance, univariate regressions with valuation ratios and term structure variables imply expected return processes with different properties. Both types of regressions can still be useful for understanding predictability, as these variables most likely carry information about future expected returns. However, they most probably do so imperfectly in the sense that the predictors only proxy for the expected return series, as described by the predictive system in Pástor and Stambaugh (2009). That is, expected return variation is only partially recovered in any given specification.

In this work, I study predictive regressions in the presence of predictor imperfection. I examine two forms of imperfection that are non mutually exclusive. First, predictors might not explain the full variation in expected returns — that is, the latent expected

return process is not a linear combination of the predictor variables. This form of imperfection reflects a fundamental lack of information in the predictive regression formulation; it cannot be fully controlled for within the model. Second, predictors and the expected returns might have different dynamic properties. I focus on this latter form of imperfection, which can be controlled for within the predictive system. I demonstrate that, based on the standard predictive regression, the explanatory power of the predictor decreases as the difference between the persistence of the expected returns and the predictor grows. This effect is particularly strong if the variables are highly persistent. In the limit, where both the predictor and the expected return are (nearly) non-stationary, the predictive regression becomes spurious (like the problems described in Ferson et al., 2003; Deng, 2013).

Figure 1: **Implied expected return processes from predictive regressions**



Notes: The figure shows the realized excess returns of the Centre for Research in Security Prices (CRSP) value-weighted index (dotted line) and expected returns implied by running univariate predictive regressions $r_{t+1} = \alpha + \beta x_t + e_{t+1}$, where x_t is either the dividend–price ratio (solid line) or the (detrended) yield on the long term government bond (dashed line). The sample runs between 1952 and 2016. Further details on the variables are provided in Section 5.

To intuitively understand why differences in the time-series structure are important, consider the simple example in Figure 1, where expected returns are calculated using univariate predictive regressions based on two different predictors: the dividend–price ratio and the (detrended) long-term bond yield.¹ Unsurprisingly, the two expected return

¹A detailed description of the specifications can be found in the description of Figure 1.

series are markedly different from each other, particularly in terms of their dynamic properties. Figure 1 thus indicates that information on the persistence of the expected returns can be useful when estimating the effect of the predictors.

To pursue this idea, I propose a persistence adjustment to the predictive regression. Incorporating the assumption that expected returns follow a first order autoregressive process, the persistence adjusted predictive regression (PAPR) improves upon the standard ordinary least squares (OLS) estimation in terms of model fit and real-time forecasting performance. The gain in explanatory power comes from the fact that the persistence adjustment disconnects the time-series dynamics of the predictor(s) from the persistence of expected returns. The persistence adjustment is operationalized by a two-step estimation framework. In the first step, the parameters governing the dynamics of the predictors are calculated using the standard least squares technique. In the second step, the latent expected return process is obtained by minimizing the variance of the unexplained returns. The method belongs to the class of extremum estimators described by, for example, Newey and McFadden (1994), and hence its properties are well-known. In particular, the standard errors can be calculated straightforwardly, accounting for the two-step nature of the estimation procedure.

The predictive system is formally represented as a state-space model. In the general case, the expected return process can be estimated by the Kalman filter. Asymptotically, this yields optimal expected return estimates, connecting the variation in expected returns to the predictor and/or to past realized returns. I show that the PAPR is a restricted version of the Kalman filter. It uses information in the predictive variables, but does not connect expected return variation to realized returns. It thus provides the optimal expected return series given the information in the predictor, but ignores information in past returns. The upside is that it requires less parameters to be estimated than the Kalman filter, which translates into less parameter uncertainty, and better out-of-sample forecasts.² In the special case wherein the predictor and the expected returns have the

²The information loss in the PAPR, relative to the Kalman filter, appears smaller. When the true parameters of the model are assumed known (i.e., no parameter uncertainty), the advantage of the Kalman-filtered expected returns is not particularly large.

same time-series dynamics, the PAPR collapses to a standard predictive regression estimated by OLS. Thus, from a practical perspective, the proposed persistence adjustment connects the structural assumptions of the state-space model with the estimation framework of the predictive regression.

The performances of the different specifications of expected returns are compared through simulations and an empirical application. A Monte Carlo experiment reveals that the PAPR outperforms both OLS and the Kalman filter in terms of real-time forecasting performance. This result suggests that the effect of ignoring past return information is dominated by the reduced parameter uncertainty. In line with the theoretical discussion, the advantage of the PAPR over OLS increases as the differences in the dynamics grow.

My empirical analysis is based on quarterly excess stock market returns and the three predictors used in Pástor and Stambaugh (2009): the dividend–price ratio, the consumption-to-wealth ratio (*cay*) by Lettau and Ludvigson (2001), and the detrended yield on the 30-year US government bond. The results confirm that the persistence adjustment involves a bias-variance trade-off compared with the least squares estimation of the predictive regression. Since more parameters are estimated using the same amount of information, the parameter estimates of the PAPR tend to have larger standard errors. Indeed, the time-series dynamics of the expected returns are estimated separately from the predictors, which is an advantage of the persistence adjustment. If the predictor has a relatively low persistence (as in the case of the univariate regression using the bond yield as a predictor), using the PAPR is useful because it can capture the potentially higher persistence of the expected returns. This becomes even clearer in the case of several predictors, where the PAPR outperforms OLS both in-sample and out-of-sample.

The remaining paper is organized as follows. I discuss the model and the properties of the least squares estimation in the predictive system in section 2. Section 3 describes the PAPR and its relationship to other estimation methods. I present the Monte Carlo simulations analysing the properties of the PAPR in section 4 and the empirical application of the method in section 5. I conclude the study in section 6. The appendix contains technical derivations and supplementary results.

2 Predictive regressions and predictive systems

The workhorse model of empirical research on return predictability is the predictive regression. That is,

$$r_{t+1} = \alpha + \beta x_t + e_{t+1}, \quad (1)$$

where r_{t+1} is an observed excess return series (usually stock market index returns in excess of a risk-free rate) and x_t is a predictive variable.³ This specification implies $E_t(r_{t+1}) = \alpha + \beta x_t$, that is, the conditional expected returns are a linear function of the predictive variable. In particular, the predictive regression implies that the dynamics of the expected returns are identical to the dynamics of the predictor; otherwise, the regression is misspecified. The key advantage of this model is that it can be squarely estimated using least squares, and standard testing procedures (potentially corrected for the persistent regressor bias described in Stambaugh, 1999) are readily available. Therefore, it is a simple and well-understood tool to decide whether certain variables predict excess returns. Many predictors have been proposed and tested in the literature, both in univariate settings and in combinations (see, for example, Goyal and Welch, 2008 and the references therein).

Pástor and Stambaugh (2009) introduced the predictive system, where the predictors are *not* perfect proxies of the expected returns. It is a convenient framework to analyze cases wherein the time-series dynamics of expected returns and the predictor differ, since it allows the dynamics of the expected return series to be defined separately. Formally, the following state-space model is used to write the predictive system,

$$r_{t+1} = \mu_t + u_{t+1}, \quad (2)$$

$$\mu_{t+1} = (1 - \gamma_\mu)\bar{\mu} + \gamma_\mu\mu_t + w_{t+1}, \quad (3)$$

$$x_{t+1} = (1 - \gamma_x)\bar{x} + \gamma_x x_t + \epsilon_{t+1}. \quad (4)$$

³In the theoretical discussion, I consider the univariate case only. The results straightforwardly extend to the multiple predictor case, unless it is discussed separately (as in Appendix C).

r_{t+1} and x_t are the same as in case of the predictive regression specification, and $\mu_t = E_t(r_{t+1})$ is the conditional expected return process, modelled separately. The innovation processes $\{u_t, w_t, \epsilon_t\}_{t=0}^{\infty}$ are assumed to be zero mean, serially independent martingale difference sequences with a finite covariance matrix. In this specification, the autoregressive parameters of the expected return (γ_μ) and the predictor process (γ_x) need not coincide. The correlation between the innovations of the expected returns and the predictor, $\rho_{w\epsilon} = \text{Corr}(w_t, \epsilon_t)$, determines the informativeness of the predictive variable. $\rho_{w\epsilon} = 0$ implies that the predictor variable is completely uninformative. In the other extreme, $\rho_{w\epsilon} = 1$, together with $\gamma_x = \gamma_\mu$, implies that expected returns are completely pinned down by the predictor. In this case, the system reduces to the predictive regression in equation (1), with equation (4) describing the evolution of the predictor.

Under the assumption that returns and the predictor are generated by equations (2), (3) and (4), the properties of the predictive regression in equation (1) can be derived. If we assume stationarity in the system ($\gamma_x < 1$, $\gamma_\mu < 1$), the OLS estimator of the slope coefficient in the predictive regression satisfies the standard result,

$$\hat{\beta}_{OLS} \xrightarrow{p} \frac{E((r_{t+1} - \bar{r})(x_t - \bar{x}))}{E((x_t - \bar{x})^2)} = b \frac{1 - \gamma_x^2}{1 - \gamma_\mu \gamma_x}, \quad (5)$$

where $b = \rho_{w\epsilon} \frac{\sigma_w}{\sigma_\epsilon}$ is the coefficient determining the relationship between the expected return and predictor innovations (hereafter the innovation slope coefficient). The formula shows that the slope coefficient of the OLS estimator depends on the relationship between the innovations and the differences in persistence. To analyze this expression further, I fix the amount of predictability, or more specifically the ratio of expected to unexpected return variation. I then define the quantity $\eta = \sigma_w / (\sigma_u \sqrt{1 - \gamma_\mu^2})$, governing the amount of predictability present in returns (the normalized beta, for example, in Wachter and Warusawitharana, 2009, 2015; Lucivjanska, 2018).⁴ Using this notation, the slope coefficient of

⁴Using the quantity η , the amount of explained return variance can be rewritten as

$$R_{true}^2 = \frac{\eta^2}{1 + \eta^2} \quad (6)$$

Note that η is unobservable, since it depends on the parameters of the latent expected return process.

the predictive regression can be decomposed into three parts and a scale factor,

$$\hat{\beta}_{OLS} \xrightarrow{p} \eta \rho_{w\epsilon} \frac{\sqrt{1-\gamma_\mu^2} \sqrt{1-\gamma_x^2} \sigma_u \sqrt{1-\gamma_x^2}}{1-\gamma_x \gamma_\mu} \frac{1}{\sigma_\epsilon} \equiv \beta_{OLS}^{plim}. \quad (7)$$

First, the asymptotic limit of the OLS estimator depends positively on the relative variation of the expected returns, η . The intuition is straightforward: the larger the amount of predictability, the stronger the regression evidence becomes. Second, β_{OLS}^{plim} depends on the correlation between the predictor and the expected returns ($\rho_{w\epsilon}$), as a better proxy for the predictor implies a larger slope coefficient in the predictive regression. The third component highlights the importance of distinguishing between the time-series properties of the expected return and the predictor series. The value of this term, which depends only on the persistence parameters, is between zero and one, and it is equal to one only if $\gamma_\mu = \gamma_x$. The strongest predictive relationship can thus be detected if the persistence of the predictor and the expected returns are aligned.⁵ The first two components (the amount of expected return variation η and the correlation $\rho_{w\epsilon}$) are “fundamental” quantities of the model; they directly determine the amount of variation a predictor can explain. Without any further information, these quantities must be viewed as given and fixed. In contrast, the difference in persistence is a feature that can be corrected for by using the structural assumptions of the model, as discussed further in section 3.

If $\gamma_x \rightarrow 1$ while $\gamma_\mu < 1$, the OLS estimator converges to zero, keeping other parameters — especially the scaling and the degree of predictability — constant. This reflects the fact that a non-stationary variable cannot be used to capture stationary variation. The same result holds if $\gamma_\mu \rightarrow 1$ and $\gamma_x < 1$, since, analogously, a stationary variable cannot capture the variation in a non-stationary variable. Furthermore, if both persistence parameters approach one, the limit is not well defined. In particular, the limit

$$\lim_{(\gamma_x, \gamma_\mu) \rightarrow (1,1)} \frac{\sqrt{1-\gamma_\mu^2} \sqrt{1-\gamma_x^2}}{1-\gamma_x \gamma_\mu}$$

⁵Similar results for the regression t-statistics and the R^2 are derived in Appendix A.

depends on the relative rates of convergence for γ_x and γ_μ . The special case, when both the expected returns and the predictor approach the non-stationary region must be analyzed separately. This case is of interest because of the extensive literature on the effect of persistent regressor bias in predictive regressions (Stambaugh, 1999; Lewellen, 2004; Campbell and Yogo, 2006; Phillips, 2014, among others), and the empirical fact that many of the important predictors (particularly, valuation ratios) exhibit high persistence. The full formal analysis is relegated to Appendix B, but the main finding is that the correlation between expected return and predictor innovations plays a crucial role when predictors are nearly non-stationary. If the correlation is not strong, the regression t-statistic is dominated by the spurious regression effect, making inference invalid. In fact, the spurious predictive regression literature (Ferson et al., 2003; Deng, 2013), where the predictor is completely uninformative about expected returns, is a special case of the results derived in Appendix B. On the other hand, when the correlation between the innovations is high, the difference in persistence does not enter the asymptotic distribution of the test statistics. In this case, the predictor and the expected returns become asymptotically equivalent. In a knife-edge case however, the difference in persistence does play a role, affecting the distribution of the t-statistic through an extra term that enters due to imperfection.

Overall, the predictive system in the (near) non-stationary case becomes tenuous, where meaningful inference is only possible in highly specific cases. That is, unless the data-generating process is in the knife-edge case described in Proposition 1 in Appendix B, the predictive system either results in spurious predictability or asymptotically reduces to the predictive regression. Therefore, in the remaining analysis I focus on the stationary case, in which the predictive system does not collapse to either of these special cases.

3 Inference under imperfect predictors

3.1 Persistence adjusted predictive regression

As discussed in the previous section, the predictive regression is misspecified if the predictor and the expected returns have different persistence. In this section, I propose

a persistence adjusted predictive regression (PAPR), which is a method that explicitly corrects the predictive regression to account for the difference in the persistence of the predictor and the expected returns.

Assume that the data-generating process is described by the predictive system in equations (2), (3), and (4). Given the parameters of the model, the innovations of the predictor, $\{\epsilon_t\}_{t=1}^T$, can be calculated by applying the dynamics in equation (4). The expected return process is formed from the innovations w_t , and although these are unobservable, a projected expected return series can be calculated using the predictor innovations and the parameters of the model. The least squares projection of the expected return innovation is given by $w_t|\epsilon_t = b\epsilon_t$, where b is the innovation slope coefficient introduced in equation (5). The projected expected return series can then be calculated as

$$\mu_t = \bar{\mu} + \sum_{s=1}^t \gamma_\mu^{t-s} b \epsilon_s. \quad (8)$$

That is, the projected innovations $b\epsilon_s$ are used in the autoregressive filter governing the dynamics of the expected return process. If $\gamma_\mu = \gamma_x$, the expected return series implied by equation (8) reduces to

$$\mu_t = \bar{\mu} + \sum_{s=1}^t \gamma_x^{t-s} b \epsilon_s = \bar{\mu} + b \sum_{s=1}^t \gamma_x^{t-s} \epsilon_s = \alpha + \beta x_t,$$

in which case the expected return process implied by the projection is identical to that of the predictive regression. Estimating (8) with γ_μ as a free parameter is an augmented version of OLS estimation, where the potentially different persistence of the predictor and the expected returns is considered.

The parameter estimation of the PAPR can be performed by minimizing the forecast error. The objective function can be written as

$$Q(\theta) = \frac{1}{T-1} \sum_{t=1}^{T-1} (r_{t+1} - \tilde{\mu}_t(\theta))^2, \quad (9)$$

where θ denotes all parameters of the model. The structure of the problem suggests

that the estimation can be accomplished in two steps. Let $\theta = (\theta_1, \theta_2)$, where $\theta_1 = \{\bar{x}, \gamma_x\}$ includes the parameters of the predictor process and $\theta_2 = \{\bar{\mu}, \gamma_\mu, b\}$ contains the parameters of the expected return process and the innovation slope coefficient. Since the predictor follows a simple autoregression, OLS can efficiently estimate its parameters, and its innovations can be calculated in the first step. In the second step, the objective function in (9) can be minimized with respect to θ_2 to obtain the estimates of the parameters of the expected return process and the innovation slope coefficient,

$$\begin{aligned} \hat{\theta}_2 &= \arg \min_{\theta_2} Q(\hat{\theta}_1, \theta_2) \\ &= \arg \min_{\theta_2} \frac{1}{T-1} \sum_{t=1}^{T-1} \left(r_{t+1} - \tilde{\mu}_t(\hat{\theta}_1, \theta_2) \right)^2 \\ &= \arg \min_{\theta_2} \frac{1}{T-1} \sum_{t=1}^{T-1} \left(r_{t+1} - \bar{\mu} - \sum_{s=1}^t \gamma_\mu^{t-s} b \hat{e}_s \right)^2 \\ &= \arg \min_{\theta_2} \frac{1}{T-1} \sum_{t=1}^{T-1} \left(r_{t+1} - \bar{\mu} - \sum_{s=1}^t \gamma_\mu^{t-s} b [x_s - (1 - \hat{\gamma}_x) \hat{x} - \hat{\gamma}_x x_{s-1}] \right)^2, \end{aligned}$$

where \hat{e}_t is the fitted residual of the predictor and $\tilde{\mu}_0 = \bar{r}$. That is, the expected return process is initialized in the long term average of realized returns, captured by the sample mean.⁶

Two-step estimators constitute a special case of extremum estimators; hence, their asymptotic properties are well known (Newey and McFadden, 1994). Since the parameters of the predictor process can be consistently estimated by OLS, the second step is also consistent. The asymptotic distribution of the estimator has the usual form, except that standard errors of the estimates in the second step must consider the estimation error of the first step (see Appendix C).

Analogous to classical regressions, the PAPR can also be easily extended to the multivariate regression case. If the variables $x_t^1, x_t^2, \dots, x_t^J$ are all potential predictors of the expected returns, the first-step innovation series $\hat{e}_t^1, \hat{e}_t^2, \dots, \hat{e}_t^J$ are obtained using a multivariate time-series model for the predictors. The expected return projection is then

⁶This initialization is not completely innocuous, since theoretically, the exact specification of the initialization can impact the estimation. This issue is explored further in Appendix D. Consequently, the bias of the current initialization is empirically negligible.

formulated as $w_t|\epsilon_t^1, \epsilon_t^2, \dots, \epsilon_t^J = \sum_{j=1}^J b_j \epsilon_t^j$. The second step is modified whereby all the innovation slope coefficients $\{b_j\}_{j=1}^J$ are jointly estimated with the parameters of the expected return process.

Another assumption of the PAPR that can be easily relaxed is the time-series dynamics of the predictor. Equation (4) can be redefined using a more general time-series model, and the predictor innovations are obtained by estimation of the defined model. Asymptotic results and further discussion on the implementation of the two-step procedure are found in Appendix C.

3.2 Comparison to Kalman filter

Pástor and Stambaugh (2009) used a Kalman filter to estimate the predictive system. Since their data-generating process is identical to the one proposed in the present study, I compare the persistence adjustment to the Kalman filter estimation of the system. The key difference between the two methods is that the PAPR contains information only from the predictor (its covariance structure and cross-correlation with the returns), while the Kalman filter connects the expected return variation not only to past predictor innovations but also to past returns. To see this, consider the regression formulation of the conditional expected returns in the state-space model (the derivation can be found in Pástor and Stambaugh, 2009),

$$\mu_t = \bar{\mu} + \sum_{s=1}^t \omega_s (r_s - \bar{\mu}) + \sum_{s=1}^t \delta_s \epsilon_s. \quad (10)$$

The parameters of the linear model, $\omega_s = m(\gamma_\mu - m)^{t-s}$ and $\delta_s = n(\gamma_\mu - m)^{t-s}$, depend on the persistence of the expected returns and the parameters m and n , which, in turn, are functions of the parameters in equation (2)–(4) and the covariance matrix of the error terms. The parameters m and n measure the degree to which (past) returns and predictors contribute to the expected return variation, respectively.⁷ These parameters need to be estimated.

⁷Their exact dependence on the parameters of the underlying data-generating process is given by equations (A36) and (A37) in Pástor and Stambaugh (2009).

The Kalman filter estimation can be viewed as estimating the parameters of equation (10) without imposing any further assumption on the parameters. In contrast, using the PAPR is equivalent to imposing $m = 0$. In this case, $\omega_s = 0$, $b = n$, and $\delta_s = b\gamma_\mu^{t-s}$.⁸ Thus, equation (10) collapses to the specification of the PAPR in equation (8). Setting $m = 0$ is an assumption, thus forcing past returns to have no effect on the expected return prediction. While the Kalman filter attributes time variation in expected returns to both past returns and predictor innovations, the proposed two-step method shuts down the channel through which past returns directly operate.

If the model is correctly specified, the Kalman filter estimated by maximum likelihood results in an asymptotically optimal estimate of the expected returns. However, the typical sample size in the current return predictability setting is relatively small compared with the number of parameters that need to be estimated in a full state-space model. Therefore, the parameter uncertainty is potentially large in the Kalman filter estimation, and the asymptotic optimality results might not be relevant in empirically occurring sample sizes. The PAPR is advantageous because it reduces the parameter uncertainty compared with the Kalman filter. That is, the (asymptotic) bias caused by imposing the restriction $m = 0$ is traded-off against the reduced number of parameters. The PAPR can thus more robustly estimate expected returns, while still considering the potential difference between the persistence of the predictor and the expected returns.

In the following two sections, I analyze the PAPR and further compare it with the Kalman filter and OLS both in Monte Carlo simulations and in an empirical application. I compare three different specifications of expected returns: $\alpha + \beta x_t$ for the standard predictive regression, equation (8) for the PAPR, and equation (10) for the Kalman filter. I focus on their performance both in-sample (how well they describe expected returns) and out-of-sample (how they perform in terms of real-time forecasting).

⁸ $n = (\sigma_{w\epsilon} - m\sigma_{u\epsilon})\sigma_\epsilon^{-2}$ in the general formulation in Pástor and Stambaugh (2009). $m = 0$ corresponds to $n = \sigma_{w\epsilon}\sigma_\epsilon^{-2} = b$.

4 Simulations

In this section I perform a Monte Carlo simulation to assess the properties of the PAPR in a predictive system. I present two sets of simulation results that closely relate to the theoretical discussion in sections 2 and 3. First, I show that the predictive regression cannot capture the persistence of the expected returns when it is separate from the predictor. Therefore, the PAPR can produce a better estimate of the expected returns (a higher in-sample fit), since it estimates the persistence parameter of the expected returns separately. Second, I carry out an analysis of real-time forecasting performance by comparing the predictive regression, the PAPR, and the Kalman filter estimation of the full system.

4.1 Simulation setup

All simulations assume that the data-generating process is given by the predictive system described in equations (2)–(4), where the innovations follow a jointly normal process. The baseline parametrization of the system is as follows. The values $\bar{\mu} = 0.018$ and $\bar{x} = 0.03$ are the unconditional means of the return and the predictor, respectively. These values correspond to the quarterly unconditional mean of the excess return and the dividend–price ratio. The expected returns are assumed to explain 5 percent of total return variation ($\eta^2 = 0.05$). The default value for the persistence of the expected returns is $\gamma_\mu = 0.9$ and the autoregressive parameter of the predictor $\gamma_x \in [0.5, 0.99]$ is specified for each simulation.

The standard deviations of the unexpected and expected returns are set such that the quarterly unconditional volatility is 8 percent. Given the parameters (particularly, η and γ_μ) above, this implies $\sigma_u = 0.081$ and $\sigma_w = 0.011$. Further, using the value of γ_x , the standard deviation of the predictor, σ_ϵ , is calculated to ensure that $\beta_{OLS}^{plim} = \rho_{we}$. This choice makes the comparison over specifications easier, since it imposes the same asymptotic limit for the OLS estimator in each specification.

The correlation structure of the innovations is chosen to reflect the presence of imperfection. The default value for predictor imperfection — that is, the correlation between

the expected return and predictor innovation — is set to $\rho_{w\epsilon} = 0.9$. The correlation between expected and unexpected returns is set to $\rho_{uw} = -0.7$ to capture the negative correlation for the dividend–price ratio that has been documented in the literature. I also assume that $\rho_{u\epsilon} = \rho_{uw}\rho_{w\epsilon}$, which implies that the unexpected returns and the predictor are only correlated through their correlation with the expected return innovations. All simulations are performed with $T = 200$, which corresponds to a typical sample size in the context of return predictability using quarterly data. The results are based on 1000 repetitions in each case. These parameter choices are retained throughout the simulations, unless otherwise noted.

4.2 In-sample results

The first set of the simulation results highlights the misspecification in the predictive regression arising from its inability to capture the potential difference in persistence between the expected return process and the predictor. The results are obtained by fixing all parameters at default values (particularly, $\gamma_\mu = 0.9$) and varying the autoregressive coefficient of the predictor between $\gamma_x = 0.5$ and $\gamma_x = 0.99$. Table 1 displays the summary statistics of the estimation results for the predictive regression and for the PAPR. Since the expected return implied by the standard predictive regression is $\hat{\mu}_{t,OLS} = \hat{\beta}x_t$, its persistence is pinned down by x_t . Therefore, the estimated persistence of the expected returns is biased, unless $\gamma_\mu = \gamma_x$ (as seen in Table 1, Panel a). The persistence adjustment is advantageous because it can estimate γ_μ with less bias and the persistence of the estimated expected returns no longer depends on γ_x (Table 1, Panel b).⁹

[Table 1 about here.]

The ability of the PAPR to capture the difference in persistence translates into better model fit. To illustrate this, I calculate the in-sample R^2 of the models. It measures the

⁹All the estimates of the autoregressive parameters are downward biased due to the small sample bias present in OLS estimation. Nevertheless, this does not influence the comparison between the standard predictive regression estimated by OLS and the PAPR.

degree to which a given model explains return variation. That is,

$$IS - R^2 = 1 - \frac{Var(r_{t+1} - \hat{\mu}_t)}{Var(r_{t+1})}, \quad (11)$$

where $\hat{\mu}_t$ is the expected return process generated by the model. In case of OLS, the measure is identical to the usual R^2 , while in a non-linear specification it is usually called the pseudo- R^2 . According to Figure 2, the PAPR is better than the standard predictive regression in terms of in-sample R^2 , and its advantage increases as the difference in persistence grows, confirming the theoretical results in section 2.

[Figure 2 about here.]

4.3 Out-of-sample results

The second set of simulations analyzes the PAPR in terms of its ability to predict expected returns in a real-time forecasting setup, and compares it to that of the predictive regression and the estimation of the full state-space system by the Kalman filter. The real-time forecasting performance of the model is measured by its out-of-sample R^2 defined by Goyal and Welch (2008),

$$OOS - R^2 = \frac{MSFE_{benchmark} - MSFE_{model}}{MSFE_{benchmark}}, \quad (12)$$

where $MSFE_{model}$ ($MSFE_{benchmark}$) is the mean squared forecasting error of the model (benchmark). The historical mean forecast (i.e., $\mu_t = \frac{1}{t} \sum_{s=1}^t r_s$) is used as the benchmark model. A positive $OOS - R^2$ implies that the model outperforms the constant expected return model. The training sample is always set equal to 200 observations and the simulations are based on 1000 one period ahead forecasts of the expected returns. Out-of-sample R^2 values are calculated for the default parametrization, and γ_x varies between 0.5 and 0.99.

[Figure 3 about here.]

Figure 3 illustrates the results of the simulations. Indeed, the PAPR typically outperforms both the standard OLS estimation and the Kalman filter. Since the PAPR involves more parameters, its advantage over the standard predictive regression is smaller if there is no large difference in persistence. However, with even a relatively small difference in persistence, the PAPR outperforms the standard predictive regression. The results in Figure 3 further indicate that the maximum likelihood estimation of the Kalman filter is not suitable for out-of-sample forecasts due to parameter uncertainty. It always underperforms compared with the other methods, but also the historical mean specification. This confirms the results in Lucivjanska (2018), that is, the predictive regression is usually better in terms of out-of-sample performance.

[Figure 4 about here.]

To demonstrate how the weak performance of the Kalman filter can be attributed to estimation uncertainty, Figure 4 presents the results for when the parameters of the expected return process are known. That is, there is no estimation error, and the differences in the models are entirely due to how expected returns are calculated ($\alpha + \beta x_t$ for the predictive regression, equation (8) for the PAPR, and the filtering equations described by Pástor and Stambaugh (2009) for the Kalman filter). In this empirically infeasible case, the Kalman filter provides optimal expected return series. This is reflected in Figure 4, with the Kalman filter generating the highest out-of-sample R^2 . However, the figure also highlights the importance of adjusting for the difference in persistence. The prediction made by the standard predictive regression is dominated by the PAPR, which, in turn, is remarkably close to the Kalman filter. The results in Figure 4 thus suggest that the advantage of the full system estimation is limited, given the similarity between PAPR and the Kalman filter.

5 Empirical analysis

I now turn to an empirical analysis using the PAPR method described above. I estimate various models to predict the quarterly returns on the Center for Research in Security

Prices (CRSP) value-weighted stock market index between 1952 and 2016. Excess returns are calculated using the 30-day Treasury bill as the risk-free rate. The predictors are the same as those in Pástor and Stambaugh (2009). The dividend–price ratio (dp) is calculated using returns on the CRSP value-weighted index with and without dividends. The consumption-to-wealth ratio (cay) is obtained from Lettau and Ludvigson (2001). The bond yield (by) variable is the difference between the 30-year government bond yield and its twelve-month moving average in the CRSP Treasuries file.

Descriptive statistics for the variables are shown in Table 2. The first-order autocorrelations in the third column show that the predictors are substantially different in terms of their time-series properties. The autoregressive parameter for the dividend–price ratio is 0.97, which implies high persistence, close to non-stationarity. On the other hand, the first-order autocorrelation of the bond yield is only 0.61, implying a relatively fast mean reversion. The cay variable is in the middle with an autoregressive parameter of 0.82. These numbers also suggest that the expected return processes implied by the univariate regressions are likely different.

[Table 2 about here.]

Panel (a) in Table 3 presents the results from univariate OLS regressions and a multivariate regression including all the variables. In this dataset, the dividend–price ratio is the weakest predictor of the expected returns, while the other two variables exhibit stronger relationships with one-quarter-ahead returns. Including all these variables in the regression leaves the coefficient on each variable largely unchanged, which suggests that the three predictor variables convey different information, and multicollinearity is not particularly large.

[Table 3 about here.]

[Table 4 about here.]

The PAPR results show similar patterns. Innovations in the first step are obtained

through a first-order autoregressive filter (shown in Table 4).¹⁰ The estimates of the second step, that is, the estimated persistence of the expected returns and the innovation slope coefficients, are shown in panel (b) of Table 3. Overall, the univariate results of the PAPR reflect the results of the standard regression estimated by OLS. The autocorrelation coefficient of the expected returns (γ_μ) is estimated with high precision. The innovation slope coefficients are estimated with larger standard errors than the corresponding OLS slope coefficients. Thus, with the PAPR, there are more parameters to estimate compared with the standard predictive regression. When all three variables are included in the model, all are significant, suggesting that they all help explain expected return innovations. The PAPR estimates of the persistence of expected returns (γ_μ) reveal a more uniform pattern over the specifications than the corresponding OLS estimates (bottom row of each panel).

Since the simulated results show a downward bias in the PAPR estimate of γ_μ , I also report the bootstrap bias-corrected estimates for this parameter in the last row of Table 3, Panel (b).¹¹ The bias-corrected estimates are larger than the baseline values, though only to a small extent. Thus, even though the downward bias is present empirically, it is not substantial. Therefore, the forecasting results in the next section are based on the baseline PAPR estimates of γ_μ .

The autoregressive coefficient of the expected returns is a key parameter of the model, and obtaining results conditional on γ_μ is also informative, given the additional parameter uncertainty of PAPR compared with the standard predictive regression. Fixing the autoregressive parameter of the expected return process decreases the number of estimated parameters, thus reducing the parameter uncertainty in the PAPR. It also eliminates the minor downward bias in the PAPR estimate of γ_μ . Table 5 shows the restricted estimation results. The first column replicates the unrestricted estimates, while the second and third columns present the restricted estimation results, imposing either $\gamma_\mu = 0.8$ or $\gamma_\mu = 0.95$.

¹⁰In unreported results, I considered alternative specifications. I fitted higher-order autoregressive models for each predictor, where the order is determined by the Akaike and Bayesian information criteria, and ARMA(1,1) models. The results based on the alternative time-series specifications remain qualitatively similar.

¹¹A residual bootstrap with 200 repetitions is performed.

The different outcomes of the estimation show variation, but the differences are not large. That is, the estimated innovation slope coefficients are not particularly sensitive to the restrictions. I return to the usefulness of imposing restrictions on γ_μ in the out-of-sample results discussed below.

[Table 5 about here.]

To evaluate the PAPR, I also calculate its in-sample fit and real-time (out-of-sample) forecasting performance. The measures I use are the in-sample and out-of-sample R^2 defined in section 4, and the Diebold–Mariano test assessing equal forecasting performance (Diebold and Mariano, 1995). Table 6 presents both in-sample and out-of-sample results for the one regressor specifications and the full model, where all the regressors are included. The results for the standard linear model estimated by OLS as well as the unrestricted PAPR and two of the restricted forms are shown. The full estimation of the system using maximum likelihood Kalman filter is also presented.¹²

[Table 6 about here.]

Table 6 shows that the unrestricted PAPR outperforms the OLS estimation in each case in terms of in-sample R^2 . This suggests that the predictive regression is misspecified; thus, adjusting for persistence differences mitigates the misspecification. The in-sample gains of the PAPR range between 0.2 and 2 percentage points, the latter implying an 18 percent improvement on the standard predictive regression in terms of in-sample R^2 . When restrictions are imposed on the PAPR, the in-sample results worsen to some extent compared with the unrestricted model.

As seen in the estimation results in Table 3, the standard errors of the PAPR estimates are relatively large, which might negatively affect the out-of-sample performance. This is at least partially supported by the out-of-sample results shown in Table 6. Imposing a pre-defined value on the persistence parameter γ_μ , as discussed above, can potentially reduce the overall parameter uncertainty and improve the out-of-sample forecasts. In fact,

¹²Note that these results are not directly comparable with the results in Pástor and Stambaugh (2009), since performing the full Bayesian estimation as in the original study is outside the scope of the current analysis.

imposing a relatively high persistence ($\gamma_\mu = 0.95$) makes sense both from an economic and econometric perspective because most evidence suggests that the time variation in expected returns is persistent. As seen in Table 6, fixing $\gamma_\mu = 0.95$, the PAPR forecasts perform the best out of sample in all cases except for the dividend–price ratio, which appears to be a weak predictor with no out-of-sample gains for any estimation method. In the multivariate specification, all PAPR forecasts (whether based on restricted or unrestricted estimates) outperform the OLS one. This reflects the fact that the expected return parameters are estimated with more precision in the multivariate case (see also in Panel (b) in Table 3).

The fit of the Kalman filter tends to be much weaker than that of the other two methods in the specifications using the dividend–price ratio (Panel a and d in Table 6). This is likely because the estimation of the Kalman filter parameters becomes unstable when the persistence of the state variables is high. Further, out-of-sample performance tends to be weak in all specifications, which echoes the results of Lucivjanska (2018) and the simulation results in section 4.

6 Conclusion

In this study, I investigated predictive regressions when the data are generated by the predictive system proposed by Pástor and Stambaugh (2009), where predictors are imperfect proxies of the expected returns. I demonstrated how predictor imperfection can be decomposed into two main terms: the imperfect correlation between the innovation of the predictor and the expected returns as well as the difference in persistence between the predictor and the expected returns. While the first type of imperfection is arguably fundamental, the second type can be controlled for within the model. To this end, I proposed a persistence adjustment to the standard predictive regression, which is based on the structural assumptions of the predictive system.

The proposed estimator was labeled PAPR. It is a two-stage method, where the expected returns and predictor processes are modelled separately, allowing for each to have distinct dynamic properties. This method involves minimal deviation from the standard

predictive regression. If the persistence parameters of the predictor and the expected returns are equal, the method is asymptotically identical to the standard predictive regression. My simulations reveal that the model fit of the predictive regression can be substantially lower if the difference in persistence is not taken into account and the persistence adjustment can significantly improve upon standard least squares results in predictive regressions. This is particularly true if the difference in persistence is large. Empirically, both in-sample and out-of-sample improvements, relative to OLS estimation, are documented in relevant cases.

The focus of the current study was to evaluate how assumptions about imperfect predictors affect predictive regression evidence on return predictability. If the data-generating process is given by the predictive system, the Kalman filter delivers asymptotically optimal expected return series. Disregarding estimation uncertainty, the PAPR is thus inferior to the Kalman filter. However, a simple persistence adjustment brings the predictive regression results remarkably close to the estimation of the full system, and in practical situations the parameter uncertainty in the Kalman filter results in poor in-sample and out-of-sample performance. The proposed method therefore provides a simple and almost efficient way of dealing with predictor imperfection.

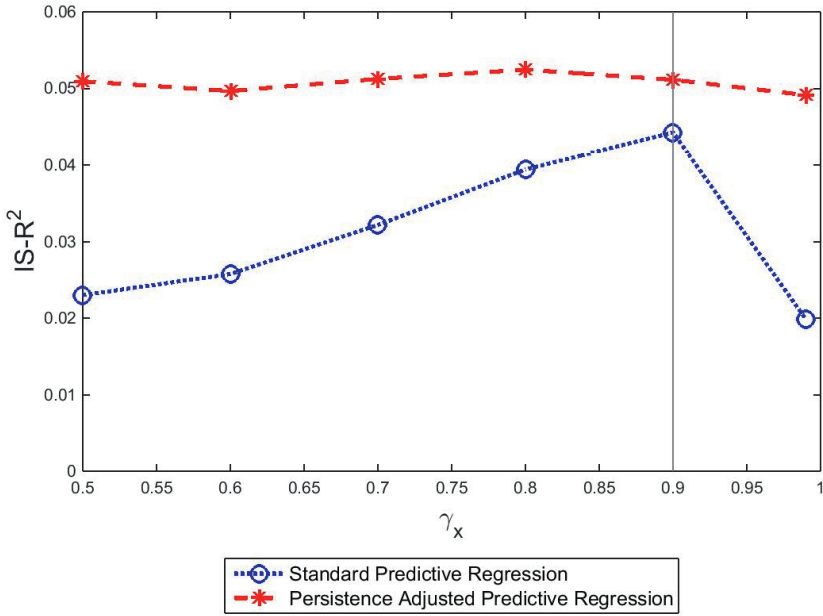
References

- Ang, A. and Bekaert, G. (2007). Stock return predictability: Is it there? *Review of Financial Studies*, 20(3):651–707.
- Campbell, J. Y. and Shiller, R. J. (1988). The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies*, 1(3):195–228.
- Campbell, J. Y. and Yogo, M. (2006). Efficient tests of stock return predictability. *Journal of Financial Economics*, 81(1):27–60.
- Cavanagh, C. L., Elliott, G., and Stock, J. H. (1995). Inference in models with nearly integrated regressors. *Econometric Theory*, 11(05):1131–1147.
- Cochrane, J. H. (2008). The dog that did not bark: A defense of return predictability. *Review of Financial Studies*, 21(4):1533–1575.
- Cochrane, J. H. (2011). Presidential address: Discount rates. *The Journal of Finance*, 66(4):1047–1108.
- Deng, A. (2013). Understanding spurious regression in financial economics. *Journal of Financial Econometrics*, 12(1):122–150.
- Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):134–144.
- Ferson, W. E., Sarkissian, S., and Simin, T. T. (2003). Spurious regressions in financial economics? *Journal of Finance*, 58(4):1393–1414.
- Goyal, A. and Welch, I. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 21(4):1455–1508.
- Jansson, M. and Moreira, M. J. (2006). Optimal inference in regression models with nearly integrated regressors. *Econometrica*, 74(3):681–714.
- Koijen, R. S. and Van Nieuwerburgh, S. (2011). Predictability of returns and cash flows. *Annual Review of Financial Economics*, 3:467–491.

- Kostakis, A., Magdalinos, T., and Stamatogiannis, M. P. (2015). Robust econometric inference for stock return predictability. *Review of Financial Studies*, 28(5):1506–1553.
- Lettau, M. and Ludvigson, S. (2001). Consumption, aggregate wealth, and expected stock returns. *the Journal of Finance*, 56(3):815–849.
- Lewellen, J. (2004). Predicting returns with financial ratios. *Journal of Financial Economics*, 74(2):209–235.
- Lucivjanska, K. (2018). Is imperfection better? evidence from predicting stock and bond returns. *Journal of Financial Econometrics*, 16(2):244–270.
- Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. *Handbook of Econometrics*, 4:2111–2245.
- Pástor, L. and Stambaugh, R. F. (2009). Predictive systems: Living with imperfect predictors. *The Journal of Finance*, 64(4):1583–1628.
- Phillips, P. C. (1988). Regression theory for near-integrated time series. *Econometrica*, 56(5):1021–1043.
- Phillips, P. C. (2014). On confidence intervals for autoregressive roots and predictive regression. *Econometrica*, 82(3):1177–1195.
- Phillips, P. C. B. (1987). Towards a unified asymptotic theory for autoregression. *Biometrika*, 74(3):535–547.
- Stambaugh, R. F. (1999). Predictive regressions. *Journal of Financial Economics*, 54(3):375–421.
- Torous, W., Valkanov, R., and Yan, S. (2004). On predicting stock returns with nearly integrated explanatory variables. *The Journal of Business*, 77(4):937–966.
- Wachter, J. A. and Warusawitharana, M. (2009). Predictable returns and asset allocation: Should a skeptical investor time the market? *Journal of Econometrics*, 148(2):162–178.

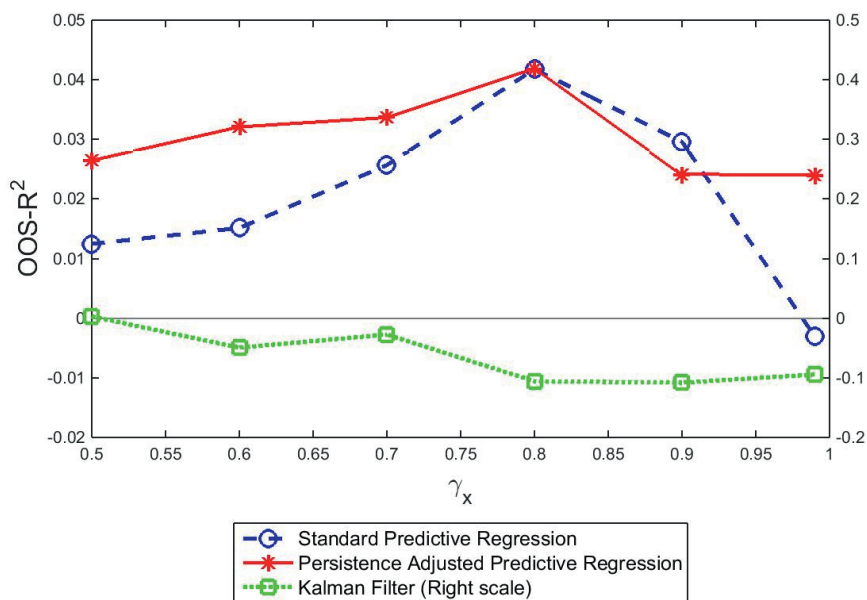
Wachter, J. A. and Warusawitharana, M. (2015). What is the chance that the equity premium varies over time? evidence from regressions on the dividend-price ratio. *Journal of Econometrics*, 186(1):74–93.

Figure 2: Simulated in-sample model fit



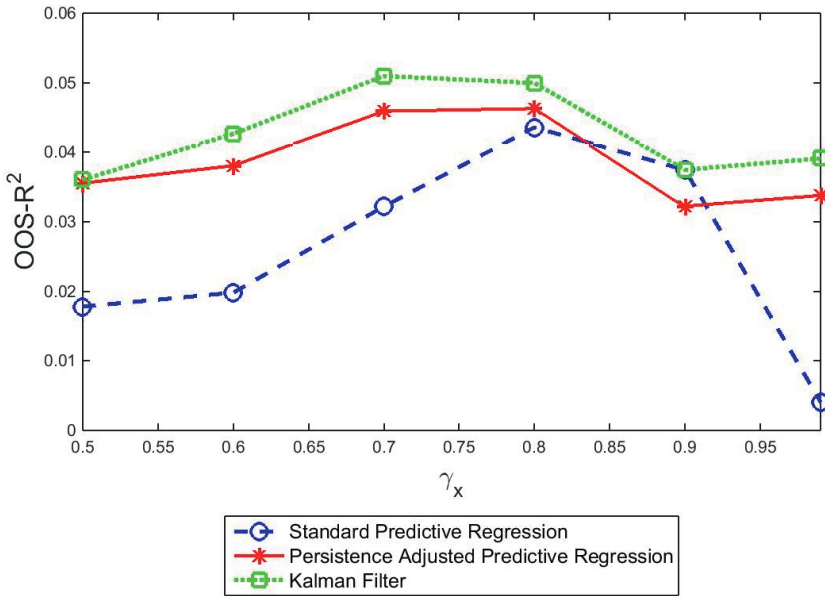
Notes: This plot shows the in-sample R^2 of the predictive regression and the PAPR as a function of the autoregressive parameter of the predictor. The results are based on a Monte Carlo simulation with 1,000 repetitions. The parameter choices are as in the description for Table 1. In particular, the vertical line indicates that the autoregressive parameter of the expected returns is set to $\gamma_\mu = 0.9$.

Figure 3: Simulated out-of-sample model fit when parameters are estimated



Notes: These plots show the out-of-sample R^2 of the standard predictive regression (dashed line), the PAPR (solid line), and the Kalman filter (dotted line). The autoregressive parameter of the expected returns is set to $\gamma_\mu = 0.9$, and results are shown as a function of the persistence parameter of the predictor. The other parameters are set to their default values as described in the text and in Table 1. The benchmark model is the historical mean forecast and the results are based on 1,000 repetitions.

Figure 4: Simulated out-of-sample model fit when parameters are imposed



Notes: This plot shows the out-of-sample R^2 of the standard predictive regression (dashed line), the PAPR (solid line), and the Kalman filter (dotted line), where the parameters are not estimated (the true parameters are imposed). The autoregressive parameter of the expected returns is set to $\gamma_\mu = 0.9$, and results are a function of the persistence parameter of the predictor. The rest of the parameters are set to their default values as described in the text and in Table 1. The benchmark model is the historical mean forecast and the results are based on 1,000 repetitions.

Table 1: **Estimated persistence of the expected returns**

This table shows how OLS (Panel a) and the persistence adjusted predictive regression (Panel b) capture the persistence of the expected returns. The columns are the mean and standard deviation of the autoregressive coefficient of the expected returns, measured by the sample first-order autocorrelation. The data-generating process is the predictive system in equation (2)–(4). Each row indicates the persistence parameter of the predictor used in the simulation. Otherwise, default parameter values are used: $\bar{\mu} = 0.018$, $\bar{x} = 0.03$, $\eta^2 = 0.05$, $\sigma_u = 0.081$, $\gamma_\mu = 0.9$, $\sigma_w = 0.011$, $\rho_{uw} = -0.7$, $\rho_{w\epsilon} = 0.9$, $\sigma_\epsilon = 0.011 \frac{1-\gamma_x^2}{1-0.9\gamma_x}$. All the results are based on a sample size of $T = 200$ and 1,000 repetitions.

(a) Predictive Regression

	Mean $\hat{\gamma}_\mu$	S.e. $\hat{\gamma}_\mu$
$\gamma_x = 0.5$	0.4862	0.0622
$\gamma_x = 0.6$	0.5843	0.0595
$\gamma_x = 0.7$	0.6846	0.0519
$\gamma_x = 0.8$	0.7817	0.0448
$\gamma_x = 0.9$	0.8810	0.0343
$\gamma_x = 0.99$	0.9772	0.0177

(b) Persistence Adjusted Predictive Regression

	Mean $\hat{\gamma}_\mu$	S.e. $\hat{\gamma}_\mu$
$\gamma_x = 0.5$	0.8051	0.2266
$\gamma_x = 0.6$	0.8224	0.2092
$\gamma_x = 0.7$	0.8395	0.1681
$\gamma_x = 0.8$	0.8495	0.1497
$\gamma_x = 0.9$	0.8571	0.1476
$\gamma_x = 0.99$	0.8083	0.1788

Table 2: **Descriptive statistics**

This table includes descriptive statistics for variables used in the main empirical analysis. The data are quarterly, running from the first quarter of 1952 until the fourth quarter of 2016. The first two columns are the mean and the standard deviation of the variables. The third column is the estimated slope coefficient of a first-order autoregressive process.

	mean	stdev	γ_x	N
dp	0.0308	0.0111	0.967	260
cay	-2.07e-05	0.0125	0.822	260
by	7.12e-05	0.00531	0.612	260
ret	0.0181	0.0824	0.0821	260

Table 3: **Estimation results**

This table presents the estimation results for the predictive regression with and without persistence adjustment. The first three columns show the results based on one predictor, while the last column shows the results when all predictors are included. Panel (a) includes the estimates of univariate (columns 1–3) and multivariate OLS regressions. The slope coefficient estimates are shown in rows one through three. The last row shows the implied autocorrelation of the expected returns, that is, the first-order autocorrelation of the process $\hat{\mu}_t = \sum_{j=1}^J \hat{\beta}_j x_j$. Standard errors are given in parentheses. Panel (b) shows results for the PAPR. The first three rows are the innovation slope coefficients, and the last row is the estimated persistence of the expected returns. Innovations in the first step are obtained through a first-order autoregressive filter. Standard errors in parentheses are calculated using the asymptotic formula given in Appendix C. The bootstrap bias-corrected version of the autoregressive parameter (based on a residual bootstrap approach with 200 repetitions) is shown in square brackets. The sample runs from the first quarter of 1952 to the last quarter of 2016. The dependent variable is the one-step ahead excess return. *, **, and *** indicate significance at the 10, 5 and 1 percent levels, respectively.

(a) Predictive Regression				
	dp	cay	by	full
dp	0.9196 (0.4649)**			0.7655 (0.4604)*
cay		1.6051 (0.4106)***		1.2921 (0.4230)***
by			2.9109 (1.1792)**	2.7117 (1.1242)**
γ_μ (implied)	0.9644 (0.0147)***	0.8248 (0.0340)***	0.6145 (0.0849)***	0.7768 (0.0423)***

(b) PAPR (Second step)				
	dp	cay	by	full
dp	2.1933 (1.5889)			3.0026 (1.0243)***
cay		1.2293 (0.3658)***		0.9136 (0.4169)**
by			2.7669 (1.2236)**	2.0009 (0.8130)**
γ_μ	0.8762 (0.1236)*** [0.8953]	0.9235 (0.0395)*** [0.9282]	0.6808 (0.2155)*** [0.7006]	0.9165 (0.0444)*** [0.9269]

Table 4: **PAPR (first stage)**

This table shows the first-step estimates of the two-step procedures described in the text. The sample runs between the first quarter of 1952 and the last quarter of 2016. Since a first-order autoregressive process is used, the estimated parameters are the intercept (first column) and the scalar autoregressive parameter (second column). Standard errors are based on the usual OLS formula. *, **, and *** indicate significance at the 10, 5 and 1 percent levels, respectively.

	Constant	AR(1)
dp	0.0009 (0.0007)	0.9671 (0.0169)*
cay	-0.0000 (0.0004)	0.8223 (0.0376)*
by	0.0000 (0.0003)	0.6121 (0.0293)*

Table 5: **Persistence Adjusted Predictive Regression (restricted estimates)**

This table shows how restrictions affect estimation results for the two-step estimates. The sample runs between the first quarter of 1952 and the last quarter of 2016. The first column is the unrestricted model, while results in the second and third columns are obtained by fixing the autoregressive coefficient of the expected returns. Innovations in the first step are obtained through a first-order autoregressive filter. The dependent variable is the one-step-ahead excess return. Standard errors in parenthesis are calculated using the asymptotic formula given in Appendix C. *, **, and *** indicate significance at the 10, 5 and 1 percent levels, respectively.

	Full	Restricted ($\gamma_\mu = 0.8$)	Restricted ($\gamma_\mu = 0.95$)
b_{dp}	3.0026 (1.0243)***	2.7873 (1.4966)*	2.6010 (0.8032)***
b_{cay}	0.9136 (0.4169)**	1.1948 (0.6610)*	0.7950 (0.3419)**
b_{by}	2.0009 (0.8130)**	2.5758 (1.0418)**	1.5699 (0.7087)**
γ_μ	0.9165 (0.0444)*** [0.9269]	0.8000	0.9500

Table 6: Model comparison

These tables present the performance measures for different model specifications. Panels (a) to (c) include the results based on univariate specifications using the dividend–price ratio, cay and bond yield, respectively. Panel (d) shows the results for when all the variables are included simultaneously. The first column in each table presents in-sample R^2 values based on the formula in equation (11). The second column is the out-of-sample R^2 defined in equation (12). Positive values indicate that the model performs better than the historical mean. The first row indicates standard predictive regression results, and the second row is the unrestricted PAPR. The third and fourth rows are restricted versions of the PAPR, where the persistence of the expected returns is fixed. The last row shows results based on the maximum likelihood estimates of the full system using the Kalman filter. Only one-step-ahead forecast horizons are considered, and the initial training sample is taken to be 40 percent of the entire sample. The third column is the test statistic of the Diebold–Mariano test as in Diebold and Mariano (1995).

(a) dividend–price ratio			
	IS- R^2	OOS- R^2	DM test statistic
OLS	0.0153	-0.0299	-7.7337
PAPR	0.0185	-0.0486	-12.8341
Restricted PAPR ($\gamma_\mu = 0.8$)	0.0150	-0.0069	-2.5791
Restricted PAPR ($\gamma_\mu = 0.95$)	0.0140	-0.0355	-9.9231
Kalman Filter	0.0069	-0.7463	-14.2686
(b) cay			
	IS- R^2	OOS- R^2	DM test statistic
OLS	0.0587	0.0401	7.3666
PAPR	0.0685	0.0367	6.9956
Restricted PAPR ($\gamma_\mu = 0.8$)	0.0559	0.0307	5.6796
Restricted PAPR ($\gamma_\mu = 0.95$)	0.0679	0.0504	10.6227
Kalman Filter	0.0870	0.0037	0.6103
(c) bond yield			
	IS- R^2	OOS- R^2	DM test statistic
OLS	0.0425	-0.0050	-0.7302
PAPR	0.0449	-0.0210	-3.0038
Restricted PAPR ($\gamma_\mu = 0.8$)	0.0444	-0.0054	-0.8058
Restricted PAPR ($\gamma_\mu = 0.95$)	0.0321	0.0047	0.9358
Kalman Filter	0.0442	-0.0806	-9.1279
(d) all			
	IS- R^2	OOS- R^2	DM test statistic
OLS	0.0927	-0.0049	-0.6252
PAPR	0.1114	0.0087	1.0897
Restricted PAPR ($\gamma_\mu = 0.8$)	0.0944	0.0069	0.8341
Restricted PAPR ($\gamma_\mu = 0.95$)	0.1093	0.0217	3.0534
Kalman Filter	0.0186	-0.8611	-15.2522

Appendix

A t-statistics and R^2 of the predictive regression

The results reflecting equation (7) can be derived for the t-statistic and the R^2 of the predictive regression. Begin with the variance of the regression error:

$$\begin{aligned}
 Var(\hat{e}_{t+1}) &= Var(r_{t+1} - \bar{r}) + \hat{\beta}_{OLS}^2 Var(x_t - \bar{x}) - 2\hat{\beta}_{OLS} Cov(r_{t+1} - \bar{r}, x_t - \bar{x}) \\
 &= \frac{\sigma_w^2}{1 - \gamma_\mu^2} + \sigma_u^2 + \hat{\beta}_{OLS}^2 \frac{\sigma_\epsilon^2}{1 - \gamma_x^2} - 2\hat{\beta}_{OLS} \frac{\sigma_{w\epsilon}}{1 - \gamma_x \gamma_\mu} \\
 &= \frac{\sigma_w^2}{1 - \gamma_\mu^2} + \sigma_u^2 + \frac{\sigma_{w\epsilon}^2}{(1 - \gamma_x \gamma_\mu)^2} \frac{(1 - \gamma_x^2)^2}{\sigma_\epsilon^4} \frac{\sigma_\epsilon^2}{1 - \gamma_x^2} - 2 \frac{\sigma_{w\epsilon}}{1 - \gamma_x \gamma_\mu} \frac{(1 - \gamma_x^2)}{\sigma_\epsilon^2} \frac{\sigma_{w\epsilon}}{1 - \gamma_x \gamma_\mu} \\
 &= \frac{\sigma_w^2}{1 - \gamma_\mu^2} + \sigma_u^2 - \frac{\sigma_{w\epsilon}^2}{(1 - \gamma_x \gamma_\mu)^2} \frac{(1 - \gamma_x^2)}{\sigma_\epsilon^2}.
 \end{aligned}$$

The variance of the OLS estimator of β is

$$\begin{aligned}
 TVar(\hat{\beta}_{OLS}) &= \frac{Var(\hat{e}_{t+1})}{Var(x_t)} = \frac{\sigma_w^2}{\sigma_\epsilon^2} \frac{1 - \gamma_x^2}{1 - \gamma_\mu^2} + \frac{\sigma_u^2(1 - \gamma_x^2)}{\sigma_\epsilon^2} - \rho_{w\epsilon}^2 \frac{\sigma_w^2}{\sigma_\epsilon^2} \frac{(1 - \gamma_x^2)^2}{(1 - \gamma_x \gamma_\mu)^2} \\
 &= \eta^2 \frac{\sigma_u^2}{\sigma_\epsilon^2} (1 - \gamma_x^2) + \frac{\sigma_u^2(1 - \gamma_x^2)}{\sigma_\epsilon^2} - \eta^2 \rho_{w\epsilon}^2 \frac{\sigma_u^2}{\sigma_\epsilon^2} \frac{(1 - \gamma_x^2)^2 (1 - \gamma_\mu^2)}{(1 - \gamma_x \gamma_\mu)^2} \\
 &= \frac{\sigma_u^2(1 - \gamma_x^2)}{\sigma_\epsilon^2} \left[1 + \eta^2 \left(1 - \rho_{w\epsilon}^2 \frac{(1 - \gamma_x^2)(1 - \gamma_\mu^2)}{(1 - \gamma_x \gamma_\mu)^2} \right) \right],
 \end{aligned}$$

which implies the following expression for the t-statistic:

$$t_{\hat{\beta}_{OLS}} = \frac{\hat{\beta}_{OLS}}{\sqrt{Var(\hat{\beta}_{OLS})}} = \sqrt{T} \frac{\eta \rho_{w\epsilon} \frac{\sqrt{1 - \gamma_\mu^2} \sqrt{1 - \gamma_x^2}}{1 - \gamma_x \gamma_\mu}}{\sqrt{1 + \eta^2 \left[1 - \rho_{w\epsilon}^2 \left(\frac{(1 - \gamma_\mu^2)(1 - \gamma_x^2)}{(1 - \gamma_x \gamma_\mu)^2} \right) \right]}}.$$

The expression above is scale-free and essentially depends on the same factors as the slope coefficient: the underlying amount of predictability, the degree of imperfection and the difference between the persistence of the predictor and the expected returns. Both η and $\rho_{w\epsilon}$ are positively related to the t-statistic. A higher underlying amount of predictability and a larger correlation between the predictor and the expected returns

suggest a stronger signal, which should imply a greater t-statistic. The effect of the autocorrelation parameters on the t-statistic is also similar. The t-statistic is maximal if the two autocorrelations coincide; otherwise, it decreases in the difference between the two values. The decrease is faster the closer the parameters are to unity. This implies that the persistence of the regressor substantially affects the inference, in line with the results in Appendix B.

A decomposition analogous to (7) can also be given for the regression R^2 ,

$$R_{OLS}^2 = \frac{\hat{\beta}_{OLS}^2 \text{Var}(x_t)}{\text{Var}(r_{t+1})} = R_{true}^2 \rho_{w\epsilon}^2 \frac{(1 - \gamma_x^2)(1 - \gamma_\mu^2)}{(1 - \gamma_x \gamma_\mu)^2}.$$

$R_{OLS}^2 \leq R_{true}^2$, where R_{true}^2 is the R^2 of the predictive regression in the absence of predictor imperfection. Therefore, the explanatory power of *any* predictive regression provides a lower bound for the degree to which expected returns explain return variation.

B Highly persistent predictors

Suppose that the autoregressive parameter of the predictor and the expected returns is given by $\gamma_x = 1 - \frac{c_x}{T}$ and $\gamma_\mu = 1 - \frac{c_\mu}{T}$, respectively, where T is the sample size and c_x and c_μ are constants. This corresponds to the specification in Cavanagh et al. (1995) and Campbell and Yogo (2006). As discussed, predictor imperfection can stem from two sources, either a smaller-than-unity correlation in innovations, or a difference in the autocorrelation of the predictor and the expected returns (which, in this case, is captured by the fact that $c_x \neq c_\mu$).

A correlation coefficient $\rho_{w\epsilon}$ strictly smaller than one implies a spurious predictive regression if the predictor and the expected returns are nearly integrated. Any non-perfect correlation means that the expected return process has a nearly integrated component that is unrelated to the predictor. In this case, the slope coefficient of the predictive regression converges to a random variable, and the t-statistic diverges for any value of $\rho_{w\epsilon}$. Therefore, it is necessary to introduce an asymptotically perfect correlation between the innovations to analyze the effect of predictor imperfection further. To this end, I

introduce the concept of a nearly perfect predictor (analogous to near integration). This implies that, asymptotically, the correlation $\rho_{w\epsilon} \rightarrow 1$ as $T \rightarrow \infty$. I specify $\rho_{w\epsilon} = 1 - \frac{d}{T^\delta}$ for some positive constants d and δ . The asymptotic results on the regression slope coefficient and the t-statistics are summarized in the following proposition.

Proposition 1 *Assume that the data-generating process is given by equations (2) – (4), with $\gamma_x = 1 - \frac{c_x}{T}$ and $\gamma_\mu = 1 - \frac{c_\mu}{T}$ for some $c_x, c_\mu \geq 0$, and the correlation between the predictor and the expected returns can be written as $\rho_{ew} = 1 - \frac{d}{T^\delta}$ for some $d > 0$ and $\delta \geq 0$. Define $\beta^* = \frac{\sigma_w}{\sigma_\epsilon}$. The following results hold:*

1. *The slope coefficient satisfies*

$$\begin{aligned} T^{\delta/2} \left(\hat{\beta}_{OLS} - \beta^* \right) &\Rightarrow \sqrt{2d} \frac{\sigma_w \int \underline{J}_{c_x} \tilde{J}_{c_\mu}}{\sigma_\epsilon \int \underline{J}_{c_x}^2} && \text{if } \delta < 2, \\ T \left(\hat{\beta}_{OLS} - \beta^* \right) &\Rightarrow \frac{\sigma_u \int \underline{J}_{c_x} dB_u}{\sigma_\epsilon \int \underline{J}_{c_x}^2} + \sqrt{2d} \frac{\sigma_w \int \underline{J}_{c_x} \tilde{J}_{c_\mu}}{\sigma_\epsilon \int \underline{J}_{c_x}^2} && \text{if } \delta = 2, \\ T \left(\hat{\beta}_{OLS} - \beta^* \right) &\Rightarrow \frac{\sigma_u \int \underline{J}_{c_x} dB_u}{\sigma_\epsilon \int \underline{J}_{c_x}^2} && \text{if } \delta > 2. \end{aligned}$$

2. *The t-statistic*

$$\begin{aligned} t_{\hat{\beta}_{OLS}} &\text{ diverges} && \text{if } \delta < 2, \\ t_{\hat{\beta}_{OLS}} &\Rightarrow (1 - \rho_{\epsilon u}^2)^{1/2} \frac{\int \underline{J}_{c_x} dW_{\tilde{u}}}{\left(\int \underline{J}_{c_x}^2 \right)^{1/2}} + \rho_{\epsilon u} \frac{\int \underline{J}_{c_x} dW_x}{\left(\int \underline{J}_{c_x}^2 \right)^{1/2}} + \sqrt{2d} \frac{\sigma_w \int \underline{J}_{c_x} \tilde{J}_{c_\mu}}{\sigma_u \left(\int \underline{J}_{c_x}^2 \right)^{1/2}} && \text{if } \delta = 2, \\ t_{\hat{\beta}_{OLS}} &\Rightarrow (1 - \rho_{\epsilon u}^2)^{1/2} \frac{\int \underline{J}_{c_x} dW_{\tilde{u}}}{\left(\int \underline{J}_{c_x}^2 \right)^{1/2}} + \rho_{\epsilon u} \frac{\int \underline{J}_{c_x} dW_x}{\left(\int \underline{J}_{c_x}^2 \right)^{1/2}} && \text{if } \delta > 2, \end{aligned}$$

where the processes J_{c_x} and \tilde{J}_{c_μ} are defined below in Lemma 1.

Before the proof of Proposition 1, I collect some useful standard results related to local-to-unity asymptotics in a lemma (based on Phillips, 1987).

Lemma 1 *Let $v_t = (u_t, w_t, \epsilon_t)'$ be a serially uncorrelated martingale difference sequence with $E(v_t v_t') = \Sigma \forall t$. Define the nearly integrated processes $\xi_t = (\tilde{u}_t, \mu_t, x_t)'$ such that*

their time-series dynamics can be written as

$$\xi_t = \begin{pmatrix} 1 - \frac{c_u}{T} & 0 & 0 \\ 0 & 1 - \frac{c_\mu}{T} & 0 \\ 0 & 0 & 1 - \frac{c_x}{T} \end{pmatrix} \xi_{t-1} + v_t \equiv \left(I - \frac{1}{T} C \right) \xi_{t-1} + v_t.$$

Then, by standard results (for example, in Phillips, 1988)

$$\begin{aligned} \frac{1}{T^{-3/2}} \sum_{t=2}^T \xi_{t-1} &\Rightarrow \Sigma^{1/2} \int_0^1 \mathbf{J}_C(r) dr \\ \frac{1}{T^{-2}} \sum_{t=2}^T \xi_{t-1} \xi'_{t-1} &\Rightarrow \Sigma^{1/2} \int_0^1 \mathbf{J}_C(r) \mathbf{J}_C(r)' dr \Sigma^{1/2} \\ \frac{1}{T^{-1}} \sum_{t=2}^T \xi_{t-1} v'_t &\Rightarrow \Sigma^{1/2} \int_0^1 \mathbf{J}_C(r) d\mathbf{W}(r)' dr \Sigma^{1/2}, \end{aligned}$$

where \mathbf{W} is a standard three-dimensional Brownian motion and \mathbf{J}_C satisfies the stochastic differential equation

$$d\mathbf{J}_C(r) = C\mathbf{J}_C(r)dr + d\mathbf{W}(r); \quad \mathbf{J}_C(0) = 0.$$

The joint convergence result also implies that the result holds for individual processes and pairs of processes, too. Note that we can use the usual orthogonal decomposition for the processes due to the lack of autocorrelation in the innovations. In particular, $W_\mu = \rho_{ew}W_x + \sqrt{1 - \rho_{ew}^2}W_{\tilde{\mu}}$, where $W_{\tilde{\mu}}$ is a standard Wiener process, independent from W_x . We can build up a \tilde{J}_{c_μ} process, which is independent from J_{c_x} , using this decomposition for W_μ . The results hold for the demeaned process $\underline{\xi}_t = \xi_t - \frac{1}{T} \sum_{t=1}^T \xi_t$, replacing $\mathbf{J}_C(r)$ with $\underline{\mathbf{J}}_C(r) = \mathbf{J}_C(r) - \int_0^1 \mathbf{J}_C(r) dr$.

Proof of Proposition 1.

Consider first the OLS estimate of the slope coefficient,

$$\begin{aligned}
\hat{\beta}_{OLS} &= \frac{\sum_{t=1}^{T-1} (r_{t+1} - \bar{r})(x_t - \bar{x})}{\sum_{t=1}^{T-1} (x_t - \bar{x})^2} = \frac{\sum_{t=1}^{T-1} (\mu_t - \bar{\mu})(x_t - \bar{x})}{\sum_{t=1}^{T-1} (x_t - \bar{x})^2} + \frac{\sum_{t=1}^{T-1} u_{t+1}(x_t - \bar{x})}{\sum_{t=1}^{T-1} (x_t - \bar{x})^2} \\
&= \frac{\sum_{t=1}^{T-1} \left(1 - \frac{d}{T^\delta}\right) \frac{\sigma_w}{\sigma_\epsilon} (x_t - \bar{x})^2}{\sum_{t=1}^{T-1} (x_t - \bar{x})^2} + \frac{\sum_{t=1}^{T-1} u_{t+1}(x_t - \bar{x})}{\sum_{t=1}^{T-1} (x_t - \bar{x})^2} + \frac{\sum_{t=1}^{T-1} \sqrt{\frac{2d}{T^\delta} - \frac{d^2}{T^{2\delta}}} (\tilde{\mu}_{t+1} - \bar{\mu})(x_t - \bar{x})}{\sum_{t=1}^{T-1} (x_t - \bar{x})^2} \\
&= \left(1 - \frac{d}{T^\delta}\right) \frac{\sigma_w}{\sigma_\epsilon} + \frac{\sum_{t=1}^{T-1} u_{t+1}(x_t - \bar{x})}{\sum_{t=1}^{T-1} (x_t - \bar{x})^2} + \frac{\sum_{t=1}^{T-1} \sqrt{\frac{2d}{T^\delta} - \frac{d^2}{T^{2\delta}}} (\tilde{\mu}_{t+1} - \bar{\mu})(x_t - \bar{x})}{\sum_{t=1}^{T-1} (x_t - \bar{x})^2},
\end{aligned}$$

where $\tilde{\mu}_{t+1}$ is a local-to-unity process with autoregressive parameter $1 - c_\mu/T$ and innovations uncorrelated with ϵ_t . Based on the results in Lemma 1,

$$\begin{aligned}
T^{\delta/2} \sqrt{\frac{2d}{T^\delta} - \frac{d^2}{T^{2\delta}}} \frac{\frac{1}{T} \sum_{t=1}^{T-1} (\tilde{\mu}_{t+1} - \bar{\mu})(x_t - \bar{x})}{\frac{1}{T} \sum_{t=1}^{T-1} (x_t - \bar{x})^2} &\Rightarrow \sqrt{2d} \frac{\sigma_w \int \underline{J}_{c_x} \tilde{J}_{c_\mu}}{\sigma_\epsilon \int \underline{J}_{c_x}^2}, \\
T \frac{\frac{1}{T} \sum_{t=1}^{T-1} u_{t+1}(x_t - \bar{x})}{\frac{1}{T} \sum_{t=1}^{T-1} (x_t - \bar{x})^2} &\Rightarrow \frac{\sigma_u \int \underline{J}_{c_x} dW_u}{\sigma_\epsilon \int \underline{J}_{c_x}^2}.
\end{aligned}$$

The joint weak convergence also holds. Defining $\beta^* = \frac{\sigma_w}{\sigma_\epsilon}$ yields the first part of the results in Proposition 1. The result implies that the OLS estimate is consistent. That is,

$$\hat{\beta}_{OLS} \xrightarrow{P} \beta,$$

and $\hat{\beta}_{OLS} - \beta = O_p(T^{-\min\{1, \delta/2\}})$, which can also be written as $\hat{\beta}_{OLS} - \beta = \frac{\zeta}{T^{\min\{1, \delta/2\}}}$, where ζ is a random variable ($O_p(1)$). The fitted residuals and their sum of squares can thus be written as

$$\begin{aligned}
\hat{u}_{t+1} &= r_{t+1} - \hat{\beta}_{OLS}(x_t - \bar{x}) \\
&= \mu_t - \bar{\mu}_t + u_{t+1} - \bar{u}_t - \hat{\beta}_{OLS}(x_t - \bar{x}) \\
&= \left(1 - \frac{d}{T^\delta}\right) \beta(x_t - \bar{x}) + \sqrt{\frac{2d}{T^\delta} - \frac{d^2}{T^{2\delta}}} (\tilde{\mu} - \bar{\mu}) - \hat{\beta}_{OLS}(x_t - \bar{x}) + (u_{t+1} - \bar{u}) \\
&= \left(\beta - \hat{\beta}_{OLS} - \frac{d}{T^\delta}\right) (x_t - \bar{x}) + \sqrt{\frac{2d}{T^\delta} - \frac{d^2}{T^{2\delta}}} (\tilde{\mu} - \bar{\mu}) + (u_{t+1} - \bar{u})
\end{aligned}$$

and

$$\begin{aligned}
\frac{1}{T} \sum_{t=1}^{T-1} \hat{u}_{t+1}^2 &= \frac{1}{T} \left(\beta - \hat{\beta}_{OLS} - \frac{d}{T^\delta} \right)^2 \sum_{t=1}^{T-1} (x_t - \bar{x})^2 + \frac{1}{T} \left(\frac{2d}{T^\delta} - \frac{d^2}{T^{2\delta}} \right) \sum_{t=1}^{T-1} (\tilde{\mu} - \bar{\mu})^2 + \frac{1}{T} \sum_{t=1}^{T-1} (u_{t+1} - \bar{u})^2 \\
&+ 2 \frac{1}{T} \left(\beta - \hat{\beta}_{OLS} - \frac{d}{T^\delta} \right) \sqrt{\frac{2d}{T^\delta} - \frac{d^2}{T^{2\delta}}} \sum_{t=1}^{T-1} (x_t - \bar{x}) (\tilde{\mu} - \bar{\mu}) \\
&+ 2 \frac{1}{T} \left(\beta - \hat{\beta}_{OLS} - \frac{d}{T^\delta} \right) \sum_{t=1}^{T-1} (x_t - \bar{x}) (u_{t+1} - \bar{u}) \\
&+ 2 \frac{1}{T} \sqrt{\frac{2d}{T^\delta} - \frac{d^2}{T^{2\delta}}} \sum_{t=1}^{T-1} (\tilde{\mu} - \bar{\mu}) (u_{t+1} - \bar{u}) \\
&= \frac{1}{T} \sum_{t=1}^{T-1} (u_{t+1} - \bar{u})^2 + o_p(1).
\end{aligned}$$

The variance of the OLS estimator of β can now be written as

$$\text{Var}(\hat{\beta}_{OLS}) = \frac{1}{T} \frac{\frac{1}{T} \sum_{t=1}^{T-1} \hat{u}_{t+1}^2}{\frac{1}{T} \sum_{t=1}^{T-1} (x_t - \bar{x})^2} = \frac{1}{T} \frac{\zeta^2 \left(o_p(1) + \frac{1}{T\zeta^2} \sum_{t=1}^{T-1} (u_{t+1} - \bar{u})^2 \right)}{\frac{1}{T} \sum_{t=1}^{T-1} (x_t - \bar{x})^2}$$

and the t-statistic (disregarding the asymptotically vanishing term) is

$$t_{\hat{\beta}_{OLS}} = \frac{\hat{\beta}_{OLS} - \beta}{\sqrt{\text{Var}(\hat{\beta}_{OLS})}} = \frac{\frac{\zeta}{T^{\min\{1, \delta/2\}}}}{\frac{\zeta}{\sqrt{T}} \sqrt{\frac{1}{T\zeta^2} \sum_{t=1}^{T-1} (u_{t+1} - \bar{u})^2}} = T^{\max\{-\frac{1}{2}, \frac{1-\delta}{2}\}} \frac{\zeta \sqrt{\frac{1}{T} \sum_{t=1}^{T-1} (x_t - \bar{x})^2}}{\sqrt{\frac{1}{T} \sum_{t=1}^{T-1} (u_{t+1} - \bar{u})^2}}.$$

If $\delta < 2$, the expression above diverges, establishing the first line of the second part of the results. Otherwise, the expression above converges to

$$t_{\hat{\beta}_{OLS}} \Rightarrow \frac{\zeta \left(\sigma_\epsilon^2 \int \underline{J}_{c_x}^2 \right)^{1/2}}{\sigma_u} = \frac{\int \underline{J}_{c_x} dW_u}{\left(\int \underline{J}_{c_x}^2 \right)^{1/2}} + \sqrt{2d} \frac{\sigma_w \int \underline{J}_{c_x} \tilde{J}_{c_\mu}}{\sigma_u \left(\int \underline{J}_{c_x}^2 \right)^{1/2}} \quad \text{if } \delta = 2,$$

and

$$t_{\hat{\beta}_{OLS}} \Rightarrow \frac{\zeta \left(\sigma_\epsilon^2 \int \underline{J}_{c_x}^2 \right)^{1/2}}{\sigma_u} = \frac{\int \underline{J}_{c_x} dW_u}{\left(\int \underline{J}_{c_x}^2 \right)^{1/2}} \quad \text{if } \delta > 2.$$

The remaining results are then established by decomposing the innovations in the last term on the right-hand side as $W_u = \rho_{eu} W_x + (1 - \rho_{eu}^2)^{1/2} W_{\bar{u}}$, where $W_{\bar{u}}$ is a standard

Wiener process independent of W_x . ■

Proposition 1 distinguishes between three cases. First, if the predictor imperfection is strong ($\delta < 2$), spurious regression dominates. The convergence of β is slow (or it does not converge in probability if $\delta = 0$), and the t-statistic diverges, indicating strong statistical significance, even if there is no relationship between the predictor and the expected returns. That is, the predictive regression is spurious if $\delta < 2$. It is analogous to the result in Ferson et al. (2003) and Deng (2013) dealing with completely spurious regressions in the context of return predictability ($\delta = 0$ and $d = 1$ nest the specification discussed in Deng, 2013). In this case, the endogeneity of the predictor has a second-order effect. The second case, $\delta > 2$, is then asymptotically equivalent to the situation, in which the predictor perfectly correlates with the expected returns. The estimation of β is consistent and the t-statistic converges to the expression derived in Cavanagh et al. (1995). Third, in the knife-edge case, $\delta = 2$, predictor imperfection has a non-trivial effect on the results. The slope coefficient can be estimated consistently, but the t-statistic becomes non-standard. Particularly, it is a linear combination of three terms. The first one is a standard normal distribution, that is, in the absence of endogeneity and imperfection ($\rho_{eu} = d = 0$), the t-statistic asymptotically follows a standard normal distribution. The second part enters because of the potential endogeneity of the regressor, and the third component indicates the effect of predictor imperfection. The relative strengths of the three terms are determined by the correlation structure of the innovations.

C The general formulation of the two-step procedure

Let the predictive system be described by

$$r_{t+1} = \mu_t + u_{t+1} \tag{13}$$

$$\mu_{t+1} = (1 - \gamma_\mu)\bar{\mu} + \gamma_\mu\mu_t + w_{t+1} \tag{14}$$

$$\mathbf{x}_{t+1} = \bar{\mathbf{x}} + A(L)\mathbf{x}_t + \boldsymbol{\epsilon}_t, \tag{15}$$

where \mathbf{x}_t is a K -dimensional vector process, and $A(L)$ is a lag polynomial satisfying assumptions that ensure the stationarity of \mathbf{x}_t . The innovations $\{u_t, w_t, \boldsymbol{\epsilon}_t\}$ jointly follow martingale difference sequences with a contemporaneous covariance matrix Σ . Let $\theta = \{\theta_1, \theta_2\}$ be the parameters of the model to be estimated, with $\theta_1 = \{\bar{\mathbf{x}}, A\}$ and $\theta_2 = \{\bar{\mu}, \gamma_\mu, \mathbf{b}\}$, as in the main text.

In the first step, the parameters of equation (15) are estimated (for example, equation-by-equation OLS), and the predictor innovations are calculated as

$$\hat{\boldsymbol{\epsilon}}_{t+1} = \mathbf{x}_{t+1} - \hat{\bar{\mathbf{x}}} - \hat{A}(L)\mathbf{x}_t.$$

The second step consists of minimization of the criterion function. That is,

$$\hat{\theta}_2 = \arg \min_{\theta_2} Q(\hat{\theta}_1, \theta_2) = \arg \min_{\theta_2} \frac{1}{T-1} \sum_{t=1}^{T-1} \left(r_{t+1} - \bar{\mu} + \sum_{s=1}^t \gamma_\mu^{t-s} \mathbf{b}' \hat{\boldsymbol{\epsilon}}_t \right)^2. \quad (16)$$

The asymptotic properties of the two-step estimator, which is a special case of extremum estimators, is well-known (for example Newey and McFadden, 1994). Given a consistent estimator for the first-step parameters $\hat{\theta}_1$, the remaining parameters (θ_2) can be consistently estimated in the second step, treating the first-step parameters as known.

The asymptotic variance of the second step estimator is, however, affected by the first step results. Especially, the standard errors of the second-step estimates must consider the estimated first-step parameters. Let $y_t = \{r_{t+1}, \mathbf{x}_t, \mathbf{x}_{t-1}, \dots\}$ denote the data, and

$$q(\theta_1, \theta_2, y_t) = \left(r_{t+1} - \bar{\mu} + \sum_{s=1}^t \gamma_\mu^{t-s} \mathbf{b}' \boldsymbol{\epsilon}_t \right)^2,$$

where the dependence on θ_1 is through the innovations $\boldsymbol{\epsilon}_t$. Let the random variable $r_t(\theta_1)$ be defined as

$$\sqrt{T}(\hat{\theta}_1 - \theta_1) = T^{-1/2} \sum_{t=1}^T r_t(\theta_1) + o_p(1).$$

Then, the asymptotic distribution of the second step estimator is given by

$$\begin{aligned}\sqrt{T}(\hat{\theta}_2 - \theta_2) &\xrightarrow{d} N(0, \Gamma_0^{-1} \Omega_0 \Gamma_0^{-1}), \\ \Gamma_0 &= E \left(\frac{\partial^2 q(\theta_1, \theta_2, y_t)}{\partial \theta_2 \partial \theta_2'} \right), \\ \Omega_0 &= E (s(\theta_1, \theta_2, y_t) s(\theta_1, \theta_2, y_t)'), \\ s(\theta_1, \theta_2, y_t) &= \frac{\partial q(\theta_1, \theta_2, y_t)}{\partial \theta_2} + \frac{\partial^2 q(\theta_1, \theta_2, y_t)}{\partial \theta_2 \partial \theta_1'} r_t(\theta_1).\end{aligned}$$

The effect of the first-step estimation appears through the second term of the $s(\theta_1, \theta_2, y_t)$ function (in a usual one-step estimator $\frac{\partial^2 q(\theta_1, \theta_2, y_t)}{\partial \theta_2 \partial \theta_1'} = 0$). When calculating the standard errors, the matrices Γ_0, Ω_0 are replaced by consistent estimators

$$\begin{aligned}\hat{\Gamma}_0 &= \frac{1}{T} \sum_{t=1}^T \frac{\partial^2 q(\hat{\theta}_1, \theta_2, y_t)}{\partial \theta_2 \partial \theta_2'} \Big|_{\theta_2 = \hat{\theta}_2}, \\ \hat{\Omega}_0 &= \frac{1}{T} \sum_{t=1}^T \hat{s}(\hat{\theta}_1, \hat{\theta}_2, y_t) \hat{s}(\hat{\theta}_1, \hat{\theta}_2, y_t)', \\ \hat{s}(\hat{\theta}_1, \hat{\theta}_2, y_t) &= \frac{\partial q(\hat{\theta}_1, \theta_2, y_t)}{\partial \theta_2} \Big|_{\theta_2 = \hat{\theta}_2} + \frac{\partial^2 q(\theta_1, \theta_2, y_t)}{\partial \theta_2 \partial \theta_1'} \Big|_{\theta_1 = \hat{\theta}_1, \theta_2 = \hat{\theta}_2} r_t(\hat{\theta}_1).\end{aligned}$$

D The effect of initialization

The definition of the PAPR estimator assumes that the expected return in $t = 0$ is equal to the sample mean of the return. The high persistence of the expected return process thus implies that the initial value may have an effect not only in the initial period, but in the subsequent periods too. Therefore, if the true initial value of the expected returns is vastly different from its long term mean, it may bias the estimated expected returns in the first part of the sample (simulations suggest that the problem may prevail for 20-30 periods, depending on the persistence of the variables). However, in sample sizes relevant for return predictions, this does not affect the real time forecasting performance, since the effect of the initial value is negligible by the end of the training sample. It does affect the in-sample fit of the model, making it weaker due to the poor fit in the initial periods. Therefore, the in-sample results of the model reported in the study become conservative

estimates of the overall performance of the PAPR estimator.

The current initialization also impacts parameter estimation. The formulation implies that, in the second step, parameters are estimated under a restriction: the projected expected return process starts from the sample mean of the returns. For a robustness check, I perform a modified version of the two-step estimation (which I label the initialization robust PAPR). In this case, the first h observations are omitted when the second-step estimation is performed (where h is an integer, which is either a fixed number or represents a fraction of the entire sample). That is, the second-step parameter estimation problem can be written as

$$\hat{\theta}_2 = \arg \min_{\theta_2} \frac{1}{T-1} \sum_{t=1}^{T-1} H(t) \left(r_{t+1} - \bar{\mu} - \sum_{s=1}^t \gamma_{\mu}^{t-s} b \hat{e}_s \right)^2, \quad (17)$$

$$H(t) = \begin{cases} 0 & \text{if } t \leq h \\ 1 & \text{otherwise} \end{cases}. \quad (18)$$

This modification of the second step ensures that estimates of the parameters of the expected return process are not affected by the observations for which the potential bias due to the initialization is the largest.

I consider the cases $h = \{0.05T, 0.1T, 0.2T\}$, which correspond to $h = 13, 26, 52$, respectively (since $T = 260$ in the empirical specification). The estimation results are presented in Table A1, while in-sample and out-of-sample performances are shown in Table A2. The results are both qualitatively and quantitatively similar to the baseline results in Tables 3 and 6. The only notable difference is that the cay variable performs slightly worse in terms of real-time forecasting after correcting for the initialization.

[Table A1 about here]

[Table A2 about here]

Table A1: **Robustness - Initialization robust PAPR**

This table shows the impact of omitting the first h observation in the second step estimation. The results are obtained based on the second-step objective function in (18). Further details on the specification can be found in the description of Table 3.

(a) $h = 13$

	dp	cay	by	full
dp	2.4475 (1.6026)			3.1689 (1.0560)***
cay		1.1770 (0.3629)***		0.8591 (0.4204)**
by			2.7230 (1.2230)**	2.1055 (0.8299)**
γ_μ	0.8627 (0.1266)*** [0.8663]	0.9247 (0.0404)*** [0.9366]	0.6798 (0.2197)*** [0.6951]	0.9119 (0.0474)*** [0.9168]

(b) $h = 26$

	dp	cay	by	full
dp	2.5523 (1.6081)			3.3080 (1.0664)***
cay		1.1314 (0.3561)***		0.8340 (0.4187)**
by			2.6317 (1.2181)**	2.1669 (0.8348)***
γ_μ	0.8564 (0.1286)*** [0.8551]	0.9273 (0.0399)*** [0.9432]	0.6851 (0.2244)*** [0.7007]	0.9110 (0.0476)*** [0.9153]

(c) $h = 52$

	dp	cay	by	full
dp	2.4458 (1.6090)			3.4455 (1.0320)***
cay		1.0899 (0.3488)***		0.7124 (0.3915)*
by			2.6710 (1.2297)**	2.2754 (0.8256)***
γ_μ	0.8572 (0.1332)*** [0.8535]	0.9302 (0.0392)*** [0.9445]	0.6645 (0.2334)*** [0.6675]	0.9198 (0.0425)*** [0.9332]

Table A2: Model comparison for the initialization robust PAPR

This table shows both in-sample and out-of-sample results for the initialization robust PAPR for different choices of h . Further details on the specification can be found in the description of Table 6.

(a) dividend-price ratio			
	IS-R^2	OOS-R^2	DM test statistic
$h = 13$	0.0183	-0.0386	-9.9980
$h = 26$	0.0181	-0.0316	-7.8336
$h = 52$	0.0182	-0.0651	-13.4851
(b) cay			
	IS-R^2	OOS-R^2	DM test statistic
$h = 13$	0.0684	-0.0024	-0.3960
$h = 26$	0.0682	-0.0016	-0.2665
$h = 52$	0.0680	-0.0020	-0.3462
(c) bond yield			
	IS-R^2	OOS-R^2	DM test statistic
$h = 13$	0.0356	-0.0202	-3.0300
$h = 26$	0.0355	-0.0146	-2.4257
$h = 52$	0.0355	-0.0222	-3.6494
(d) all			
	IS-R^2	OOS-R^2	DM test statistic
$h = 13$	0.1113	0.0136	1.7331
$h = 26$	0.1112	0.0141	1.7815
$h = 52$	0.1106	0.0051	0.6433

Chapter II

Testing Return Predictability with the Dividend-Growth Equation: An Anatomy of the Dog*

Erik Hjalmarsson

Tamás Kiss[†]

Abstract

The dividend-growth based test of return predictability, proposed by Cochrane [2008, Review of Financial Studies 21, 1533-1575], is similar to a likelihood-based test of the standard return-predictability model, treating the autoregressive parameter of the dividend-price ratio as known. In comparison to standard OLS-based inference, both tests achieve power gains from a strong use of the exact value postulated for the autoregressive parameter. When compared to the likelihood-based test, there are no power advantages for the dividend-growth based test. In common implementations, with the autoregressive parameter set equal to the corresponding OLS estimate, Cochrane's test also suffers from severe size distortions.

Keywords: Predictive regressions; Present-value relationship; Stock-return predictability.

JEL classification: C22, G1.

*We have benefitted from comments by Daniel Buncic, Paolo Giordani, Randi Hjalmarsson, Ron Kaniel, Jesper Lindé, Anders Wilhelmsson, Xin Zhang, and Pär Österholm, as well as by seminar participants at Sveriges Riksbank, Örebro University, and the European Summer Meeting of the Econometric Society, Cologne 2018. Hjalmarsson gratefully acknowledges financial support from the Nasdaq Nordic Foundation.

[†]Both authors are with the Department of Economics and the Centre for Finance, University of Gothenburg. Email: erik.hjalmarsson@economics.gu.se and tamas.kiss@cff.gu.se.

1 Introduction

A standard empirical framework for testing return predictability consists of a predictive regression for returns and an autoregressive (AR) process for the dividend–price ratio. In an influential study, Cochrane (2008) adds a predictive regression for the dividend-growth rate to this standard model, and links the three equations through the Campbell and Shiller (1988) present value identity. Cochrane’s key insight is that under the identity that links returns, prices, and dividends, a given degree of predictability in returns corresponds exactly to some degree of predictability in dividend growth. Thus, if one is testing the null hypothesis of no return predictability, using the dividend–price ratio as predictor, this null hypothesis has an exact translation in terms of predictability in dividend growth.¹

In this paper, we show that Cochrane’s dividend-growth based test is very similar to a test based on the full information maximum likelihood (ML) estimator for the standard bi-variate system, where the AR parameter of the dividend–price ratio is treated as known. Cochrane’s test can thus be viewed as an economically motivated (approximate) derivation of the ML estimator in the case of a known AR parameter. This finding explains why the dividend-growth based test, as implicitly formulated by Cochrane, appears more powerful than the return-based test using the simple OLS estimator. However, if one were to use the same assumptions when formulating the return-based test, one could use an ML procedure that (asymptotically) dominates the dividend-growth based test. That is, if one compares testing approaches based on the same information set, or the same set of assumptions (i.e., treating the AR parameter as known and equal to some given value),

¹Cochrane’s proposed modelling framework and testing approach has received great interest in the profession, and there is now a host of papers that evaluate both return predictability as well as dividend-growth predictability. Papers explicitly using Cochrane’s (2008) approach to test for return predictability include Chen (2009); Engsted and Pedersen (2010); Golez and Koudijs (2018). Other related works that consider both return and dividend-growth predictability include Lettau and Ludvigson (2005); Boudoukh et al. (2007); Lettau and Van Nieuwerburgh (2008); Binsbergen and Koijen (2010); Lacerda and Santa-Clara (2010); Lettau and Ludvigson (2010); Chen et al. (2012); Kelly and Pruitt (2013); Golez (2014); Bollerslev et al. (2015); Maio and Santa-Clara (2015); Detzel and Strauss (2016). Koijen and Van Nieuwerburgh (2011) provides a review on return and dividend-growth predictability.

there are no apparent gains from using the dividend-growth regression in testing for return predictability.²

The ML procedure is, in fact, extremely sensitive to the specific value used for the AR parameter of the dividend–price ratio, and by implication so is Cochrane’s procedure. Specifically, the choice to treat the OLS estimate of the AR parameter as the “true” value, which appears to have been adopted in subsequent empirical studies (e.g., Chen, 2009; Engsted and Pedersen, 2010; Golez and Koudijs, 2018), leads to severe size distortions. A test with a nominal size of five percent is shown to have actual rejection rates in excess of 20 percent under the null hypothesis. On the other hand, if the value of the AR parameter is set high enough, such that it is greater than or equal to the true parameter value in the data, the size of the resulting test can be controlled. For instance, under an assumption that the dividend–price ratio is stationary, setting the AR parameter equal to unity would ensure that it is greater than the true parameter value in the data. In this case, Cochrane’s test becomes similar to the conservative sup-bound test developed in Lewellen (2004) and analyzed further in Campbell and Yogo (2006).³

To form some intuition for our results, note that Cochrane’s predictive model is made up of three regressions, where the equations are linked together by the Campbell and Shiller (1988) present value identity, which implies an exact relationship among the slope coefficients as well as the error terms in the three regressions. Thus, any one of the three equations is redundant in the model formulation and the standard bi-variate predictive system, consisting of a predictive regression for returns and an AR process for the dividend–price ratio, must contain exactly the same information as the tri-variate system.

²Within the same framework, Cochrane also discusses tests of long-run predictability. We do not consider the properties of these tests here, as their formulation is based on a re-scaled version of the return coefficient, and does not explicitly rely on inference in the dividend-growth regression.

³We make no claim to provide an exhaustive analysis of the relative merits of the many inferential methods that exist for predictive regressions (recent examples include Chen and Deo, 2009, Phillips and Chen, 2014, and Kostakis et al., 2015). Rather, we focus solely on the properties of the dividend-growth based procedure proposed by Cochrane (2008), and how it relates to likelihood-based inference.

Our results essentially confirm this basic intuition: Adding a fully redundant equation to a regression system should not lead to any statistical gains. From the perspective of empirical research, our findings imply that if one is interested in testing for *return* predictability, there is no extra information available in the dividend-growth equation, and equally or more powerful tests can be formulated from the standard predictive regression setup.

An empirical application to aggregate U.S. stock returns illustrates our main theoretical results. Specifically, we show that the dividend-growth based test provides very similar results to an ML-based test, and that both tests depend strongly upon the assumption on the maximum feasible value for the AR parameter in the dividend-price ratio. If one is not willing to impose any stronger assumption than stationarity of the dividend-price ratio, both tests fail to reject the null hypothesis of no return predictability at the five percent significance level. On the other hand, if one is willing to assume that the AR parameter in the dividend-price ratio is below about 0.97 in annual data, the evidence would point in favor of return predictability. Seemingly small changes in the assumptions on the AR parameter can thus lead to rather drastic changes in inference.

2 Testing return predictability

2.1 Model formulation

Our predictive model is identical to the one used by Cochrane (2008). Let r_t denote the log-returns from period $t-1$ to t , d_t the time t log-dividends, and $d_t - p_t$ the corresponding log dividend-price ratio. The joint model of return and dividend-growth predictability is

formulated as the following restricted first-order VAR system,

$$r_t = \alpha_r + \beta_r (d_{t-1} - p_{t-1}) + \epsilon_t^r, \quad (\text{E1})$$

$$\Delta d_t = \alpha_d + \beta_d (d_{t-1} - p_{t-1}) + \epsilon_t^d, \quad (\text{E2})$$

$$d_t - p_t = \alpha_{dp} + \phi (d_{t-1} - p_{t-1}) + \epsilon_t^{dp}. \quad (\text{E3})$$

By Campbell and Shiller (1988), the following (approximate) present value identity holds,

$$r_t = \rho (p_t - d_t) + \Delta d_t - (p_{t-1} - d_{t-1}). \quad (1)$$

The identity is obtained through a log-linearization of returns around the long-run mean of the dividend–price ratio, denoted by ρ and empirically defined as

$$\rho = \frac{e^{-(\overline{d-p})}}{1 + e^{-(\overline{d-p})}}, \quad (2)$$

where $\overline{d-p}$ is the average dividend–price ratio. The parameter ρ is subsequently treated as a fixed and “known” quantity. In the CRSP data used by Cochrane, $\rho = 0.9638$, and this is the value that we use throughout this study as well.

The present value identity in equation (1) implies the following restrictions on the coefficients and error terms in the predictive equations (E1)–(E3),

$$\beta_r = \beta_d + (1 - \rho\phi), \quad (\text{R1})$$

and

$$\epsilon_t^r = \epsilon_t^d - \rho\epsilon_t^{dp}. \quad (\text{R2})$$

The restrictions in (R1) and (R2) imply that any one of the three model equations is redundant, and an equivalent model formulation would be retained by dropping any one

of the equations (E1)–(E3).⁴

Let $\epsilon_t = (\epsilon_t^r, \epsilon_t^d, \epsilon_t^{dp})'$ denote the vector of mean zero innovations, and let $\Sigma \equiv E[\epsilon_t \epsilon_t']$ be the covariance matrix for ϵ_t , where σ_{ij} , $i, j = 1, 2, 3$, denotes the elements of Σ . By restriction (R2), the covariance matrix Σ can be written as

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{11} + \rho\sigma_{13} & \sigma_{13} \\ \sigma_{11} + \rho\sigma_{13} & \sigma_{11} + \rho^2\sigma_{33} + 2\rho\sigma_{13} & \sigma_{13} + \rho\sigma_{33} \\ \sigma_{13} & \sigma_{13} + \rho\sigma_{33} & \sigma_{33} \end{bmatrix}. \quad (3)$$

It is easily seen that Σ is not full rank, reflecting the redundancy in the equation system (E1)–(E3).

2.2 Standard OLS-based inference

Cochrane's (2008) key idea is that the absence of return predictability must imply the presence of dividend-growth predictability. That is, the coefficient restriction,

$$\beta_r = \beta_d + (1 - \rho\phi), \quad (4)$$

implies that if $\beta_r = 0$,

$$\beta_d = -(1 - \rho\phi) < 0, \quad (5)$$

provided $\rho\phi < 1$. Thus, under the assumption of $\phi < 1$ (or at a minimum $\rho\phi < 1$), it follows that $\beta_r = 0$ implies $\beta_d = -(1 - \rho\phi) \neq 0$ and $\beta_d = 0$ implies $\beta_r = (1 - \rho\phi) \neq 0$.

Cochrane therefore suggests that instead of testing just the usual simple null, $\beta_r = 0$,

⁴The identity in equation (1) is only approximate, and the restrictions stated in equations (R1) and (R2) are therefore also approximate. However, as shown by Cochrane (2008), as well as in many subsequent papers (e.g., Binsbergen and Koijen, 2010; Koijen and Van Nieuwerburgh, 2011; Engsted et al., 2012; Kelly and Pruitt, 2013), the restrictions in (R1) and (R2) hold very closely empirically, and we will therefore treat them as exact throughout the paper. This is also in line with how Cochrane deals with them, and all his simulation results make explicit use of their exact identity.

one should also test whether $\beta_d = -(1 - \rho\phi)$. Thus, the “joint” null hypothesis can be formulated as

$$H_0 : \beta_r = 0 \quad \text{and} \quad \beta_d + (1 - \rho\phi) = 0. \quad (6)$$

Under the maintained model specification, the coefficient restriction $\beta_r = \beta_d + (1 - \rho\phi)$ is exact and the restriction also applies to the estimated values, such that

$$\hat{\beta}_{r,LS} = \hat{\beta}_{d,LS} + (1 - \rho\hat{\phi}_{LS}). \quad (7)$$

Since this equality holds numerically for the OLS estimates in any sample, the distribution of $\hat{\beta}_{r,LS}$ and $\hat{\beta}_{d,LS} + (1 - \rho\hat{\phi}_{LS})$ must also be identical (see Appendix A). Therefore, if one uses the OLS estimates of β_r, β_d , and ϕ to test either of the simple nulls, $\beta_r = 0$ or $\beta_d + (1 - \rho\phi) = 0$, the two tests using $\hat{\beta}_{r,LS}$ or $\hat{\beta}_{d,LS} + (1 - \rho\hat{\phi}_{LS})$ must have the same rejection regions for a given significance level. Appendix A illustrates this point, by showing that the individual standard t -statistics for testing $\beta_r = 0$ or $\beta_d + (1 - \rho\phi) = 0$ are numerically identical.

Thus, provided all three parameters, β_r, β_d , and ϕ are estimated (by OLS), there is no distinction between testing the joint null vis-à-vis the standard simple null of $\beta_r = 0$, and no power gains can therefore be achieved through such an approach.

2.3 Cochrane’s simulation approach

In contrast, Cochrane (2008) proposes a simulation-based approach—from which finite-sample distributions of the estimators are obtained—and reports substantially stronger evidence of return predictability when considering a test based on β_d rather than β_r . The simulations are intended to replicate the estimated model as closely as possible, while imposing the null of no return predictability.

Specifically, the simulated model is specified as follows. The AR parameter ϕ is set

equal to the OLS estimate from the data being used (below, we also discuss alternatives to this parameterization). Given this value of ϕ , β_d is set such that $\beta_r = \beta_d + (1 - \rho\phi) = 0$, thus imposing the null of no return predictability. That is, $\beta_d = -(1 - \rho\phi)$. The parameter ρ is set to 0.9638 and the covariance matrix Σ is also set equal to the empirical estimate from the data.

Samples from this model are simulated, and β_r and β_d are estimated in each of these samples. In particular, equations (E2) and (E3) are simulated, and the values in the return equation (E1) are inferred from the identity (1). The present value identity thus holds exactly in the simulations, and by implication the restrictions (R1) and (R2) also hold exactly.

In each draw i of the simulation, coefficient estimates $\hat{\beta}_{r,LS}^{i,sim}$, $\hat{\beta}_{d,LS}^{i,sim}$, and $\hat{\phi}_{LS}^{i,sim}$ are obtained through OLS estimation. For each simulated sample, these are related as

$$\hat{\beta}_{r,LS}^{i,sim} = \hat{\beta}_{d,LS}^{i,sim} + (1 - \rho\hat{\phi}_{LS}^{i,sim}). \quad (8)$$

Let $b_r^{Data} \equiv \hat{\beta}_{r,LS}^{Data}$ and $b_d^{Data} \equiv \hat{\beta}_{d,LS}^{Data}$ denote, respectively, the OLS coefficient estimates of β_r and β_d in the actual data. The simulations are then used to evaluate how rarely the events $\hat{\beta}_{r,LS}^{i,sim} \geq b_r^{Data}$ and $\hat{\beta}_{d,LS}^{i,sim} \geq b_d^{Data}$ occur. That is, treating b_r^{Data} and b_d^{Data} as given, the probabilities

$$\Pr\left(\hat{\beta}_{r,LS}^{i,sim} \geq b_r^{Data}\right) \equiv p_r, \quad (9)$$

and

$$\Pr\left(\hat{\beta}_{d,LS}^{i,sim} \geq b_d^{Data}\right) \equiv p_d, \quad (10)$$

are determined based on the empirical distributions of $\hat{\beta}_{r,LS}^{i,sim}$ and $\hat{\beta}_{d,LS}^{i,sim}$. These p-values capture the likelihood of observing the empirically estimated coefficients b_r^{Data} and b_d^{Data} ,

if the null of no return predictability was true.⁵ Cochrane finds that $p_d \approx 2\%$ whereas $p_r \approx 22\%$, which is interpreted as the dividend-growth test providing much stronger evidence against the null of no return predictability than the predictability test for the actual returns. That is, the b_d^{Data} outcome is highly unlikely to have been generated under the null model, whereas the b_r^{Data} outcome is not that unlikely. Based on these results, it is concluded that the test based on β_d is more powerful.⁶

Why are the two p-values, p_r and p_d different? According to the discussion in the previous sub-section, one would expect the tests of β_r and β_d to be identical. Consider first the case where the value of ϕ used to simulate the dividend-price ratio process is set to $\phi^{Data} \equiv \hat{\phi}_{LS}^{Data}$. This specification is the one that seems to have been adopted in subsequent empirical studies (Engsted and Pedersen, 2010; Golez and Koudijs, 2018), although Cochrane also considers other scenarios that we discuss further below. To see the implications of this formulation of the simulated model, write

$$b_d^{Data} = b_r^{Data} - (1 - \rho\phi^{Data}) = b_r^{Data} - \left(1 - \rho\hat{\phi}_{LS}^{i,sim}\right) + \rho\left(\phi^{Data} - \hat{\phi}_{LS}^{i,sim}\right), \quad (11)$$

where the first step uses the restriction on the coefficients from the data and the second step simply adds and subtracts $\rho\hat{\phi}_{LS}^{i,sim}$.⁷ Using the restriction on the estimated coefficients

⁵Cochrane also considers the empirical distribution of the t -ratios $\hat{\beta}_{r,LS}^{i,sim} / \sqrt{Var(\hat{\beta}_{r,LS}^{i,sim})}$ and $\hat{\beta}_d^{i,sim} / \sqrt{Var(\hat{\beta}_d^{i,sim})}$. The subsequent literature seems to have primarily adopted the tests based directly on the coefficients (e.g., Golez and Koudijs, 2018), and we similarly focus on these in our analysis. In non-reported Monte Carlo simulations, we find results for the t -ratio tests that are very similar to those we document for the coefficient tests in Section 3.2 below.

⁶The word “power” is used here in a somewhat imprecise sense. Formally, power is defined as the probability of rejecting the null under a given alternative. In Cochrane’s simulations, the rejection probabilities are all obtained under the null and are therefore not, in the true sense, a measure of power.

⁷The restriction on the coefficients (equation (7)) does not hold exactly for the OLS estimates in the actual data, but the discrepancy is empirically very small.

in the simulations stated in equation (8), it follows that

$$\begin{aligned}
& \Pr\left(\hat{\beta}_{d,LS}^{i,sim} \geq b_d^{Data}\right) \\
&= \Pr\left(\hat{\beta}_{r,LS}^{i,sim} - \left(1 - \rho\hat{\phi}_{LS}^{i,sim}\right) \geq b_r^{Data} - \left(1 - \rho\hat{\phi}_{LS}^{i,sim}\right) + \rho\left(\phi^{Data} - \hat{\phi}_{LS}^{i,sim}\right)\right) \\
&= \Pr\left(\hat{\beta}_{r,LS}^{i,sim} - \rho\left(\phi^{Data} - \hat{\phi}_{LS}^{i,sim}\right) \geq b_r^{Data}\right) \\
&\neq \Pr\left(\hat{\beta}_{r,LS}^{i,sim} \geq b_r^{Data}\right)
\end{aligned} \tag{12}$$

whenever $\hat{\phi}_{LS}^{i,sim} \neq \phi^{Data}$.

How does this simulation-based testing approach differ from the standard OLS inference discussed in the previous sub-section, where the β_r - and β_d -based tests of return predictability were shown to be identical? The key difference stems from the fact that if one wishes to use the estimate of β_d to test a null of $\beta_r = 0$, the relevant null hypothesis for β_d is in fact not fully known since $\beta_r = 0$ is equivalent to $\beta_d = -(1 - \rho\phi)$. In the simulation approach, the p-values for the β_d -based test reflect the sampling uncertainty in the estimates of β_d , but ignore the uncertainty coming from the fact that the value of the AR parameter in the original data is indeed unknown. In effect, the p-values correspond to a test of the null hypothesis $\beta_d = -(1 - \rho\phi^{Data})$, which postulates that the true AR parameter in the data is known and equal to ϕ^{Data} .⁸

Because of the downward bias in the OLS estimator of the AR coefficient ϕ , it follows that on average, $\phi^{Data} - \hat{\phi}_{LS}^{i,sim} > 0$, and it is therefore reasonable to assume that

$$\Pr\left(\hat{\beta}_{d,LS}^{i,sim} \geq b_d^{Data}\right) = \Pr\left(\hat{\beta}_{r,LS}^{i,sim} - \rho\left(\phi^{Data} - \hat{\phi}_{LS}^{i,sim}\right) \geq b_r^{Data}\right) \leq \Pr\left(\hat{\beta}_{r,LS}^{i,sim} \geq b_r^{Data}\right), \tag{13}$$

⁸The simulation based test can alternatively be interpreted as a parametric bootstrap approach (E1)–(E3), where the parameters are treated asymmetrically. In particular, β_r and β_d is estimated in each bootstrap sample, while ϕ is implicitly held constant at ϕ^{Data} by the definition of the null of the indirect test. A more standard way to perform the bootstrap would be to re-estimate all parameters in each bootstrap sample. In this case, the dividend-growth based test can be formulated as $\hat{\beta}_{d,LS}^{i,sim} + 1 - \rho\hat{\phi}_{LS}^{i,sim} = 0$, which is then equivalent to the direct test based on $\hat{\beta}_{r,LS}^{i,sim} = 0$ per the discussion in Section 2.2.

which goes some way towards explaining the results in Cochrane (2008). In fact, one can get a stronger result. The full information ML estimator of β_r with a known ϕ (see Campbell and Yogo, 2006) is given by

$$\hat{\beta}_{r,ML}(\phi) = \hat{\beta}_{r,LS} + \frac{\sigma_{13}}{\sigma_{33}} (\phi - \hat{\phi}_{LS}). \quad (14)$$

In Cochrane's data, $\sigma_{13}/\sigma_{33} \approx -0.9$ and $\hat{\beta}_{r,ML} \approx \hat{\beta}_{r,LS} - 0.9 (\phi - \hat{\phi}_{LS})$. With $\rho \approx 0.96$, it follows that

$$\hat{\beta}_{r,LS}^{i,sim} - \rho (\phi^{Data} - \hat{\phi}_{LS}^{i,sim}) \approx \hat{\beta}_{r,LS}^{i,sim} - 0.96 (\phi^{Data} - \hat{\phi}_{LS}^{i,sim}) \approx \hat{\beta}_{r,ML}^{i,sim}(\phi^{Data}), \quad (15)$$

where $\hat{\beta}_{r,ML}^{i,sim}(\phi^{Data})$ is calculated using the true autoregressive parameter, ϕ^{Data} , in the simulated model. Further, note that the ML estimator of β_r , with the least squares estimate of ϕ treated as the known true value, trivially reduces to the OLS estimator of β_r . That is, $b_r^{Data} = \hat{\beta}_{r,ML}^{Data}(\phi^{Data})$, where $\hat{\beta}_{r,ML}^{Data}(\phi^{Data})$ is the ML estimator of β_r in the original data, treating $\phi^{Data} = \hat{\phi}_{LS}^{Data}$ as the true value of ϕ . Defining $b_{r,ML}^{Data}(\phi^{Data}) \equiv \hat{\beta}_{r,ML}^{Data}(\phi^{Data}) = b_r^{Data}$, equations (13) and (15) thus give,

$$\begin{aligned} \Pr(\hat{\beta}_{d,LS}^{i,sim} \geq b_d^{Data}) &\approx \\ \Pr(\hat{\beta}_{r,ML}^{i,sim}(\phi^{Data}) \geq b_r^{Data}) &= \Pr(\hat{\beta}_{r,ML}^{i,sim}(\phi^{Data}) \geq b_{r,ML}^{Data}(\phi^{Data})). \end{aligned} \quad (16)$$

The test based on $\hat{\beta}_{d,LS}$ in this setup of Cochrane's simulation-based approach therefore turns out to be very similar to a test using the ML estimator of β_r , when ϕ is treated as fixed and known and equal to the OLS estimate in the data.

2.4 Altering the value of ϕ in the simulations

What happens if the value for ϕ used in the simulations is changed from $\phi^{Data} = \hat{\phi}_{LS}^{Data}$ to some other value? As shown in Appendix B, a similar result to that derived above holds, except the β_d -based test now corresponds to a test based on the ML estimator using this alternative value for the AR parameter as the true value of ϕ . In particular, if one sets the AR parameter in the simulations equal to some maximum feasible value for ϕ , say ϕ^{Max} , one ends up with a test that is similar to Lewellen's (2004) test, interpreted by Campbell and Yogo (2006) as a sup-bound test. Provided the assumption $\phi \leq \phi^{Max}$ indeed holds, the resulting test will generally be conservative, in the sense that if $\phi < \phi^{Max}$, the rejection rate under the null hypothesis will be smaller than the nominal significance level of the test. That is, analogous to the actual ML-based tests, setting the AR parameter large enough in Cochrane's simulation-based test is a way of constructing tests that do not over-reject the null.

2.5 Is the similarity with ML coincidental?

The near numerical equality with the ML estimator is seemingly somewhat accidental, and depends on the specific values of σ_{13} and σ_{33} . Should one in general expect the two to be close? Note that by the restrictions in the covariance matrix (3), $-\sigma_{13}/\sigma_{33} = \rho$ is equivalent to $\sigma_{23} = 0$. Therefore, the near-equivalence of the dividend-growth based test and the ML-based test is a result of the near zero correlation between the dividend-price ratio and dividend-growth innovations observed in the data used by Cochrane (2008).

Specifically, given $\sigma_{23} = 0$ ($-\sigma_{13}/\sigma_{33} = \rho$),

$$\begin{aligned}
\hat{\beta}_{r,ML}(\phi) &= \hat{\beta}_{r,LS} + \frac{\sigma_{13}}{\sigma_{33}}(\phi - \hat{\phi}_{LS}) \\
&= \hat{\beta}_{d,LS} + 1 - \rho\hat{\phi}_{LS} + \frac{\sigma_{13}}{\sigma_{33}}(\phi - \hat{\phi}_{LS}) \\
&= \hat{\beta}_{d,LS} + 1 - \rho\phi.
\end{aligned}$$

Treating the autoregressive parameter ϕ as fixed and known, and provided $\sigma_{23} = 0$, inference based on $\beta_{d,LS}^{i,sim}$ is thus equivalent to using $\hat{\beta}_{r,ML}(\phi)$.

As Cochrane (2008) and the subsequent literature—summarized in Kojien and Van Nieuwerburgh (2011)—suggest, $\sigma_{23} = 0$ is a relatively robust feature of the post-war U.S. equity data. However, this results is not general. In fact, the empirical literature has documented a relatively strong correlation (ranging between 0.3 and 0.5) between the dividend-growth and dividend-price ratio innovations in the pre-war U.S. data (Chen, 2009), and internationally (Engsted and Pedersen, 2010). In these contexts, σ_{23} is significantly different from zero, therefore the dividend-growth based test may differ substantially from the ML test. Since the latter makes asymptotically efficient use of the information on ϕ , using the dividend-growth based test in these cases will (asymptotically) result in a loss of power compared to the ML-based test.

3 Size of the test

3.1 Lessons from the ML estimator

The simulation-based test, using $\hat{\beta}_{d,LS}$, was shown to be similar to an ML-based test, where the value of the AR parameter ϕ specified in the simulation design is treated as the “true” AR parameter in the ML estimator. Cochrane’s dividend-growth based test therefore (approximately) inherits the properties of the ML estimator, for a given

specification of the AR parameter.

In case the AR parameter in the simulation design is set to the OLS estimate from the data, the result is a test that is severely over-sized. We illustrate this below by reporting actual rejection rates, under the null of no return predictability, in a Monte Carlo simulation. However, the general idea can quite easily be understood by considering the properties of the ML estimator. As remarked above, the ML estimator using $\hat{\phi}_{LS}$ instead of the true value ϕ , reduces to the OLS estimator. However, if one now proceeds *as if* $\hat{\phi}_{LS}$ was indeed the true known value for ϕ , the perceived (asymptotic) variance of the estimator would erroneously be calculated as (see Appendix C)

$$Var\left(\hat{\beta}_{r,ML}\left(\hat{\phi}_{LS}\right)\right) = (1 - \delta_{13}^2) Var\left(\hat{\beta}_{r,LS}\right) \leq Var\left(\hat{\beta}_{r,LS}\right), \quad (17)$$

where $\delta_{13} = \sigma_{13} / \sqrt{\sigma_{11}\sigma_{33}}$ is the correlation between the return and the dividend–price ratio innovations. Since $\hat{\beta}_{r,ML}\left(\hat{\phi}_{LS}\right) = \hat{\beta}_{r,LS}$, the variance of the ML estimator would therefore be severely under-estimated for δ_{13} close to unity. Resulting test statistics based on this (erroneous) result would be over-sized.^{9,10}

In general, one never has full knowledge of ϕ , but one might be willing to impose some upper limit on the range of possible values for ϕ . Most prominently, one might assume that the dividend–price ratio is a stationary process, such that $\phi < 1$. More generally, suppose one imposes the assumption that $\phi \leq \phi^{Max}$. In that case, provided $\sigma_{13} < 0$, one can form a “conservative” (downward biased) estimator of β_r ,

$$\hat{\beta}_{r,ML}\left(\phi^{Max}\right) = \hat{\beta}_{r,LS} + \frac{\sigma_{13}}{\sigma_{33}}\left(\phi^{Max} - \hat{\phi}_{LS}\right). \quad (18)$$

⁹The OLS estimator is also biased, which further invalidates inference.

¹⁰Amihud and Hurvich (2004) and Amihud et al. (2008, 2010) analyze how bias-corrected versions of the OLS estimator of ϕ can be used to achieve bias-corrections in the estimator of β_r . Campbell and Yogo (2006) use a bonferroni approach to obtain a feasible version of ML-based inference, although Phillips (2014) has subsequently leveled a critique against the reliability of this type of procedure.

This is essentially the approach taken by Lewellen (2004), although his motivation comes from the finite sample bias result for $\hat{\beta}_{r,LS}$, derived in Stambaugh (1999). Campbell and Yogo (2006) also discuss this type of conservative estimator and associated test statistics. Under the assumption that $\phi \leq \phi^{Max}$, and provided $\sigma_{13} < 0$, tests based on the resulting estimator are conservative (i.e., under-sized).

3.2 Monte Carlo simulations

The above results are for the actual ML procedure. Given the close similarity between the ML-based test and Cochrane's dividend-growth based test, similar results should also apply to the latter. To verify this, we perform a Monte Carlo simulation where the actual rejection rates under the null of no return predictability are obtained. To be clear, Cochrane's testing procedure is in itself a simulation-based procedure, and here we evaluate the finite sample properties of that procedure in a controlled Monte Carlo simulation. The details of the Monte Carlo simulation are given in Appendix D, but the basic setup is as follows.

The dividend-growth and dividend-price ratio equations ((E2) and (E3)) are simulated, and the return equation (E1) is inferred from the present-value identity in equation (1). The null hypothesis $\beta_r = 0$ is imposed and for a given value for ϕ , the dividend-growth coefficient β_d is implicitly determined by the parameter restriction in equation (R1). The innovations, ϵ_t , are drawn from an *iid* normal distribution and their covariance matrix Σ satisfies the restrictions stated in equation (3). Intercepts are fitted in all regressions, although these are not commented on below. All simulation results are based on 10,000 repetitions. The unknown AR parameter ϕ plays the key role in determining the sampling properties of the OLS estimators of the main parameters β_r and β_d in the model, therefore we show results for alternative values of ϕ . Other parameters are set to the empirical estimates presented in Table 2 in Cochrane (2008) and we use the same

sample size of $T = 78$ that was available in Cochrane’s original sample, in each simulation draw. As noted earlier, in the U.S. CRSP data set used by Cochrane, the parameter σ_{23} is close to zero and the approximation $\rho \approx -\sigma_{13}/\sigma_{33}$ therefore holds well (specifically, $\sigma_{13}/\sigma_{33} \approx -0.9$ and $\rho \approx 0.96$).

We fix the nominal size of Cochrane’s test procedure to five percent in a one-sided test against a positive alternative. That is, in each round of the simulations, we reject the null hypothesis of no return predictability for p-values less than or equal to 0.05. As a comparison to Cochrane’s tests, we also calculate rejection rates for the standard OLS t -test and the ML-based test described in Campbell and Yogo (2006), which takes the shape of an adjusted t -test (referred to as the Q-test in Campbell and Yogo’s notation). For these tests, the empirical rejection frequencies are calculated using standard critical values (i.e., reject for test statistics greater than 1.65).

The results from the Monte Carlo simulations are reported in Figure 1, with the size of the tests plotted as functions of the true parameter value ϕ in the data generating process. The dividend-growth based test, using the least squares estimate of ϕ as the “true” AR parameter, performs very poorly in terms of size. This is true for any underlying AR parameter ϕ , with rejection rates always in excess of 20 percent for a nominal-sized 5 percent test. The ML-based test implemented in the analogous manner—using the least squares estimate of ϕ as the “true” AR parameter—suffers from even larger size distortions than the dividend-growth based test, as also illustrated in Figure 1. This is to be expected, since the ML-based test makes even stronger use of the value for the AR parameter.¹¹

¹¹In addition to Cochrane’s dividend-growth based test and the ML-based test, we also obtained Monte Carlo rejection rates for Cochrane’s test based on the return coefficient (equation (9)) as well as the standard OLS t -test. Cochrane’s test based on the return coefficient was found not to be sensitive to the exact value of ϕ in the simulations, and somewhat under-sized when one parametrizes the simulation with the OLS estimate of the autoregressive parameter. The standard OLS t -test suffers from well-known size distortions, with rejection rates ranging from around 12 to 25 percent depending on the true value of ϕ for a nominal-sized 5 percent test.

Figure 1 also presents results for the conservative ML-based test, using equation (18), and the dividend-growth based test performed in an analogous conservative way, where the upper bound for the autoregressive parameter is set to $\phi^{Max} = 0.995$. In this conservative implementation, both tests exhibit rejection rates that are typically well below the nominal five percent significance level. As the true parameter ϕ gets closer to the specified upper bound ϕ^{Max} , the rejection rates approach five percent.

[Insert Figure 1 here]

3.3 Power

Given the large size distortions of the dividend-growth based test that uses the OLS estimate of ϕ as the “true” AR parameter, there is little use in analyzing the power properties of this specific implementation of Cochrane’s dividend-growth based test. The conservative test maintains size well. However, as argued above, this test is very similar to Lewellen’s (2004) test and Campbell and Yogo’s sup-bound test. The latter study provides extensive results on the power of this test, and there is little reason to report very similar results here.

It is clear that power gains can be achieved by using the dividend-growth equation, but only *viz-à-viz* standard OLS tests, not against ML-type tests that use information on the AR parameter. In fact, the ML-based tests are asymptotically the most powerful (see Campbell and Yogo, 2006), and the dividend-growth based tests can therefore at best (asymptotically) achieve the power of the ML tests. We verify that these results hold also for finite sample sizes in non-reported simulations, where we show that under the parameter estimates in Cochrane’s study, the power of the ML and dividend-growth based tests are very close. However, if one changes the parameters somewhat (e.g., setting $\delta_{13} = \sigma_{13} / \sqrt{\sigma_{11}\sigma_{33}} = -0.95$), such that the very close correspondence between the ML-based and the dividend-growth based tests is relaxed, the ML tests can achieve some

power advantages.

4 Empirical results

We use the annual value-weighted CRSP returns, including and excluding dividends, to calculate the dividend–price ratio and the dividend growth. The sample period is 1927 to 2016, and we use real returns calculated with inflation defined as the monthly change in the Consumer Price Index.¹²

Our focus is on Cochrane’s dividend-growth based test, and we compare it to the ML-based test. Specifically, we estimate equations (E1)-(E3), using the full 1927 to 2016 sample, and calculate the p-values for each test over a range of different values postulated for ϕ . That is, we do not use the OLS estimator of ϕ in creating either of the test statistics, since this test suffers large size distortions. Instead we calculate the range of p-values obtained for $\phi \in [0.95, 0.995]$.

Figure 2 illustrates the sensitivity of the test results with respect to the assumptions made on the AR coefficient ϕ . In particular, the figure shows the p-values of the ML-based and the dividend-growth based tests as a function of the maximum value specified for ϕ . As is seen, if one is willing to assume that $\phi \leq 0.97$, one starts to find significant results.

[Insert Figure 2 here]

The empirical results presented in Figure 2 are very much in line with many previous studies. Without imposing additional assumptions, we find that tests that are robust to the bias inherent in these types of predictive regressions are at best borderline significant. Figure 2 clearly illustrates how the empirical results can be viewed as conditional on one’s beliefs regarding the autocorrelation in the dividend–price ratio. If one is willing to make stronger assumptions than merely assuming that the dividend–price ratio is stationary,

¹²Using excess returns (over the 3-month Treasury Bill rate) or restricting the sample to end in 2004, as in the original study by Cochrane, leads to very similar results to those reported here.

more significant results are obtained. It is worth pointing out that the OLS estimate of ϕ is equal to 0.94, and using this value as the “true” AR value in the test procedures would thus lead to a strong rejection of the null hypothesis.

5 Conclusion

We provide a detailed analysis of the properties of the dividend-growth based test of return predictability in Cochrane (2008). We show that Cochrane’s test is similar to a full-information maximum likelihood test, using an explicit assumption on the degree of persistence in the predictor variable. Using this assumption gives both Cochrane’s test and the ML-based test additional power over the standard OLS-based test. Cochrane’s test can be viewed as an economically motivated proxy for the statistically motivated efficient ML method, and as such Cochrane’s test does not add power over and above the previously existing ML test. Importantly, we also show that unless one specifically imposes a conservative approach to the formulation of Cochrane’s procedure, the test will tend to over reject the null hypothesis of no return predictability. Our findings highlight that while extending the simple predictive regression to more elaborate present-value frameworks helps provide a deeper economic understanding of return predictability, it does not help escape the associated inferential issues.

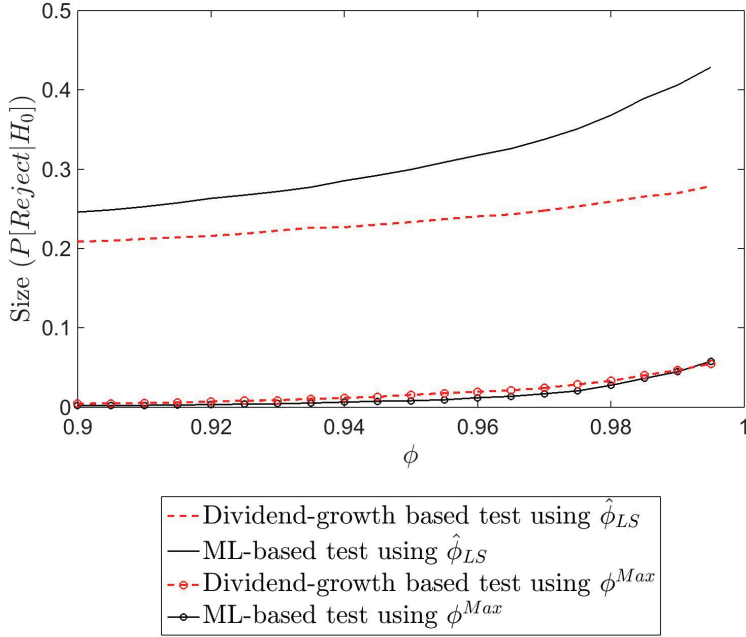
References

- Amihud, Y. and Hurvich, C. M. (2004). Predictive regressions: A reduced-bias estimation method. *Journal of Financial and Quantitative Analysis*, 39(04):813–841.
- Amihud, Y., Hurvich, C. M., and Wang, Y. (2008). Multiple-predictor regressions: Hypothesis testing. *Review of Financial Studies*, 22(1):413–434.
- Amihud, Y., Hurvich, C. M., and Wang, Y. (2010). Predictive regression with order-p autoregressive predictors. *Journal of Empirical Finance*, 17(3):513–525.
- Binsbergen, V., Jules, H., and Koijen, R. S. (2010). Predictive regressions: A present-value approach. *The Journal of Finance*, 65(4):1439–1471.
- Bollerslev, T., Xu, L., and Zhou, H. (2015). Stock return and cash flow predictability: The role of volatility risk. *Journal of Econometrics*, 187(2):458–471.
- Boudoukh, J., Michaely, R., Richardson, M., and Roberts, M. R. (2007). On the importance of measuring payout yield: Implications for empirical asset pricing. *The Journal of Finance*, 62(2):877–915.
- Campbell, J. Y. and Shiller, R. J. (1988). The dividend–price ratio and expectations of future dividends and discount factors. *Review of Financial Studies*, 1(3):195–228.
- Campbell, J. Y. and Yogo, M. (2006). Efficient tests of stock return predictability. *Journal of Financial Economics*, 81(1):27–60.
- Cavanagh, C. L., Elliott, G., and Stock, J. H. (1995). Inference in models with nearly integrated regressors. *Econometric Theory*, 11(05):1131–1147.
- Chen, L. (2009). On the reversal of return and dividend growth predictability: A tale of two periods. *Journal of Financial Economics*, 92(1):128–151.

- Chen, L., Da, Z., and Priestley, R. (2012). Dividend smoothing and predictability. *Management Science*, 58(10):1834–1853.
- Chen, W. W. and Deo, R. S. (2009). Bias reduction and likelihood-based almost exactly sized hypothesis testing in predictive regressions using the restricted likelihood. *Econometric Theory*, 25(5):1143–1179.
- Cochrane, J. H. (2008). The dog that did not bark: A defense of return predictability. *Review of Financial Studies*, 21(4):1533–1575.
- Detzel, A. L. and Strauss, J. (2016). The dog has barked for a long time: Dividend growth is predictable. Technical report.
- Engsted, T. and Pedersen, T. Q. (2010). The dividend–price ratio does predict dividend growth: International evidence. *Journal of Empirical Finance*, 17(4):585–605.
- Engsted, T., Pedersen, T. Q., and Tanggaard, C. (2012). The log-linear return approximation, bubbles, and predictability. *Journal of Financial and Quantitative Analysis*, 47(03):643–665.
- Golez, B. (2014). Expected returns and dividend growth rates implied by derivative markets. *Review of Financial Studies*, 27(3):790–822.
- Golez, B. and Koudijs, P. (2018). Four Centuries of Return Predictability. *Journal of Financial Economics*, 127(2):248–263.
- Kelly, B. and Pruitt, S. (2013). Market Expectations in the Cross-Section of Present Values. *Journal of Finance*, 68(5):1721–1756.
- Kendall, M. G. (1954). Note on bias in the estimation of autocorrelation. *Biometrika*, 41(3-4):403–404.

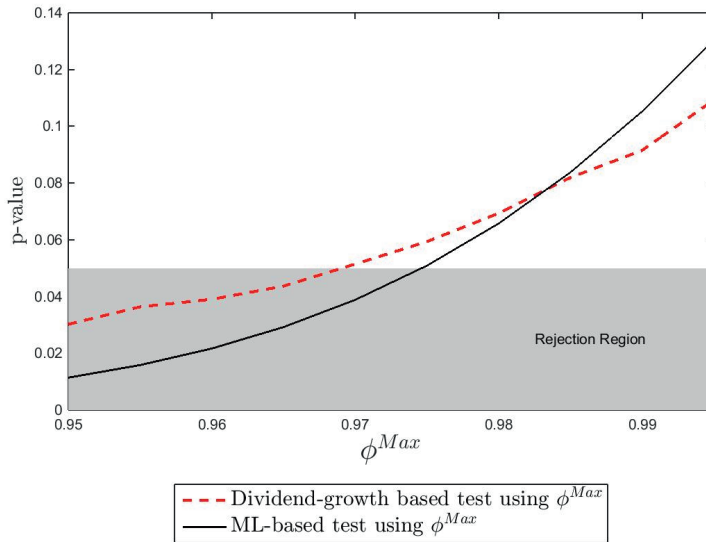
- Koijen, R. S. and Van Nieuwerburgh, S. (2011). Predictability of returns and cash flows. *Annual Review of Financial Economics*, 3:467–491.
- Kostakis, A., Magdalinos, T., and Stamatogiannis, M. P. (2015). Robust econometric inference for stock return predictability. *Review of Financial Studies*, 28(5):1506–1553.
- Lacerda, F. and Santa-Clara, P. (2010). Forecasting dividend growth to better predict returns. *Manuscript Universidade Nova de Lisboa*.
- Lettau, M. and Ludvigson, S. (2010). Measuring and Modelling Variation in the Risk-Return Trade-off. CEPR Discussion Papers 3105, C.E.P.R. Discussion Papers.
- Lettau, M. and Ludvigson, S. C. (2005). Expected returns and expected dividend growth. *Journal of Financial Economics*, 76(3):583–626.
- Lettau, M. and Van Nieuwerburgh, S. (2008). Reconciling the Return Predictability Evidence. *Review of Financial Studies*, 21(4):1607–1652.
- Lewellen, J. (2004). Predicting returns with financial ratios. *Journal of Financial Economics*, 74(2):209–235.
- Maio, P. and Santa-Clara, P. (2015). Dividend yields, dividend growth, and return predictability in the cross section of stocks. *Journal of Financial and Quantitative Analysis*, 50(1-2):33–60.
- Phillips, P. C. (2014). On confidence intervals for autoregressive roots and predictive regression. *Econometrica*, 82(3):1177–1195.
- Phillips, P. C. and Chen, Y. (2014). Restricted likelihood ratio tests in predictive regression. *Cowles Foundation Discussion Paper*, No. 1968.
- Stambaugh, R. F. (1999). Predictive regressions. *Journal of Financial Economics*, 54(3):375–421.

Figure 1: **Size distortions using $\hat{\phi}_{LS}$**



N Notes: The figure presents the actual size of one-sided nominal-sized five percent tests of the null hypothesis of $\beta_r = 0$ against a positive alternative. That is, the graphs indicate the average rejection rates, for the corresponding tests, under the null hypothesis of no return predictability. The dashed (solid) line is the dividend-growth (ML-) based test using the OLS estimate $\hat{\phi}_{LS}^{Data}$ as the value for the AR parameter. The dashed (solid) line with circles represents the rejection rates for the conservative dividend-growth (ML-) based test using $\phi^{Max} = 0.995$ as the value for the AR parameter. The ML-based tests use standard normal critical values (i.e., reject for test statistics greater than 1.65). The results are based on the Monte Carlo simulation described in the main text with 10,000 repetitions.

Figure 2: Empirical p-values as a function of ϕ^{Max}



Notes: The graph shows the p-values of tests of return predictability. The p-values are plotted as functions of the value specified for the upper bound, ϕ^{Max} , on the AR parameter. The solid line represents the ML-based test (using critical values from the standard normal distribution), and the dashed line represents the simulation-based test using β_d . Both tests specify that the value for the autoregressive parameter is set to the corresponding value for ϕ^{Max} .

Appendix

A Properties of the OLS estimators

For ease of notation, define $x_t = d_t - p_t$, and let X_{-1} denote the vector of stacked lagged observations for x_t . Similarly, denote r as the vector of observations on returns, and let ϵ^r , ϵ^d , and ϵ^{dp} , denote the stacked innovations. We treat the model without intercepts, but the results generalize immediately to regressions with fitted intercepts by replacing all variables by their demeaned versions.

The OLS estimator of $\beta = (\beta_r, \beta_d, \phi)'$ is now equal to

$$\hat{\beta}_{LS} = \begin{bmatrix} \hat{\beta}_{r,LS} \\ \hat{\beta}_{d,LS} \\ \hat{\phi}_{LS} \end{bmatrix} = \beta + (X'_{-1}X_{-1})^{-1} \begin{bmatrix} X'_{-1}\epsilon^r \\ X'_{-1}\epsilon^d \\ X'_{-1}\epsilon^{dp} \end{bmatrix}. \quad (19)$$

Using $\beta_r = \beta_d + (1 - \rho\phi)$ and $\epsilon^r = \epsilon^d - \rho\epsilon^{dp}$,

$$\begin{aligned} \hat{\beta}_{r,LS} &= \beta_r + (X'_{-1}X_{-1})^{-1} X'_{-1}\epsilon^r \\ &= \left(\beta_d + (X'_{-1}X_{-1})^{-1} X'_{-1}\epsilon^d \right) + \left(1 - \rho \left(\phi + (X'_{-1}X_{-1})^{-1} X'_{-1}\epsilon^{dp} \right) \right) \\ &= \hat{\beta}_{d,LS} + \left(1 - \rho\hat{\phi}_{LS} \right). \end{aligned} \quad (20)$$

Further, letting $M_{X_{-1}} = I - X_{-1}(X'_{-1}X_{-1})^{-1}X'_{-1}$, and using standard results, the fitted

residuals satisfy,

$$\begin{aligned}
\hat{\epsilon}^r &\equiv r - X_{-1}\hat{\beta}_{r,LS} \\
&= M_{X_{-1}}\epsilon^r \\
&= M_{X_{-1}}(\epsilon^d - \rho\epsilon^{dp}) \\
&= \hat{\epsilon}^d - \rho\hat{\epsilon}^{dp}.
\end{aligned} \tag{21}$$

In order to motivate the formulation of the standard t -statistics below, suppose that $\phi < 1$ and that ϵ_t is an *iid* or martingale difference sequence.¹³ Under classical asymptotic results, it follows straightforwardly that as the sample size $T \rightarrow \infty$,

$$\sqrt{T}(\hat{\beta}_{LS} - \beta) \xrightarrow{d} N(0, (\text{Var}(x_{t-1}^2))^{-1}\Sigma) \equiv N(0, \Omega), \tag{22}$$

where Σ is given in (3), and $\text{Var}(x_{t-1}^2) = \frac{\sigma_{33}}{1-\phi^2}$. For ϕ close to one, the asymptotic distribution result stated in (22) does not hold up well in finite samples. However, our main purpose for stating this result is to motivate the standard t -statistics considered below, and show that their standard formulation leads to an equivalence between the return based and the dividend-growth based tests, analogous to that for the actual coefficients seen in equation (20) above.

Let $\hat{\epsilon} = (\hat{\epsilon}^r, \hat{\epsilon}^d, \hat{\epsilon}^{dp})$ be the matrix of fitted residuals. A standard estimator of the asymptotic covariance matrix in (22) is given by

$$\hat{\Omega} = \left(\frac{1}{T}\hat{\epsilon}'\hat{\epsilon}\right) \left(\frac{1}{T}X'_{-1}X_{-1}\right)^{-1} = \begin{bmatrix} \hat{\epsilon}'_r\hat{\epsilon}_r & \hat{\epsilon}'_r\hat{\epsilon}_d & \hat{\epsilon}'_r\hat{\epsilon}_{dp} \\ \hat{\epsilon}'_r\hat{\epsilon}_d & \hat{\epsilon}'_d\hat{\epsilon}_d & \hat{\epsilon}'_d\hat{\epsilon}_{dp} \\ \hat{\epsilon}'_r\hat{\epsilon}_{dp} & \hat{\epsilon}'_d\hat{\epsilon}_{dp} & \hat{\epsilon}'_{dp}\hat{\epsilon}_{dp} \end{bmatrix} \left(\sum_{t=2}^T x_{t-1}^2\right)^{-1}, \tag{23}$$

¹³The stationarity condition ($\phi < 1$) is by no means necessary for our main analysis, but it enables us to frame the main properties of the OLS estimator of equations (E1)-(E3) in terms of classical asymptotic results for stationary models.

where the first diagonal element of $\hat{\Omega}$ corresponds to the asymptotic variance of $\hat{\beta}_r$. The standard t -statistic for testing the null hypothesis of $\beta_r = 0$ is thus given by,

$$t_{\beta_r} = \frac{\hat{\beta}_{r,LS}}{\sqrt{\frac{1}{T} (\hat{\epsilon}'\hat{\epsilon}^r) (X'_{-1}X_{-1})^{-1}}} = \frac{\hat{\beta}_{r,LS}}{\sqrt{\frac{1}{T} (\hat{\epsilon}^{d'}\hat{\epsilon}^d + \rho^2\hat{\epsilon}^{dp'}\hat{\epsilon}^{dp} - 2\rho\hat{\epsilon}^{d'}\hat{\epsilon}^{dp}) (X'_{-1}X_{-1})^{-1}}}, \quad (24)$$

where the last equality follows from the result in (21). To formulate the dividend-growth based t -statistic of the null of no return predictability, let $c = (0, 1, -\rho)$. The t -statistic is then given by

$$\begin{aligned} t_{\beta_{d+1-\rho\phi}} &= \frac{\hat{\beta}_{d,LS} + (1 - \rho\hat{\phi}_{LS})}{\sqrt{\frac{1}{T} (c' (\hat{\epsilon}'\hat{\epsilon}) c) (X'_{-1}X_{-1})^{-1}}} \\ &= \frac{\hat{\beta}_{r,LS}}{\sqrt{\frac{1}{T} (\hat{\epsilon}^{d'}\hat{\epsilon}^d + \rho^2\hat{\epsilon}^{dp'}\hat{\epsilon}^{dp} - 2\rho\hat{\epsilon}^{d'}\hat{\epsilon}^{dp}) (X'_{-1}X_{-1})^{-1}}} = t_{\beta_r}. \end{aligned} \quad (25)$$

The standard t -statistics, t_{β_r} and $t_{\beta_{d+1-\rho\phi}}$, are thus numerically identical, provided the restrictions in (R1) and (R2) hold.

B The simulation-based test parametrized with ϕ^{Max}

Before analyzing the simulation-based test, recall first the ‘‘conservative’’ ML estimator in equation (18),

$$\hat{\beta}_{r,ML}(\phi^{Max}) = \hat{\beta}_{r,LS} + \frac{\sigma_{13}}{\sigma_{33}} (\phi^{Max} - \hat{\phi}_{LS}). \quad (26)$$

In a conservative test, the value of $\hat{\beta}_{r,ML}(\phi^{Max})$ is evaluated against the critical value that would apply if the AR parameter in the data was indeed equal to ϕ^{Max} . Provided the true value of the AR parameter in the data is less than or equal to ϕ^{Max} , and $\sigma_{13} < 0$, the resulting estimator will be downward biased, and tests based on this estimator will be conservative against a positive alternative.

Consider now the simulation-based test, where the simulated model is parametrized with an AR parameter ϕ^{Max} .¹⁴ Let $\hat{\beta}_{r,ML}^{i,sim}(\phi^{Max})$ be the ML estimator of β_r in the simulations, and let $b_{r,ML}^{Data}(\phi^{Max}) \equiv \hat{\beta}_{r,ML}^{Data}(\phi^{Max})$ be the conservative ML estimate of β_r in the actual data. In the simulated data $\hat{\beta}_{r,ML}^{i,sim}(\phi^{Max})$ is the “correct” ML estimator, since the simulated model has an AR parameter equal to ϕ^{Max} , whereas in the original data the estimator is conservative (provided the true AR parameter is less than or equal to ϕ^{Max}). In order to show that the simulation-based dividend-growth test is similar to the conservative ML test in this case, we need to show that

$$\Pr\left(\hat{\beta}_{d,LS}^{i,sim} \geq b_d^{Data}\right) \approx \Pr\left(\hat{\beta}_{r,ML}^{i,sim}(\phi^{Max}) \geq b_{r,ML}^{Data}(\phi^{Max})\right). \quad (27)$$

That is, $\Pr\left(\hat{\beta}_{r,ML}^{i,sim}(\phi^{Max}) \geq b_{r,ML}^{Data}(\phi^{Max})\right)$ compares the conservative ML estimate from the data, $b_{r,ML}^{Data}(\phi^{Max})$, to the distribution of the ML estimator in a setting where ϕ^{Max} is indeed the true value for the AR parameter.

For $\rho \approx -\sigma_{13}/\sigma_{33}$, the ML estimator of β_r in the simulations can be written as

$$\hat{\beta}_{r,ML}^{i,sim}(\phi^{Max}) = \hat{\beta}_{r,LS}^{i,sim} + \frac{\sigma_{13}}{\sigma_{33}}\left(\phi^{Max} - \hat{\phi}_{LS}^{i,sim}\right) \approx \hat{\beta}_{r,LS}^{i,sim} - \rho\left(\phi^{Max} - \hat{\phi}_{LS}^{i,sim}\right). \quad (28)$$

As in Section 2.3, let $b_r^{Data} \equiv \hat{\beta}_{r,LS}^{Data}$ and $b_d^{Data} \equiv \hat{\beta}_{d,LS}^{Data}$. By restriction (R1),

$$b_d^{Data} = b_r^{Data} - \left(1 - \rho\hat{\phi}_{LS}^{Data}\right) = b_r^{Data} - \left(1 - \rho\hat{\phi}_{LS}^{i,sim}\right) + \rho\left(\hat{\phi}_{LS}^{Data} - \hat{\phi}_{LS}^{i,sim}\right), \quad (29)$$

and $\hat{\beta}_{d,LS}^{i,sim} = \hat{\beta}_{r,LS}^{i,sim} - \left(1 - \rho\hat{\phi}_{LS}^{i,sim}\right)$. The conservative ML estimator in the actual data, using $\rho \approx -\sigma_{13}/\sigma_{33}$, can be written as

$$b_{r,ML}^{Data}(\phi^{Max}) \approx b_r^{Data} - \rho\left(\phi^{Max} - \hat{\phi}_{LS}^{Data}\right). \quad (30)$$

¹⁴This formulation is without loss of generality, as the actual derivations make no use of the assumption that ϕ^{Max} is indeed larger than the true value for the AR parameter in the data.

It follows that

$$\begin{aligned}
& \Pr\left(\hat{\beta}_{d,LS}^{i,sim} \geq b_d^{Data}\right) \\
&= \Pr\left(\hat{\beta}_{r,LS}^{i,sim} - \left(1 - \rho\hat{\phi}_{LS}^{i,sim}\right) \geq b_r^{Data} - \left(1 - \rho\hat{\phi}_{LS}^{i,sim}\right) + \rho\left(\hat{\phi}_{LS}^{Data} - \hat{\phi}_{LS}^{i,sim}\right)\right) \\
&= \Pr\left(\hat{\beta}_{r,LS}^{i,sim} \geq b_r^{Data} + \rho\left(\hat{\phi}_{LS}^{Data} - \hat{\phi}_{LS}^{i,sim}\right)\right) \\
&= \Pr\left(\hat{\beta}_{r,LS}^{i,sim} - \rho\left(\phi^{Max} - \hat{\phi}_{LS}^{i,sim}\right) \geq b_r^{Data} - \rho\left(\phi^{Max} - \hat{\phi}_{LS}^{Data}\right)\right) \\
&\approx \Pr\left(\hat{\beta}_{r,ML}^{i,sim}\left(\phi^{Max}\right) \geq b_{r,ML}^{Data}\left(\phi^{Max}\right)\right). \tag{31}
\end{aligned}$$

C Variance of the estimators

Using the joint convergence for $\hat{\beta}_{LS} = \left(\hat{\beta}_{r,LS}, \hat{\beta}_{d,LS}, \hat{\phi}_{LS}\right)$ in equation (22), along with the shape of Σ given in (3), the asymptotic variance of the ML estimator of β_r can be expressed as follows,

$$\begin{aligned}
Var\left(\hat{\beta}_{r,ML}\right) &= Var\left(\hat{\beta}_{r,LS} + \frac{\sigma_{13}}{\sigma_{33}}\left(\phi - \hat{\phi}_{LS}\right)\right) \\
&= Var\left(\hat{\beta}_{r,LS}\right) + \frac{\sigma_{13}^2}{\sigma_{33}^2}Var\left(\hat{\phi}_{LS}\right) - 2\frac{\sigma_{13}}{\sigma_{33}}Cov\left(\hat{\beta}_{r,LS}, \hat{\phi}_{LS}\right) \\
&= Var\left(\hat{\beta}_{r,LS}\right) + \frac{\sigma_{13}^2}{\sigma_{33}^2}\frac{\sigma_{33}}{\sigma_{11}}Var\left(\hat{\beta}_{r,LS}\right) - 2\frac{\sigma_{13}}{\sigma_{33}}\frac{\sigma_{13}}{\sigma_{11}}Var\left(\hat{\beta}_{r,LS}\right) \\
&= Var\left(\hat{\beta}_{r,LS}\right)\left(1 - \frac{\sigma_{13}^2}{\sigma_{33}\sigma_{11}}\right) \\
&= Var\left(\hat{\beta}_{r,LS}\right)\left(1 - \delta_{13}^2\right). \tag{32}
\end{aligned}$$

D Implementation of Monte Carlo simulations

In order to implement Cochrane’s procedure in a repeated Monte Carlo simulation, a “two-layered” simulation is implemented. For a given set of true parameter values, the procedure can most easily be summarized by the below steps. For a simulation with

10,000 repetitions, steps 1-4 are repeated 10,000 times.

1. Simulate a single sample of size $T = 78$ of the system in equations (E1)-(E3), under restrictions (R1) and (R2).
2. From the simulated sample, obtain OLS estimates of all regression coefficients, as well as estimates of all parameters in the covariance matrix Σ . Denote the OLS estimates of β_r , β_d , and ϕ as $\hat{\beta}_{r,LS}^{sim}$, $\hat{\beta}_{d,LS}^{sim}$, and $\hat{\phi}_{LS}^{sim}$, respectively.
3. Parametrize the same system as in step (1), imposing the null of no return predictability and replacing all other parameters of the model with the *empirical estimates* from step (2). That is, in this parametrization, $\beta_r = 0$, $\beta_d = -\left(1 - \rho\hat{\phi}_{LS}^{sim}\right)$, and all other parameters are set equal to the empirical estimates obtained in step (2). Simulate 50,000 sample paths of size $T = 78$ from this empirically parametrized system that imposes the null of no return predictability. For each of the simulated samples, obtain OLS estimates of β_r and β_d .
4. Based on the resulting empirical distributions of the OLS estimators in step (3), calculate the p-values for the coefficient estimates $\hat{\beta}_{r,LS}^{sim}$ and $\hat{\beta}_{d,LS}^{sim}$, obtained from the initial simulated sample in steps (1) and (2). Classify as rejection/non-rejection depending on whether the respective p-value is less or greater than 0.05.

The above steps refer to the implementation of Cochrane's procedure where the parametrization in step (3) uses the OLS estimate of ϕ obtained in step (2). Alternative implementations differ only in their treatment of ϕ in step (3). Specifically, we also consider a conservative test where ϕ is set equal $\phi^{Max} = 0.995$. Thus, in this alternative implementation, the value of ϕ used to parametrize the simulated system in step (3) is set to ϕ^{Max} and the value of β_d is adjusted accordingly to ensure that the null of no return predictability is imposed (i.e., $\beta_d = -\left(1 - \rho\phi^{Max}\right)$).

Chapter III

1 Introduction

In this paper, I propose that the high persistence of the predictor and the parameter instability of predictive regressions are two interconnected phenomena. Analogous to the work on noisy predictors (Torous et al., 2004), I assume that the information in the predictor, captured by a stationary process, is confounded by an uninformative, non-stationary component. If the variability in the informative part of the predictor is large enough, then it is possible to detect the predictive relationship in small samples. However, as the number of observations increases, the non-stationary component inevitably becomes dominant in the regressor, and therefore, the estimated slope coefficient of the predictive regression converges to zero. This implies that the power of the predictive regression decreases over time, even if there is an underlying relationship between the predictor information and the returns. That is, predictability asymptotically vanishes under these assumptions.

Since several prominent predictors of excess stock market returns have serial correlation close to unity, the literature on inference and estimation based on such variables is abundant. In particular, the persistent regressor bias has been extensively discussed.¹ In these studies, predictability implies that expected returns and the predictor have common time series characteristics. In particular, if the predictor has high serial correlation, then the expected returns must share this property. To reconcile this feature with the stylized fact that realized returns are nearly serially uncorrelated, the usual approach is to assume that the persistent expected return component is small, and the unexpected returns dominate (c.f., Moon et al., 2005; Moon and Velasco, 2014).

In contrast, I propose that the predictor information need not be highly persistent; that is, expected returns are stationary and remain potentially predictable by a non-stationary variable. This weakens the predictability evidence in larger samples, which is in line with two related forms of parameter instability. First, the evidence of predictability is usually stronger in sub-samples than in century-long datasets (Ang and Bekaert, 2007;

¹Goetzmann and Jorion (1993); Cavanagh et al. (1995); Nelson and Kim (1993); Stambaugh (1999); Valkanov (2003); Lewellen (2004); Campbell and Yogo (2006); Ang and Bekaert (2007); Cochrane (2008); Boudoukh et al. (2008)

Lettau and Van Nieuwerburgh, 2008; Koijen and Van Nieuwerburgh, 2011). Second, predictors appear to lose power because the evidence of predictability weakens over time (Ferson et al., 2003; Goyal and Welch, 2008; Deng, 2013). In the current study, these forms of parameter instability are not surprising. In fact, based on the model I propose, if highly persistent (non-stationary) predictors are used, then (i) predictability is stronger in smaller samples, and (ii) the loss in predictive power is due to the increase in the sample size. This conclusion is parallel to the discussion in Osterrieder et al. (2015) who discuss the properties of the predictive equation with fractionally integrated regressors.

I derive the asymptotic results for the ordinary least squares (OLS) estimator to show that predictability disappears under a general set of assumptions on the dynamics of the time-varying expected returns. Finite sample properties indicate that the high persistence of the regressor results in a substantially biased slope coefficient for relevant sample sizes. The bias is especially large if the persistence of the regressor is so high that the variable is indistinguishable from a non-stationary process.

To mitigate the bias caused by the decreasing predictive power, I propose a simple and flexible estimation framework, subsample fixed effects (SFE). It builds on the idea that the bias increases with the sample size because the non-stationary component becomes dominant in larger samples. The problem can therefore be reduced by dividing the full sample and pooling the information from different subsamples via a fixed effects estimator. By limiting the subsample size, one effectively puts a bound on the variance accumulation within the regressor. Therefore, the extent of the bias in the estimator decreases and the estimated slope coefficient no longer vanishes asymptotically. I derive the exact relationship between the bias and the subsample size under the assumption of independent, identically distributed (i.i.d.) innovations. Simulations show that the proposed subsampling estimator is robust, since it reduces the bias caused by the non-stationary component, even under a more general set of assumptions.

Since the non-stationary component is more dominant in larger samples, the bias in the subsampling estimator is positively related to the size of a given subsample. Choosing a smaller subsample is thus more favorable to reduce bias. However, including more

fixed effects results in a loss of estimation precision. This translates into an efficiency-bias trade-off for the choice of subsample size in the SFE estimator. The simulation results suggest that the optimal choice of subsample size depends on the (potentially unobservable) parameters of the data generating process. Therefore, one can ensure robustness empirically by considering several subsample sizes simultaneously.

To test the proposed model empirically, I investigate the predictors of the excess returns on the S&P 500 stock market index. I focus on highly persistent variables as in this case, the model with a non-stationary component in the predictor can potentially be a good approximation. Looking at how regression estimates change over time, predictors of excess returns (including the dividend-price ratio, treasury bill rate, and book-to-market value) appear to exhibit vanishing predictability. Their slope coefficients approach zero as the sample size grows. Applying SFE shows an overall increase in the significance of these predictors. The estimated slope coefficients are the smallest in magnitude for the OLS (no subsampling) and they grow as one introduces subsampling and moves towards smaller subsamples. All these empirical observations support the predictions of the proposed model.

The rest of the paper is organized as follows. Section 2 describes the modelling framework and presents the main theoretical results, along with the proposed estimator, SFE. Section 4 presents the Monte Carlo simulations to analyze the performance of the model. Section 5 discusses subsample size selection. I provide the empirical results based on several important predictors of excess returns in Section 6, and Section 7 concludes. Technical derivations are presented in the Appendix.

2 The model

Consider returns that are stationary and potentially predictable. However, the explanatory variable has two components: the informative component (stationary) and a unit-root component (non-stationary) unrelated to the dependent variable. The data generating process can then be written in the following form:

$$y_t = \alpha_0 + \beta_0 \eta_{t-1} + u_t \quad (1)$$

$$x_t = \eta_t + \xi_t \quad (2)$$

$$\xi_t = \xi_{t-1} + \varepsilon_t \quad (3)$$

The stationary parts of the dependent variable, η_t and u_t , and the innovation of the unit root process, ε_t , are assumed to be linear processes with zero unconditional mean. Specifically, define $\mathbf{w}_t = (\eta_t, u_t, \varepsilon_t)'$. Then $\mathbf{w}_t = \sum_{j=0}^{\infty} C_j \zeta_{t-j}$, where $\{C_j\}_{j=1}^{\infty}$ is a sequence of matrices and ζ_{t-j} is a martingale difference sequence with $E(\zeta_t) = 0$, $E(\zeta_t \zeta_t') = \Sigma_{\zeta} \in \mathbb{R}_+^{3 \times 3}$. Furthermore, $E(\mathbf{w}_t \mathbf{w}_t') = \Sigma < \infty$ with $\text{diag}(\Sigma) = (\sigma_{\eta}^2, \sigma_u^2, \sigma_{\varepsilon}^2)$.

The data generating process is fairly standard in the return predictability literature with noisy predictors. The non-standard element is the fact that the noise component (the component unrelated to the predictive signal) is a unit-root process. Although I do not define it in more detail, the unit-root component can be thought of as capturing the (persistent) structural changes in the economy that affect the level of the predictor, but not the predictive relationship. Lettau and Van Nieuwerburgh (2008) proposes one such framework, in which changes in the long term mean of the dividend-price ratio create instability in the predictive relationship. The authors argue that these jumps in the mean reflect changes in the structure of the economy, going from one steady-state to another. In this context, the unit-root component in the present model is an alternative interpretation of these (gradual) shifts between steady states.

The key parameter of interest when assessing predictability is the slope coefficient in equation (1), β_0 . In most empirical work, the OLS estimator is applied to calculate the parameter estimates of the standard predictive regression,

$$y_t = \alpha + \beta x_{t-1} + e_t. \quad (4)$$

Under the assumptions in (1) – (3), the predictive power of x_t is masked by the non-stationary component. Since the left hand side of equation (4) is stationary in this case,

while the right hand side has a unit root, the slope coefficient and the predictive power disappear at the limit (converge to zero) by construction. The following proposition states this result:

Proposition 1. *Let the data generating process be described by equation (1) – (3). Furthermore let $\mathbf{w}_t = (\eta_t, u_t, \varepsilon_t) = \sum_{j=0}^{\infty} C_j \zeta_{t-j}$, where $\{C_j\}_{j=1}^{\infty}$ is a sequence of matrices and ζ_{t-j} is a martingale difference sequence with $E(\zeta_t) = 0$, $E(\zeta_t \zeta_t') = \Sigma_{\zeta} \in \mathbb{R}_+^{3 \times 3}$, and $E(\mathbf{w}_t \mathbf{w}_t') = \Sigma < \infty$.*

Then $\hat{\beta}_{OLS} \xrightarrow{p} 0$, where $\hat{\beta}_{OLS}$ is the OLS estimate of the slope coefficient of the regression (4). If $\alpha_0 = 0$ is also imposed, then $\hat{\alpha}_{OLS} \xrightarrow{p} 0$, otherwise it converges to a random variable.

Proof. In the Appendix. □

Intuitively, the result is straightforward. If a stationary variable is regressed on a non-stationary predictor, the variation in x_t becomes arbitrarily large as the sample size grows, and the only way to reconcile it with the finite variation in the stationary dependent variable is that the slope coefficient converges to zero.

Since no specific assumptions are made about the autocovariance structure of the y_t and η_t process, this argument is widely applicable. From a practical perspective, the advantage of this generality is that it allows one to specify η_t as a weakly dependent series, such as a stationary autoregressive–moving-average (ARMA) process. Proposition 1 also covers the case of endogenous regressors, as the error term of the dependent variable and either component of the regressor are allowed to be correlated.

If the predictors are tested for a unit root in a given sample, the test statistics often suggest a narrow rejection of non-stationarity. This can happen if the unit-root component is small relative to the informative part. In this case, statistical testing concludes that there is no unit root in the series. Therefore, running the predictive regression (4) seems to be appropriate and it may produce a good estimate for the slope coefficient and predictive power (c.f. Moon et al. (2005)). However, Proposition 1 suggests that increasing the sample size leads to more biased point estimates that eventually converge to zero. This

is in sharp contrast with the general notion that a growing sample size results in a better estimate of the slope coefficient.

3 Local demeaning and subsample fixed effects

Removing the local (subsample) mean from the regressor can mitigate the effect of the non-stationary component. I can show this by fixing a subsample size, denoted by $M \in \mathbb{N}$, and assuming that the entire sample size can be written as $T = KM$, where $K \in \mathbb{N}$. The processes x_t and y_t can then be written as sequences of processes $\{x_{k,t}\}_{k=1}^K$ and $\{y_{k,t}\}_{k=1}^K$, where $x_{k,t} = x_{(k-1)M+t}$ and $y_{k,t} = y_{(k-1)M+t}$. Define the locally demeaned regressor as

$$\tilde{x}_{k,t} = x_{k,t} - \frac{1}{M} \sum_{m=1}^M x_{k,m} \quad (5)$$

for all k and t . If $\{x_{k,t}\}_{k=1}^K$ is generated according to equations (2)-(3), where independence and identical distribution is imposed on the error terms, one can calculate the variance of the demeaned regressor.

Lemma 1. *Let x_t be generated by equations (2)-(3), and ε_t and η_t be i.i.d. random variables with variances $\sigma_\varepsilon^2, \sigma_\eta^2 < \infty$, respectively. Then*

$$\text{Var}(\tilde{x}_{k,t}) = A(M)\sigma_\eta^2 + B(M,t)\sigma_\varepsilon^2,$$

where the expressions for $A(M)$ and $B(M,t)$ are given in the proof.

Proof. In the Appendix. □

The key observation of Lemma 1 is that the variance of the demeaned regressor grows linearly in the subsample size. Based on this result, it is possible to characterize the properties of the least squares estimator using the locally demeaned explanatory variable (hereinafter, the “subsample fixed effects” estimator).² To obtain the exact asymptotic

²The notation comes from the observation that the estimator can be computed simply by including

results the regression error u_t is assumed to be independent of the other error variables. Then, the asymptotic results for the locally demeaned regressor are characterized by the following proposition.

Proposition 2. *Let the data generating process be described by equations (1-3) with $\{\varepsilon_t, \eta_t, u_t\}_{t=0}^\infty$ i.i.d. sequences with unconditional variances $\sigma_\varepsilon^2, \sigma_\eta^2, \sigma_u^2 < \infty$, respectively. Moreover, fix the subsample size $M \in \mathbb{N}$ such that $M = \frac{T}{K}$, where T is the total number of observations and K is the number of subsamples. Define $\hat{\alpha}_{SFE}$ and $\hat{\beta}_{SFE}$ as the OLS estimator of the coefficients in regression*

$$y_t = \alpha + \beta \tilde{x}_{t-1} + e_t, \quad (6)$$

where \tilde{x}_t is the locally demeaned regressor given by equation (5) using $\tilde{x}_{(k-1)M+t} = \tilde{x}_{k,t}$. Then, as $T \rightarrow \infty$ (and therefore $K \rightarrow \infty$ as M is fixed)

$$\begin{aligned} \hat{\beta}_{SFE} &\xrightarrow{p} \beta_0 \frac{\sigma_\eta^2}{\sigma_\eta^2 + \frac{M+1}{6}\sigma_\varepsilon^2} \\ \hat{\alpha}_{SFE} &\xrightarrow{p} E(y_t) = \alpha_0 \end{aligned}$$

Proof. In the Appendix. □

The formulation with fixed effects is especially useful since its estimation is straightforward. The result for the slope coefficient is similar to the classical measurement error attenuation bias formula, as the bias enters the estimate as a multiplicative factor. However, for the subsampling fixed effects estimator, the extent of the bias depends on the number of observations in each subsample, M . Since their relationship is positive, a larger M implies a more biased estimation.³ On the other hand, the variance of the estimator a fixed effect for each subsample, and using an OLS estimation. To see the equivalence note that if $\{y_{k,t}\}_{k=1}^K$ is generated according to equation (5), then

$$\frac{\sum_{k=1}^K \sum_{t=1}^M \tilde{y}_{k,t} \tilde{x}_{k,t-1}}{\sum_{k=1}^K \sum_{t=1}^M \tilde{x}_{k,t-1}^2} = \frac{\sum_{k=1}^K \sum_{t=1}^M y_{k,t} \tilde{x}_{k,t-1}}{\sum_{k=1}^K \sum_{t=1}^M \tilde{x}_{k,t-1}^2} - \frac{\sum_{k=1}^K \bar{y}_{k,t} \sum_{t=1}^M \tilde{x}_{k,t-1}}{\sum_{k=1}^K \sum_{t=1}^M \tilde{x}_{k,t-1}^2} = \frac{\sum_{k=1}^K \sum_{t=1}^M y_{k,t} \tilde{x}_{k,t-1}}{\sum_{k=1}^K \sum_{t=1}^M \tilde{x}_{k,t-1}^2},$$

where the last equality follows from the fact that $\sum_{t=1}^M \tilde{x}_{k,t-1} = 0$.

³ $M \rightarrow \infty$ implies $\hat{\beta}_{SFE} \xrightarrow{p} 0$, which is in line with Proposition 1.

within a single subsample is decreasing as M increases, since the number of observations in a subsample grows. Therefore choosing the number of subsamples involves a bias-variance trade-off.

To obtain the asymptotic results in Proposition 2, I need to impose fairly strong assumptions. In particular, Proposition 2 allows only strictly exogenous regressors. However, I show in a simulation exercise in the following section that the method also appears robust with endogenous regressors.

4 Simulations

This section carries out a Monte Carlo experiment to analyze finite sample properties of the model and the proposed estimator. The simulations suggest that the presence of a non-stationary component in the regressor substantially biases OLS estimates even for moderate sample sizes. Using SFE mitigates this bias. However, the results of the subsampling estimator are sensitive to the choice of subsample size.

The simulation set-up is based on the data generating assumption in equations (1)-(3) and the predictive regression (4). The common assumptions in these specifications is that the information part of the predictor, η_{t-1} , is i.i.d., and $\sigma_\eta = \sigma_u = 1$ for normalization. The underlying slope coefficient of equation (1) is set to $\beta_0 = 0.2$ and no intercept is used.

For the remaining parameters of the model, I analyze a set of difference scenarios. In particular, I consider three values of the signal-to-noise ratio $\lambda = \frac{\sigma_\eta}{\sigma_\varepsilon}$, $\lambda = \{1, 3, 10\}$, representing different levels of persistence caused by the non-stationary component.⁴ For the correlation between u_t and ε_t , the simulations use two values, $\rho_{u,\varepsilon} = 0$ and $\rho_{u,\varepsilon} = -0.8$, which represent whether the predictor is strictly exogenous, or there is a negative correlation between the error terms that causes endogeneity (and hence a persistent regressor bias). The results using 1000 repetitions are presented in Table 1 for sample sizes of $T = \{100, 300, 800\}$. These sample sizes represent the number of available observations in yearly, quarterly or monthly datasets generally used to assess return predictability.

⁴On a sample of 1000 observations, these choices of λ correspond to first order autocorrelations of approximately 0.99, 0.95 and 0.7, respectively.

[Insert Table 1 here]

The first three columns in Table 1 represent the case of an exogenous regressor, $\rho_{u,\varepsilon} = 0$. The first observation is that the extent of bias for different sample sizes strongly depends on the persistence of the predictor. If the informative component is small ($\lambda = 1$), then the sample autocorrelation is almost completely driven by the non-stationary component, and there is already a serious bias for a small sample ($T = 100$). This bias decreases as the informative component of the predictor strengthens (λ becomes larger), which implies a relatively weaker non-stationary component, and thus, a lower persistence. The convergence of the slope coefficient to zero is apparent in each case, and it happens quickly. The coefficients are substantially closer to zero for a sample size of $T = 800$.

Columns (5)–(7) in Table 1 show the results based on simulations in which the strict exogeneity assumption of the regressor is violated. $\rho_{u,\varepsilon} = -0.8$ means a strong negative relationship between the innovations u_t and ε_t . As Stambaugh (1999) argued, this correlation creates an upward bias in the estimation, which one can observe in columns (5)–(7). This results in an over-rejection of the null hypothesis of no predictability in the absence of non-stationarity. However, if the autocorrelation of the regressor is substantial, then the parameter estimates are considerably biased towards zero, even though the estimated slope coefficients are larger in absolute value. In fact, the endogeneity bias enters the estimates nearly additively and the sample size does not affect its extent. Thus, as the sample size increases, the effect of the non-stationary component becomes stronger than that of the endogeneity bias, and therefore the slope coefficients eventually converge to zero (the vanishing predictability phenomenon dominates).

The findings of the simulations have important practical implications for predictive regressions. In particular, even if statistically significant predictive power is found in a given sample, it is not certain that the estimation precision of the relationship improves when using more observations. In fact, the contrary holds in the present case. The larger the sample is, the more biased the least squares estimation becomes.

Next, the properties of the subsampling fixed effects method are analyzed. The simulations use the same data generating process as above and apply the SFE estimator with

subsample sizes of $M = \{10, 25, 50\}$. Given M and the sample size $T(\geq M)$, the subsampling partition $\{T_1, T_2, \dots, T_K\}$ is uniquely determined. Panel I in Table 2 presents the mean of the simulated SFE estimates. The standard deviation of the simulated empirical distribution (Panel II), or the standard error of the SFE estimator based on the simulations results, is calculated based on 1000 repetitions.

[Insert Table 2 here]

As Table 2 shows, the SFE is robust to the size of the entire sample. The point estimates for $T = 100$ and $T = 800$ are almost identical. This is unsurprising given the theoretical results, as the accumulated variance of the non-stationary component is constrained to a given subsample and thus the overall bias is similar to the bias that appears in one subsample. To further illustrate the point, Figure 1 plots the SFE estimates as a function of the sample size T , for a fixed subsample size $M = 50$.

[Insert Figure 1 here]

Proposition 2 implies that the subsample size plays a key role in determining the bias in the SFE estimator. This is also confirmed by the simulation results, showing that the subsample size M does have a significant effect on the estimation results even in finite samples. Comparing results with $M = 10$ to $M = 50$ in Table 2, it is clear that the smaller subsample size results in a less biased, but more imprecise estimation. The point estimates are closer to the true value if the subsample size is small (Panel I), but their standard errors are larger (Panel II). This is independent of the fact whether there is correlation between the unexpected return and the innovation of the non-stationary component. In particular, the endogeneity bias enters as an additional factor to the point estimates.

5 Choice of subsample size

Both the theoretical and simulation results point to the importance of the efficiency-bias trade-off when choosing the subsample size M . To elaborate on this point further,

note that the mean squared error (MSE) of the subsampling estimator is

$$MSE = (E(\beta_{SFE}) - \beta_0)^2 + Var(\beta_{SFE}), \quad (7)$$

for a given sample size T . Now, for large samples (i.e., $T \rightarrow \infty$) and for fixed M , the variance of the estimator shrinks arbitrarily small, while the first term that captures the bias, does not disappear per Proposition 2. Therefore,

$$MSE \xrightarrow{T \rightarrow \infty} \beta_0^2 \left(\frac{\sigma_\eta^2}{\sigma_\eta^2 + \frac{M+1}{6}\sigma_\varepsilon^2} - 1 \right)^2;$$

that is, the error is driven completely by the asymptotic bias. The above expression increases in the absolute value of M . Therefore, for sufficiently large T , the best choice of M is as small as possible. This is somewhat at odds with the usual notion that the subsample size should be chosen as $M \rightarrow \infty$, such that $M/T \rightarrow 0$. However, the result mirrors the basic intuition behind the estimator: the smaller the subsample is, the less the non-informative component can affect estimation.

The asymptotic results provide poor guidance on how to choose a subsample size in finite samples. This follows from the fact that for a small M , the number of parameters to estimate (relative to the sample size) is large, and hence, the variance of the estimator is substantial. To see how the bias and variance of the estimator interact in sample sizes relevant for the predictability context, I calculate the MSE of the subsampling estimator based on the simulated mean and the standard deviation of $\hat{\beta}_{SFE}$ using equation (7). The specifications are the same as in Table 2, except that I consider a finer set of subsample sizes $M \in [10, 50]$.

[Insert Figure 2 here]

Figure 2 shows that the trade-off between the bias and variance highly depends on the signal-to-noise ratio. In Panel (a), where the informative component explains a very small part of the predictor variation, the error grows monotonically with the subsample size. However, in the other two scenarios, in which the non-stationary component is less strong, the optimal choice of subsample size depends on the size of the sample. In

particular, for $\lambda = 3$, the MSE minimizing subsample size is around 20 for annual samples. Additionally, for $\lambda = 10$ — which corresponds to a sample autocorrelation of around 0.7 in monthly samples — the optimal subsample size is even larger, between 25 and 35 for all (empirically relevant) sample sizes considered.

From a practical point of view, the main implication of Figure 2 is that the choice of subsample size depends on the data generating process. *A priori*, we do not have information about all unobservable components of the model, and λ cannot be identified by looking at the predictor process only.⁵ Therefore, an empirically feasible and robust approach is to use several subsample sizes (where a relatively small subsample size is chosen, such as below 50), and compare the results based on the set of estimates.

6 Empirical results

I use the monthly dataset compiled by Goyal and Welch (2008), who perform a comprehensive analysis of the predictors of the excess return on the S&P 500 stock market index.⁶ The time window of the analysis is between January 1952 and December 2017.⁷ Since the focus here is on prediction with highly persistent variables, I consider variables with high serial correlation (their estimated first order autocorrelation is above 0.95). The analysis is further restricted to the variables with monthly observations available for the full sample period.⁸ Consequently, I assess the predictive capacity for seven time series, including the dividend–price ratio (dp), earnings–price ratio (ep), dividend payout ratio (de), book-to-market value (bm), three-month treasury bill rate (tbl), term spread on government bonds (tms), and default yield spread (dfy).⁹ Table 3 presents the summary statistics of the predictors and the excess returns.

⁵For example, if the persistence of the predictor is partly due to a (mild) autocorrelation in the predictor information, then λ can be small, even if the informative component is relatively strong.

⁶<http://www.hec.unil.ch/agoyal/>

⁷Since the analysis includes the variables related to the short rate, the sample starts in 1952, when independent monetary policy became possible, and the interest rate variables became informative. Campbell and Yogo (2006) and Pástor and Stambaugh (2009), among others, start their samples in the same year for this reason.

⁸Although the results in this section are based on monthly data, I also considered the prediction on a quarterly basis. The empirical findings remain qualitatively unchanged, which suggests that aggregation to lower frequencies does not change the results.

⁹The variable descriptions can be found in Goyal and Welch (2008).

[Insert Table 3 here]

Since the persistence of the explanatory variables is key to the analysis, Table 3 report the OLS estimate of the largest autoregressive root and the p-value of an Augmented Dickey-Fuller test. The autoregressive roots are close to unity, and the existence of a unit root cannot be rejected (except for the term spread, which represents a borderline case: the null of a unit root is rejected at the five percent significance level, but not at the one percent level). The high persistence makes these variables good candidates that fulfil the assumptions of the model, namely, that the stationary informative component is masked by a non-stationary noise.

6.1 Vanishing predictability in the data

Although the results of Proposition 1 are asymptotic, the implied bias is also substantial in finite samples. Empirically, this is testable by looking at the changes in the estimated slope coefficient for different sample sizes. To obtain a set of slope coefficients that correspond to different sample sizes, I perform an extending window analysis of the standardized values of the variables.¹⁰ First, I consider 100 observations of the dataset and estimate equation (4) with OLS. Then, the sample is extended by adding one more observation, and equation (4) is re-estimated. This procedure is iterated until no new data points are available. I carry out this analysis for each of the predictors discussed above.

The sequences of the estimated slope coefficients for four variables (the dividend–price ratio, book-to-market value and interest-related variables, treasury-bill rate, and term spread) are presented in Figure 3. Even though the overall evidence for predictability is weak, these results show gradually decreasing predictability, in line with the theoretical results.

[Insert Figure 3 here]

In contrast, the remaining three variables — earnings–price ratio, dividend payout ratio and default yield spread — appear to be essentially non-predictors. Their slope

¹⁰I standardize the variables to obtain comparable results across predictors.

coefficients cannot be statistically distinguished from zero at any sample size, even with the relatively small OLS standard errors. This suggests that the information component in these variables is negligible, and they do not predict excess returns. In particular, estimates of the slope coefficients are statistically indistinguishable from zero for all sample sizes.¹¹

The empirical findings so far are conditional on the selected start date and the specific sample. Proposition 1 does not assume or require a specific initial condition. In fact, results are independent of the start date and value of the processes, which means that vanishing predictability does not depend on the start date of the sample according to the model. This can be tested by using subsets of observations. First, the time series between January 1952 and December 1996 are considered, omitting the last twenty years of observations. The sequences of the slope coefficients are obtained by the extending window analysis described above (this gives the same sequences as before, truncated at 1996). Then, the start and end of the sample are shifted ten years forward in time and the same exercise is carried out, giving a new set of sequences of slope coefficients for the shifted sample. This procedure is repeated twice, resulting in three sets of results, presented in Figure 4. Although the coefficient series vary significantly over time, the tendency of decreasing predictive power prevails in the shifted samples. This confirms the intuition from the baseline results overall: non-stationary predictors that potentially have predictive power for the excess returns tend to lose power over time.

[Insert Figure 4 here]

6.2 Applying subsample fixed effects

The results based on the OLS estimation suggest that vanishing predictability is present in the data. Therefore, one can expect the SFE estimator to improve the results by reducing the bias in the point estimation. I estimate the slope coefficients of the predictive regression using the SFE based on equation (5)-(6). The variables are those that are subject to vanishing predictability: dividend–price ratio, book-to-market value,

¹¹All results using this group of variables are shown in the Appendix.

treasury bill rate, and term spread. Since the results are sensitive to the choice of subsample size, the models are estimated with subsample sizes of 10, 25 and 50. The coefficient estimates and standard errors are reported in Table 4 (The OLS results are presented as a benchmark in the last column).

[Insert Table 4 here]

The findings in Table 4 are in line with the theoretical results. First, while the OLS results for the entire sample are barely significant, the subsampling results are remarkably stronger. In particular, most of the subsampling coefficients are significant at the one percent level. Second, the point estimates become larger in absolute value as the subsample size decreases. This is completely in line with the theoretical prediction of Proposition 2, which suggests that for smaller subsamples the bias caused by the non-stationary component decreases. This, in turn, makes the relationship between the informative component of the predictor and the excess return easier to reveal. The results therefore suggest that the underlying predictive power of the variables is stronger than the results based on a standard least squares estimation using the full sample.

Although the SFE estimator is defined using subsamples of equal size, the idea of subsampling is more general. If a persistent variable exhibits vanishing predictability, estimating subsamples can help mitigate the problem. An alternative approach to equal-sized subsamples is to estimate structural breaks in the time series of the explanatory variable and define the subsamples as observations between two breaks. This is analogous to the approach Lettau and Van Nieuwerburgh (2008) follow to examine the dividend-price ratio by identifying structural breaks with the method in Bai and Perron (1998).¹²

To see how the results change when using estimated cut-off points, I estimate the model by Bai and Perron (1998) to identify breaks between subsamples. The number of breaks in each series is specified in advance, and I let the method determine their location. This facilitates comparison with the SFE estimator, since by specifying the number of cut-off points, the average subsample size is also defined. Using the break-adjusted explanatory

¹²For a given number of breaks, the method of (Bai and Perron, 1998) estimates a linear model with subsample dummies, where the cut-offs between subsamples are determined such that the resulting equation model has the smallest MSE.

variables based on the estimated cut-off values, I estimate the predictive regression (6). Table 5 presents the results based on this approach. They are qualitatively similar to the ones using SFE. In particular, the results are generally strongly significant, and the coefficients seem to be even further away from zero for a given (average) subsample size in the estimated cut-off case. Overall, the evidence based on an estimated cut-off is in line with the theoretical predictions of the model, suggesting that estimation using subsamples can provide stronger evidence of predictability.

[Insert Table 5 here]

6.3 Vanishing or time-varying predictability?

The empirical evidence described so far is consistent with the time-invariant predictive relationship and the presence of a highly persistent uninformative component, as the model (1)-(3) describes. However, the shrinking predictive power of the variables can also occur because periods with stronger predictability are blended with periods with weaker predictive relationship (as Farmer et al., 2018 suggest). In theory, it is simple to disentangle these two scenarios: if the coefficients become consistently larger in absolute terms as smaller subsamples are used, then the results support vanishing predictability. On the other hand, a large variation in the estimated slope coefficients between subsamples suggests that the underlying predictive relationship is time-varying.

To investigate this question, Figure 5 shows rolling window estimates of the slope coefficient of the univariate predictive regression for each variable discussed.¹³ First, there is a substantial variation in the slope coefficients, suggesting support for the time-varying predictability argument. However, for the financial ratios (dividend–price ratio and the book-to-market value), the coefficients on smaller subsamples are consistently higher than those using larger subsamples or the entire sample, which suggests that overall, using a larger sample results in finding weaker predictability. Therefore, vanishing predictability seems to explain at least part of the variation in the slope coefficients of these variables.

¹³To partly account for the large estimation uncertainty introduced by using very small samples, I present “smoothed” rolling window results (see the description in Figure 5).

[Insert Figure 5 here]

7 Conclusion

Several predictors of stock market returns (such as financial ratios or interest-related variables) are highly persistent. This is in contrast to excess returns, which prior works usually find to be weakly dependent, almost white noise processes. I reconcile the potential non-stationarity in the explanatory variables and a stationary expected return by assuming a noisy predictor. The main result is that if stationary returns are regressed on the lagged values of a non-stationary explanatory variable, then the slope coefficient and the predictive power approach zero as the sample size increases. This observation is in line with the empirical evidence of weakening predictive power of several regressors presented in Section 6.

The key result of the model holds for a general set of assumptions about the innovations in the regressor and the unexpected returns. Using Monte Carlo simulations, I also show that the convergence of the slope coefficients to zero happens quickly; therefore, the estimates are biased, even for moderate sample sizes.

The proposed SFE estimator puts a bound on the variance of the non-stationary component, and it therefore reduces the bias caused by the high persistence of the explanatory variable. I derive exact theoretical results for the extent of the bias for a fixed subsample size for a restrictive set of assumptions. The simulations show that the estimator also works well for more general assumptions. Applying this estimator to highly persistent predictors of the returns shows that the point estimates improve and tend to become significant.

References

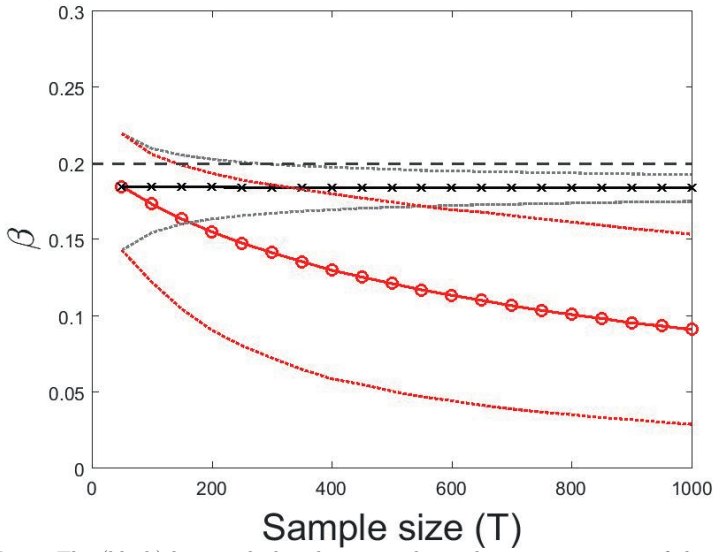
- Ang, A. and Bekaert, G. (2007). Stock return predictability: Is it there? *Review of Financial Studies*, 20(3):651–707.
- Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, pages 47–78.
- Boudoukh, J., Richardson, M., and Whitelaw, R. F. (2008). The myth of long-horizon predictability. *Review of Financial Studies*, 21(4):1577–1605.
- Campbell, J. Y. and Yogo, M. (2006). Efficient tests of stock return predictability. *Journal of Financial Economics*, 81(1):27–60.
- Cavanagh, C. L., Elliott, G., and Stock, J. H. (1995). Inference in models with nearly integrated regressors. *Econometric Theory*, 11(05):1131–1147.
- Cochrane, J. H. (2008). The dog that did not bark: A defense of return predictability. *Review of Financial Studies*, 21(4):1533–1575.
- Deng, A. (2013). Understanding spurious regression in financial economics. *Journal of Financial Econometrics*, 12(1):122–150.
- Farmer, L., Schmidt, L., and Timmermann, A. (2018). Pockets of predictability. Technical report, SSRN.
- Ferson, W. E., Sarkissian, S., and Simin, T. T. (2003). Spurious regressions in financial economics? *Journal of Finance*, 58(4):1393–1414.
- Goetzmann, W. N. and Jorion, P. (1993). Testing the predictive power of dividend yields. *The Journal of Finance*, 48(2):663–679.
- Goyal, A. and Welch, I. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 21(4):1455–1508.
- Koijen, R. S. and Van Nieuwerburgh, S. (2011). Predictability of returns and cash flows. *Annual Review of Financial Economics*, 3:467–491.

- Lettau, M. and Van Nieuwerburgh, S. (2008). Reconciling the return predictability evidence. *Review of Financial Studies*, 21(4):1607–1652.
- Lewellen, J. (2004). Predicting returns with financial ratios. *Journal of Financial Economics*, 74(2):209–235.
- Moon, R., Rubia, A., and Valkanov, R. (2005). Long horizon regressions when the predictor is slowly varying. *Working Paper*.
- Moon, S. and Velasco, C. (2014). On the properties of regression tests of stock return predictability using dividend-price ratios. *Journal of Financial Econometrics*, 12(1):151–173.
- Nelson, C. R. and Kim, M. J. (1993). Predictable stock returns: The role of small sample bias. *The Journal of Finance*, 48(2):641–661.
- Osterrieder, D., Ventosa-Santaulria, D., and Vera-Valds, J. E. (2015). Unbalanced regressions and the predictive equation. CREATES Research Papers 2015-09, Department of Economics and Business Economics, Aarhus University.
- Pástor, L. and Stambaugh, R. F. (2009). Predictive systems: Living with imperfect predictors. *The Journal of Finance*, 64(4):1583–1628.
- Phillips, P. C. B. (1986). Understanding spurious regressions in econometrics. *Journal of Econometrics*, 33(3):311–340.
- Phillips, P. C. B. (1987). Time series regression with a unit root. *Econometrica*, 55(2):277–301.
- Stambaugh, R. F. (1999). Predictive regressions. *Journal of Financial Economics*, 54(3):375–421.
- Torous, W., Valkanov, R., and Yan, S. (2004). On predicting stock returns with nearly integrated explanatory variables. *The Journal of Business*, 77(4):937–966.

Valkanov, R. (2003). Long-horizon regressions: theoretical results and applications. *Journal of Financial Economics*, 68(2):201–232.

Figures

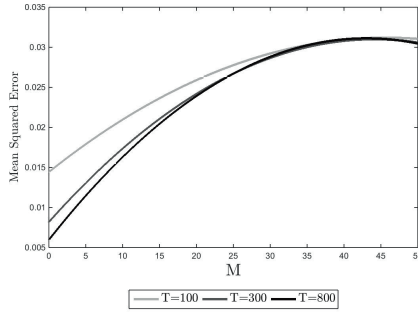
Figure 1: Example of the subsample fixed effects estimator



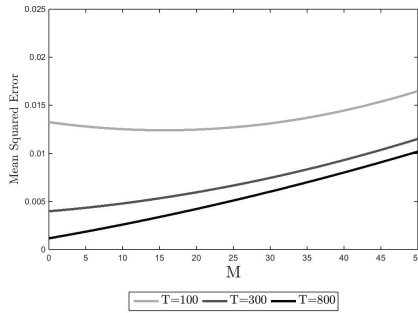
Notes: The (black) line marked with crosses shows the point estimates of the slope coefficient of the SFE estimator in the simulated model described by equations (1)-(4) with parameter choices $\sigma_\varepsilon = \sigma_u = 1$, $\sigma_\eta = 10$. The chosen true value of the slope coefficient is $\beta_0 = 0.2$ (marked with a dashed line in the figure), and the subsamples have 50 observations ($M = 50$). The (red) line marked with circles represents the OLS results based on the same samples. The dotted lines represent the respective 95 percent confidence intervals based on the simulated sampling distributions.

Figure 2: Mean squared errors as a function of subsample size

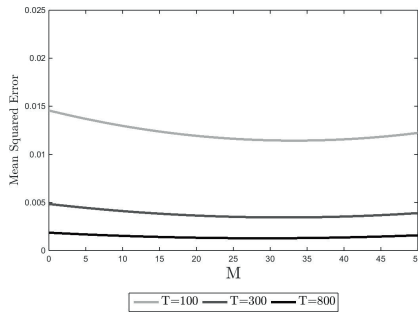
(a) $\lambda = 1$



(b) $\lambda = 3$

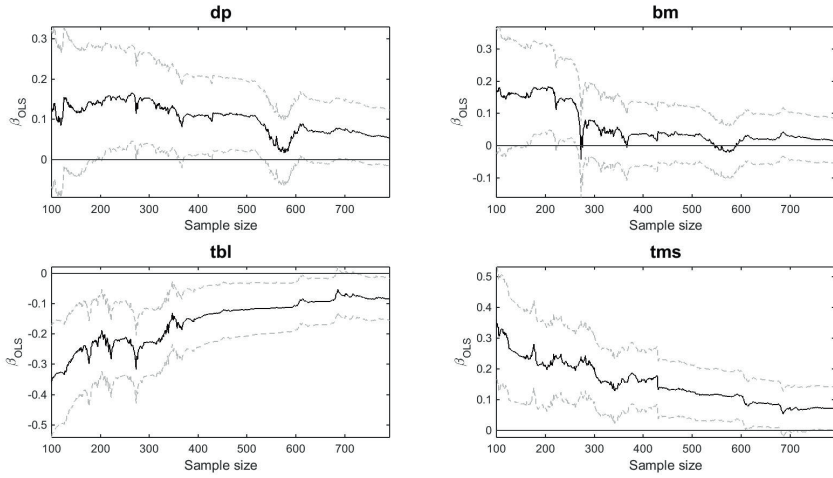


(c) $\lambda = 10$



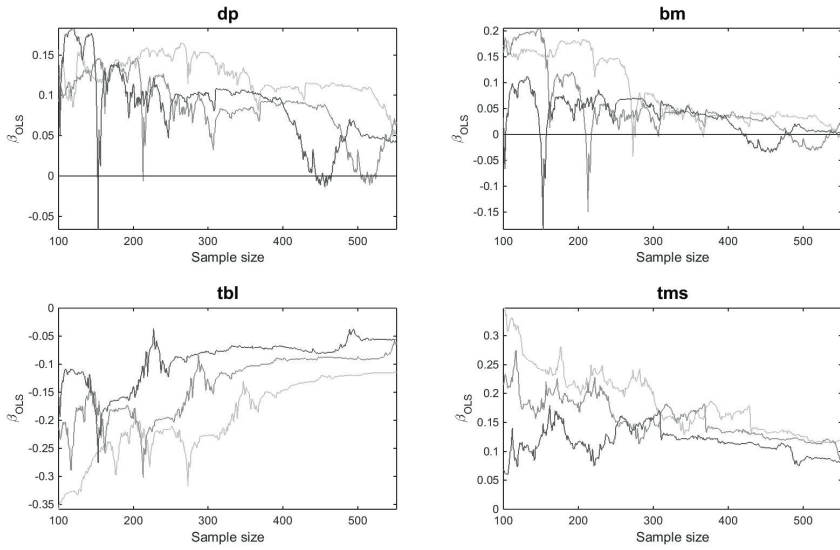
Notes: The plots show the simulated mean squared errors based on the simulation setup described in equations (1)-(4). The chosen true value of the slope coefficient is $\beta_0 = 0.2$. Each panel represents different choices of signal-to-noise ratio, while the different lines correspond to various sample sizes (T). The subsample size M is depicted on the horizontal axis. The results are based on 1000 repetitions.

Figure 3: Extending window analysis of the predictors



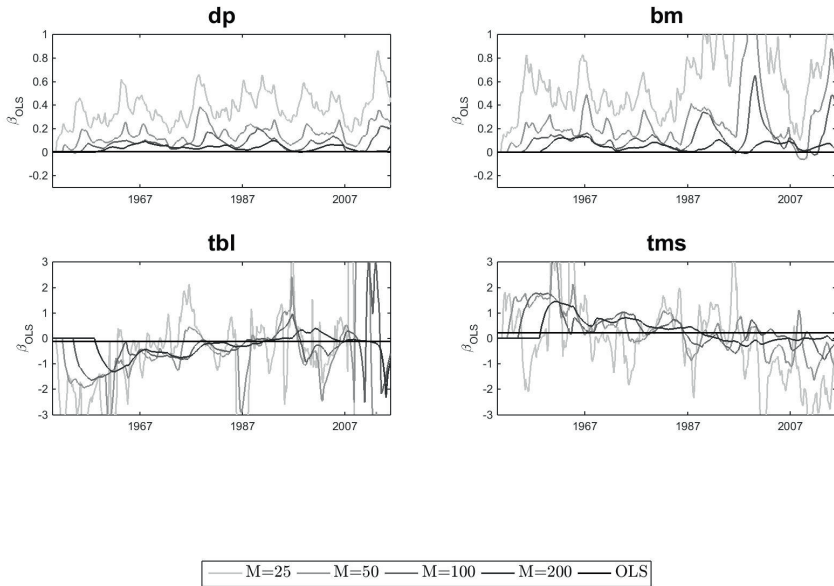
Notes: The figures show the estimated slope coefficients of the univariate predictive regressions as a function of the sample size. All samples start in January 1952. The explanatory variables in the figures are the dividend–price ratio (**dp**), book-to-market value (**bm**), three-month treasury bill rate (**tbl**), and term spread on the government bonds (**tms**).

Figure 4: Sensitivity of the slope coefficient to the sample start date



Notes: The lines represent the estimated slope coefficients of the univariate predictive regressions as a function of the sample size, with different starting points. The lightest grey line represents January 1952, and darker shades mean later starting points. Variable descriptions are given in Figure 3.

Figure 5: Rolling window slope coefficients



Notes: The lines represent the smoothed rolling window estimates of the standard univariate predictive regression using different subsample sizes. That is, $\beta_t^{Smooth} = \omega\beta_{t-1}^{Smooth} + (1 - \omega)\beta_t^{Rolling}$, where $\beta_t^{Rolling}$ is estimated on the rolling window $[t, t - M]$ and $\omega = 0.95$. The variable descriptions are given in Figure 3.

Tables

Table 1: Slope coefficients of the predictive regression

	$\rho_{u,\varepsilon} = 0$			$\rho_{u,\varepsilon} = -0.8$		
	$\lambda = 1$	$\lambda = 3$	$\lambda = 10$	$\lambda = 1$	$\lambda = 3$	$\lambda = 10$
T=100	0.0185	0.0836	0.1792	0.0527	0.1429	0.2028
T=300	0.0062	0.0432	0.1411	0.0201	0.0736	0.1712
T=800	0.0027	0.0191	0.1010	0.0077	0.0343	0.1183

Notes: This table presents the OLS estimates of the slope coefficient β in the regression $y_t = \alpha + \beta x_{t-1} + e_t$ with different sample sizes (T). (y_t, x_t) are generated using equations (1)-(2) with $\beta_0 = 0.2$. The information component of the predictor, η_t is i.i.d. with $\sigma_\eta (= \sigma_u) = 1$. λ is the signal-to-noise ratio that determines the persistence of the explanatory variable. Columns (2)-(4) represent the exogenous cases, in which $\rho_{u,\varepsilon} = \text{Corr}(u, \varepsilon) = 0$, while columns (5)-(7), $\rho_{u,\varepsilon} = -0.8$, correspond to endogenous regressors.

Table 2: Subsampling fixed effects estimator

I: Mean of $\hat{\beta}_{SFE}$

	$\rho_{u,\varepsilon} = 0$			$\rho_{u,\varepsilon} = -0.8$		
	$\lambda = 1$	$\lambda = 3$	$\lambda = 10$	$\lambda = 1$	$\lambda = 3$	$\lambda = 10$
$M = 10$						
T=100	0.0688	0.1610	0.1985	0.2176	0.2786	0.2336
T=300	0.0694	0.1663	0.1973	0.2158	0.2757	0.2385
T=800	0.0709	0.1655	0.1966	0.2117	0.2765	0.2363
$M = 25$						
T=100	0.0418	0.1342	0.1853	0.1246	0.2307	0.2316
T=300	0.0388	0.1357	0.1933	0.1180	0.2250	0.2290
T=800	0.0370	0.1364	0.1908	0.1149	0.2258	0.2305
$M = 50$						
T=100	0.0256	0.1149	0.1771	0.0790	0.1856	0.2188
T=300	0.0229	0.1069	0.1859	0.0695	0.1756	0.2178
T=800	0.0223	0.1052	0.1848	0.0652	0.1720	0.2229

II: Standard deviation of $\hat{\beta}_{SFE}$

	$\rho_{u,\varepsilon} = 0$			$\rho_{u,\varepsilon} = -0.8$		
	$\lambda = 1$	$\lambda = 3$	$\lambda = 10$	$\lambda = 1$	$\lambda = 3$	$\lambda = 10$
$M = 10$						
T=100	0.0651	0.0967	0.1088	0.0661	0.0954	0.1017
T=300	0.0376	0.0565	0.0591	0.0363	0.0515	0.0612
T=800	0.0234	0.0342	0.0367	0.0224	0.0341	0.0374
$M = 25$						
T=100	0.0503	0.0876	0.0993	0.0559	0.0835	0.0959
T=300	0.0274	0.0510	0.0571	0.0326	0.0499	0.0573
T=800	0.0167	0.0308	0.0348	0.0195	0.0300	0.0367
$M = 50$						
T=100	0.0393	0.0814	0.0957	0.0477	0.0865	0.0986
T=300	0.0211	0.0470	0.0547	0.0270	0.0467	0.0565
T=800	0.0122	0.0279	0.0345	0.0152	0.0284	0.0334

Notes: This table presents the means (Panel I) and standard deviations (Panel II) of the simulated SFE estimates in Equation (4) with subsample sizes of $M = \{25, 50, 100\}$. The data generating process with further parameter specifications are given in the description of Table 1. The simulation is repeated 1000 times.

Table 3: Summary statistics of the predictors and excess returns

	mean	stdev	ϕ	p-value	N
dp	-3.5391	1.4000	0.9933	0.8772	792
bm	0.5143	0.8547	0.9943	0.2121	792
tbl	0.0432	0.1072	0.9914	0.2497	792
tms	0.0172	0.0485	0.9572	0.0098	792
ep	-2.8075	1.4460	0.9891	0.7279	792
de	-0.7316	1.0211	0.9865	0.3981	792
dfy	-0.0097	0.0152	0.9702	0.1518	792
ret	0.0596	0.1444	0.0519	0.0010	792

Notes: **stdev** indicates the standard deviation of the variable and **N** is the number of observations. Column (3) presents the first order autocorrelation of the variables, while column (4) shows the empirical significance level of the Augmented Dickey Fuller test (without deterministic trend and drift). The table presents the dividend–price ratio (**dp**), earnings–price ratio (**ep**), dividend payout ratio (**de**), book-to-market value (**bm**), three-month treasury bill rate (**tbl**), term spread on the government bonds (**tms**), and default yield spread (**dfy**) variables. The variable **ret** is the excess return on the S&P 500 index.

Table 4: Regression results from the one-period-ahead forecasts using a fixed subsample size

	Subsample size (M)			
	10	25	50	OLS
dp	0.2070 (0.0247) ^{***}	0.0886 (0.0163) ^{***}	0.0538 (0.0129) ^{***}	0.0055 (0.0040)
bm	0.1927 (0.0401) ^{***}	0.1131 (0.0248) ^{***}	0.0709 (0.0207) ^{***}	0.0024 (0.0064)
tbl	-0.4905 (0.2071) ^{**}	-0.4534 (0.1479) ^{***}	-0.2441 (0.0934) ^{***}	-0.1149 (0.0493) ^{**}
tms	0.1770 (0.2462)	0.4538 (0.1777) ^{**}	0.2437 (0.1418) [*]	0.2158 (0.1179) [*]

Notes: The table presents the slope coefficients and standard errors of the univariate predictive regression. The SFE estimator is used. Standard errors are calculated using residual block bootstrapping, where the length of the blocks is $O(T^{1/3})$. The column header specifies the size of the subsample. The table presents the dividend–price ratio (**dp**), book-to-market value (**bm**), three-month treasury bill rate (**tbl**), and term spread on the government bonds (**tms**) variables. ^{***}, ^{**}, and ^{*} represent statistical significance at the 1%, 5%, and 10% level, respectively.

Table 5: Regression results from the one-period-ahead forecasts using the estimated subsample cut-offs

	10	8	4	2
dp	0.0559 (0.0140)***	0.0462 (0.0126)***	0.0319 (0.0079)***	0.0219 (0.0066)***
bm	0.1142 (0.0269)***	0.0848 (0.0231)***	0.0521 (0.0161)***	0.0115 (0.0099)
tbl	-0.5346 (0.1419)***	-0.4870 (0.1314)***	-0.3223 (0.0837)***	-0.1632 (0.0580)***
tms	0.6124 (0.1840)***	0.3878 (0.1696)**	0.2532 (0.1388)*	0.2177 (0.1268)*
Av. size	79	99	198	396

Notes: The table presents the slope coefficients and standard errors of the univariate predictive regression. Subsample fixed effects are used together with the approach in Bai and Perron (1998) to estimate cut-off values for the subsamples. Standard errors are from the classical OLS formula. The column header specifies the number of subsamples (which also determine the average sample size, shown in the last row). The table presents the dividend–price ratio (**dp**), book-to-market value (**bm**), three-month treasury bill rate (**tbl**), and term spread on the government bonds (**tms**) variables. ***, **, and * represent statistical significance at the 1%, 5%, and 10% level, respectively.

Appendix

A Proofs

A.1 Proof of Proposition 1

First, consider the slope coefficient (without loss of generality assume that the first observation of x is x_0),

$$\begin{aligned}\hat{\beta}_{OLS} &= \frac{\frac{1}{T} \sum_{t=1}^T y_t x_{t-1} - \left(\frac{1}{T} \sum_{t=1}^T y_t \right) \left(\frac{1}{T} \sum_{t=1}^T x_{t-1} \right)}{\frac{1}{T} \sum_{t=1}^T (x_{t-1} - \bar{x}_{t-1})^2} \\ &= \frac{1}{\sqrt{T}} \frac{\frac{1}{T^{3/2}} \sum_{t=1}^T y_t x_{t-1} - \left(\frac{1}{T} \sum_{t=1}^T y_t \right) \left(\frac{1}{T^{3/2}} \sum_{t=1}^T x_{t-1} \right)}{\frac{1}{T^2} \sum_{t=1}^T (x_{t-1} - \bar{x}_{t-1})^2}.\end{aligned}\quad (8)$$

Using the results $\frac{1}{T^{3/2}} \sum_{t=1}^T y_t \xi_{t-1} \xrightarrow{p} 0$ in Phillips (1987), $\frac{1}{T} \sum_{t=1}^T y_t \xrightarrow{p} E(y_t) = E(\alpha_0 + \beta_0 \eta_{t-1} + u_t) = \alpha_0$, $\frac{1}{T^{3/2}} \sum_{t=1}^T \eta_{t-1} \xrightarrow{p} 0$ and $\frac{1}{T^{3/2}} \sum_{t=1}^T y_t \eta_{t-1} \xrightarrow{p} 0$ by the Weak Law of Large Numbers, and $\frac{1}{T^{3/2}} \sum_{t=1}^T \xi_{t-1} \Rightarrow \int_0^1 W(r) dr$ in Phillips (1986), where $W(r)$ is a standard Brownian motion. The numerator can be written as

$$\begin{aligned}&\frac{1}{T^{3/2}} \sum_{t=1}^T y_t x_{t-1} - \left(\frac{1}{T} \sum_{t=1}^T y_t \right) \left(\frac{1}{T^{3/2}} \sum_{t=1}^T x_{t-1} \right) = \\ &\frac{1}{T^{3/2}} \sum_{t=1}^T y_t \xi_{t-1} + \frac{1}{T^{3/2}} \sum_{t=1}^T y_t \eta_{t-1} - \left(\frac{1}{T} \sum_{t=1}^T y_t \right) \left(\frac{1}{T^{3/2}} \sum_{t=1}^T \xi_{t-1} + \frac{1}{T^{3/2}} \sum_{t=1}^T \eta_{t-1} \right) = \\ &\frac{1}{T^{3/2}} \sum_{t=1}^T y_t \xi_{t-1} + \frac{1}{T^{3/2}} \sum_{t=1}^T y_t \eta_{t-1} - \left(\frac{1}{T} \sum_{t=1}^T y_t \right) \left(\frac{1}{T^{3/2}} \sum_{t=1}^T \xi_{t-1} \right) - \left(\frac{1}{T} \sum_{t=1}^T y_t \right) \left(\frac{1}{T^{3/2}} \sum_{t=1}^T \eta_{t-1} \right) \\ &\Rightarrow \alpha_0 \int_0^1 W(r) dr.\end{aligned}$$

Furthermore, using $\frac{1}{T^2} \sum_{t=1}^T (\xi_{t-1} - \bar{\xi}_{t-1})^2 \Rightarrow \int_0^1 W(r)^2 dr - \left(\int_0^1 W(r) dr \right)^2$ in Phillips (1986), the denominator converges to

$$\begin{aligned} \frac{1}{T^2} \sum_{t=1}^T (x_{t-1} - \bar{x}_{t-1})^2 &= \frac{1}{T^2} \sum_{t=1}^T (\xi_{t-1} - \bar{\xi}_{t-1})^2 + \frac{1}{T^2} \sum_{t=1}^T (\eta_{t-1} - \bar{\eta}_{t-1})^2 \\ &+ 2 \frac{1}{T^2} \sum_{t=1}^T (\eta_{t-1} - \bar{\eta}_{t-1})(\xi_{t-1} - \bar{\xi}_{t-1}) \Rightarrow \int_0^1 W(r)^2 dr - \left(\int_0^1 W(r) dr \right)^2. \end{aligned}$$

Therefore, by the joint convergence results in Phillips (1987), I obtain

$$\sqrt{T} \hat{\beta} \Rightarrow \frac{\alpha_0 \int_0^1 W(r) dr}{\int_0^1 W(r)^2 dr - \left(\int_0^1 W(r) dr \right)^2},$$

which establishes the result for $\hat{\beta}$. For the intercept,

$$\hat{\alpha} = \frac{1}{T} \sum_{t=1}^T y_t - \hat{\beta} \frac{1}{T} \sum_{t=1}^T x_t = \frac{1}{T} \sum_{t=1}^T y_t + o_p(1) \xrightarrow{p} E(y_t) = \alpha_0.$$

A.2 Proof of Lemma 1

First, write out $\tilde{x}_{k,t}$.

$$\begin{aligned} \tilde{x}_{k,t} &= x_{k,t} - \frac{1}{M} \sum_{m=1}^M x_{k,m} = \eta_{k,t} + \xi_{k,t} - \frac{1}{M} \sum_{m=1}^M \eta_{k,m} - \frac{1}{M} \sum_{m=1}^M \xi_{k,m} = \\ &\eta_{k,t} - \frac{1}{M} \sum_{m=1}^M \eta_{k,m} + \xi_{k,0} + \sum_{s=1}^t \varepsilon_{k,s} - \frac{1}{M} \sum_{m=1}^M \left(\sum_{s=1}^m \varepsilon_{k,s} + \xi_{k,0} \right) = \\ &\eta_{k,t} - \frac{1}{M} \sum_{m=1}^M \eta_{k,m} + \sum_{s=1}^t \varepsilon_{k,s} - \frac{1}{M} \sum_{m=1}^M \sum_{s=1}^m \varepsilon_{k,s} = \\ &\eta_{k,t} - \frac{1}{M} \sum_{m=1}^M \eta_{k,m} + \sum_{s=1}^t \varepsilon_{k,s} - \sum_{m=1}^M \frac{M-m+1}{M} \varepsilon_{k,m}. \end{aligned}$$

Note that $\tilde{x}_{k,t}$ is independent of the initial value $\xi_{k,0}$. The variance is then given by

$$\begin{aligned} \text{Var}(\tilde{x}_{k,t}) &= \text{Var}\left(\eta_{k,t} - \frac{1}{M} \sum_{m=1}^M \eta_{k,m} + \sum_{s=1}^t \varepsilon_{k,s} - \sum_{m=1}^M \frac{M-m+1}{M} \varepsilon_{k,m}\right) = \\ &= \text{Var}\left(\eta_{k,t} - \frac{1}{M} \sum_{m=1}^M \eta_{k,m}\right) + \text{Var}\left(\sum_{s=1}^t \varepsilon_{k,s} - \sum_{m=1}^M \frac{M-m+1}{M} \varepsilon_{k,m}\right) \\ &\quad - 2\text{Cov}\left(\eta_{k,t} - \frac{1}{M} \sum_{m=1}^M \eta_{k,m}, \sum_{s=1}^t \varepsilon_{k,s} - \sum_{m=1}^M \frac{M-m+1}{M} \varepsilon_{k,m}\right) = \\ &= I + II + III. \end{aligned}$$

I consider the three parts separately.

$$\begin{aligned} I &= \text{Var}\left(\eta_{k,t} - \frac{1}{M} \sum_{m=1}^M \eta_{k,m}\right) = \text{Var}(\eta_{k,t}) + \text{Var}\left(\frac{1}{M} \sum_{m=1}^M \eta_{k,m}\right) - 2\text{Cov}\left(\eta_{k,t}, \frac{1}{M} \sum_{m=1}^M \eta_{k,m}\right) = \\ &= \sigma_\eta^2 + \frac{M}{M^2} \sigma_\eta^2 - 2 \frac{1}{M} \sum_{m=1}^M \text{Cov}(\eta_{k,t}, \eta_{k,m}) = \frac{M-1}{M} \sigma_\eta^2, \end{aligned}$$

where the last equality follows from the fact that $\text{Cov}(\eta_{k,t}, \eta_{k,m}) = \sigma_\eta^2$ if $m = t$, and $\text{Cov}(\eta_{k,t}, \eta_{k,m}) = 0$ otherwise.

$$\begin{aligned} II &= \text{Var}\left(\sum_{s=1}^t \varepsilon_{k,s} - \sum_{m=1}^M \frac{M-m+1}{M} \varepsilon_{k,m}\right) = \\ &= \text{Var}\left(\sum_{s=1}^t \varepsilon_{k,s}\right) + \text{Var}\left(\sum_{m=1}^M \frac{M-m+1}{M} \varepsilon_{k,m}\right) - 2\text{Cov}\left(\sum_{s=1}^t \varepsilon_{k,s}, \sum_{m=1}^M \frac{M-m+1}{M} \varepsilon_{k,m}\right) = \\ &= t\sigma_\varepsilon^2 + \sum_{m=1}^M \left(\frac{M-m+1}{M}\right)^2 \sigma_\varepsilon^2 - 2 \sum_{s=1}^t \text{Cov}\left(\varepsilon_{k,s}, \sum_{m=1}^M \frac{M-m+1}{M} \varepsilon_{k,m}\right) = \\ &= \left(t + \sum_{m=1}^M \left(\frac{M-m+1}{M}\right)^2 - 2 \sum_{s=1}^t \frac{M-s+1}{M}\right) \sigma_\varepsilon^2, \end{aligned}$$

where the last equality follows from the fact that $\text{Cov}(\varepsilon_{k,s}, \varepsilon_{k,m}) = \sigma_\varepsilon^2$ if $m = s$, and $\text{Cov}(\varepsilon_{k,s}, \varepsilon_{k,m}) = 0$ otherwise. Lastly, since $\text{Cov}(\eta_{k,t}, \varepsilon_{k,s}) = 0$ for all $t, s = 1, 2, \dots, M$, I obtain

$$III = -2Cov \left(\eta_{k,t} - \frac{1}{M} \sum_{m=1}^M \eta_{k,m}, \sum_{s=1}^t \varepsilon_{k,s} - \sum_{m=1}^M \frac{M-m+1}{M} \varepsilon_{k,m} \right) = 0.$$

Then, the variance can be written as

$$\begin{aligned} Var(\tilde{x}_{k,t}) &= A(M)\sigma_\eta^2 + B(M,t)\sigma_\varepsilon^2, \quad \text{where} \\ A(M) &= \frac{M-1}{M} \\ B(M,t) &= t + \sum_{m=1}^M \left(\frac{M-m+1}{M} \right)^2 - 2 \sum_{s=1}^t \frac{M-s+1}{M}. \end{aligned}$$

Furthermore,

$$\begin{aligned} B(M,t) &= t + \sum_{m=1}^M \left(\frac{M-m+1}{M} \right)^2 - 2 \sum_{s=1}^t \frac{M-s+1}{M} = \\ &= t + \sum_{m=1}^M \frac{M^2 + m^2 + 1 - 2mM - 2m + 2M}{M^2} - 2 \left(t - \frac{1}{M} \sum_{s=1}^t s + \frac{t}{M} \right) = \\ &= t + M + \frac{M(M+1)(M+\frac{1}{2})}{3M^2} + \frac{1}{M} - \frac{2M(M+1)}{2M} + 2 - \frac{2M(M+1)}{2M^2} - 2t + \frac{t^2-t}{M} = \\ &= 1 - t + \frac{t^2-t}{M} + \frac{(M+1)(M+\frac{1}{2}) + 3 - 3(M+1)}{3M} = 1 - t + \frac{t^2-t}{M} + \frac{M^2 - \frac{3}{2}M + \frac{1}{2}}{3M}. \end{aligned}$$

A.3 Proof of Proposition 2

Consider $\hat{\beta}_{SFE}$. Using the fact that $\bar{\tilde{x}}_{t-1} = 0$ the estimator can be written as

$$\hat{\beta}_{SFE} = \frac{\sum_{t=1}^T (y_t - \bar{y}) \tilde{x}_{t-1}}{\sum_{t=1}^T \tilde{x}_{t-1}^2} = \frac{\sum_{t=1}^T y_t \tilde{x}_{t-1}}{\sum_{t=1}^T \tilde{x}_{t-1}^2} = \frac{\sum_{k=1}^K \sum_{t=1}^M y_{k,t} \tilde{x}_{k,t-1}}{\sum_{k=1}^K \sum_{t=1}^M \tilde{x}_{k,t-1}^2} = \frac{\sum_{t=1}^M \frac{1}{K} \sum_{k=1}^K y_{k,t} \tilde{x}_{k,t-1}}{\sum_{t=1}^M \frac{1}{K} \sum_{k=1}^K \tilde{x}_{k,t-1}^2}.$$

Due to the i.i.d. assumptions on the error terms both $\{y_{k,t} \tilde{x}_{k,t-1}\}_{k=1}^K$ and $\{\tilde{x}_{k,t-1}^2\}_{k=1}^K$ are i.i.d. sequences with finite variance. Therefore, the Weak Law of Large Numbers applies:

$$\begin{aligned} \frac{1}{K} \sum_{k=1}^K y_{k,t} \tilde{x}_{k,t-1} &\xrightarrow{P} E(y_{k,t} \tilde{x}_{k,t-1}) \\ \frac{1}{K} \sum_{k=1}^K \tilde{x}_{k,t-1}^2 &\xrightarrow{P} E(\tilde{x}_{k,t-1}^2) \end{aligned}$$

Furthermore, the Slutsky theorem implies that

$$\frac{\sum_{t=1}^M \text{plim} \frac{1}{K} \sum_{k=1}^K y_{k,t} \tilde{x}_{k,t-1}}{\sum_{t=1}^M \text{plim} \frac{1}{K} \sum_{k=1}^K \tilde{x}_{k,t-1}^2} = \frac{\text{plim} \sum_{t=1}^M \frac{1}{K} \sum_{k=1}^K y_{k,t} \tilde{x}_{k,t-1}}{\text{plim} \sum_{t=1}^M \frac{1}{K} \sum_{k=1}^K \tilde{x}_{k,t-1}^2} = \text{plim} \frac{\sum_{t=1}^M \frac{1}{K} \sum_{k=1}^K y_{k,t} \tilde{x}_{k,t-1}}{\sum_{t=1}^M \frac{1}{K} \sum_{k=1}^K \tilde{x}_{k,t-1}^2};$$

Therefore,

$$\hat{\beta}_{SFE} \xrightarrow{p} \frac{\sum_{t=1}^M E(y_{k,t} \tilde{x}_{k,t-1})}{\sum_{t=1}^M E(\tilde{x}_{k,t-1}^2)},$$

where

$$\begin{aligned} E(y_{k,t} \tilde{x}_{k,t-1}) &= E((\alpha_0 + \beta_0 \eta_{k,t-1} + u_{k,t}) \tilde{x}_{k,t-1}) = \alpha_0 E(\tilde{x}_{k,t-1}) + \beta_0 E(\eta_{k,t-1} \tilde{x}_{k,t-1}) + E(u_{k,t} \tilde{x}_{k,t-1}) \\ &= \beta_0 E\left(\eta_{k,t-1} \left(\eta_{k,t-1} - \frac{1}{M} \sum_{m=1}^M \eta_{k,m-1} + \sum_{s=1}^t \varepsilon_{k,s-1} - \sum_{m=1}^M \frac{M-m+1}{M} \varepsilon_{k,m-1}\right)\right) \\ &= \beta_0 \left(\sigma_\eta^2 - \frac{1}{M} \sigma_\eta^2 + 0 + 0\right) = \beta_0 \sigma_\eta^2 \frac{M-1}{M} \end{aligned}$$

and

$$E(\tilde{x}_{k,t-1}^2) = \text{Var}(\tilde{x}_{k,t-1}) = \frac{M-1}{M} \sigma_\eta^2 + \left(1-t + \frac{t^2-t}{M} + \frac{M^2 - \frac{3}{2}M + \frac{1}{2}}{3M}\right) \sigma_\varepsilon^2$$

by Lemma 1. Summing over M gives

$$\begin{aligned} \sum_{t=1}^M E(y_{k,t} \tilde{x}_{k,t-1}) &= \sum_{t=1}^M \beta_0 \sigma_\eta^2 \frac{M-1}{M} = \beta_0 \sigma_\eta^2 (M-1) \\ \sum_{t=1}^M E(\tilde{x}_{k,t-1}^2) &= \sum_{t=1}^M \frac{M-1}{M} \sigma_\eta^2 + \sum_{t=1}^M \left(1-t + \frac{t^2-t}{M} + \frac{M^2 - \frac{3}{2}M + \frac{1}{2}}{3M}\right) \sigma_\varepsilon^2 = \\ &= (M-1) \sigma_\eta^2 + \frac{M^2-1}{6} \sigma_\varepsilon^2, \end{aligned}$$

where the last equality holds because

$$\begin{aligned} \sum_{t=1}^M B(M, t) &= \sum_{t=1}^M \left(1 - t + \frac{t^2 - t}{M} + \frac{M^2 - \frac{3}{2}M + \frac{1}{2}}{3M} \right) = \\ &= M - \frac{M(M+1)}{2} + \frac{(M+1)(M+\frac{1}{2})}{3} - \frac{M+1}{2} + \frac{M^2}{3} - \frac{M}{2} + \frac{1}{6} = \\ &= \frac{(-3+2+2)M^2 + (6-3+3-3-3)M + 1-3+1}{6} = \frac{M^2-1}{6} \end{aligned}$$

The result for $\hat{\alpha}_{SFE}$ completes the proof:

$$\hat{\alpha}_{SFE} = \bar{y}_t - \beta_{SFE} \bar{x}_{t-1} = \bar{y}_t \xrightarrow{p} E(y_t) = \alpha_0$$

B Supplementary Tables and Figures

Table B1: Regression results for variables with no predictive power on small samples

	Subsample size (M)			
	10	25	50	OLS
ep	0.0201 (0.0114)*	0.0011 (0.0071)	0.0077 (0.0064)	0.0033 (0.0038)
de	0.0338 (0.0113)***	0.0184 (0.0068)***	0.0055 (0.0057)	0.0037 (0.0053)
dfy	-2.2487 (0.9653)**	-2.2115 (0.5459)***	-1.2741 (0.5074)**	-0.1748 (0.3644)

Notes: The table presents the slope coefficients and standard errors of the univariate predictive regression for predictors with no predictive power even on small samples. This group includes the earnings–price ratio (**ep**), dividend payout ratio (**de**), and default yield spread (**dfy**). The SFE estimator is used. Standard errors are calculated using residual block bootstrapping, where the length of the blocks is $O(T^{1/3})$. The column header specifies the size of the subsamples. ***, **, and * represent statistical significance at the 1%, 5%, and 10% level, respectively.

Figure B1: Monthly time series of the predictors and excess returns

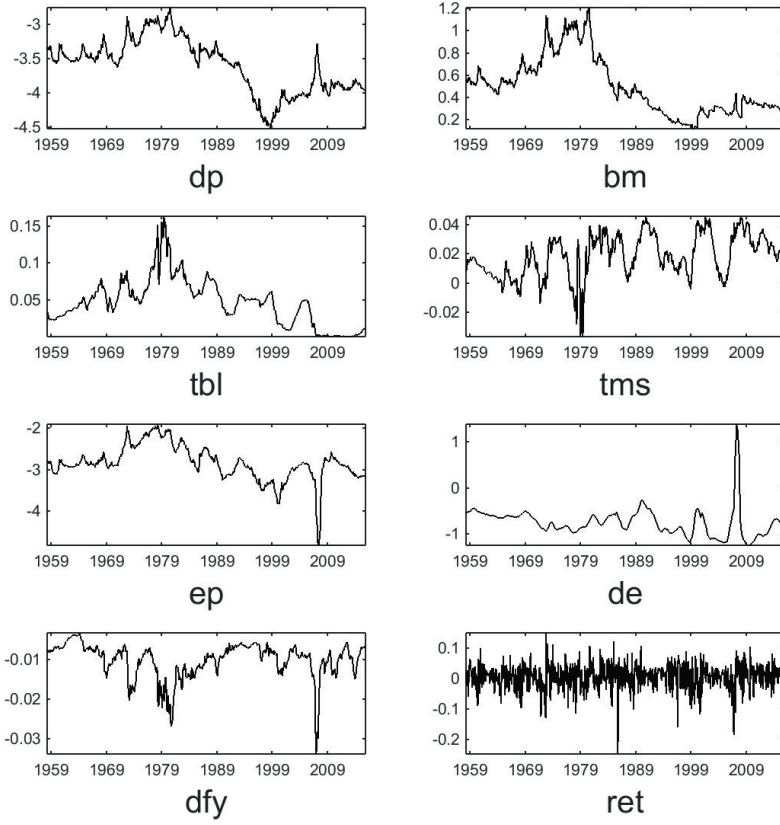
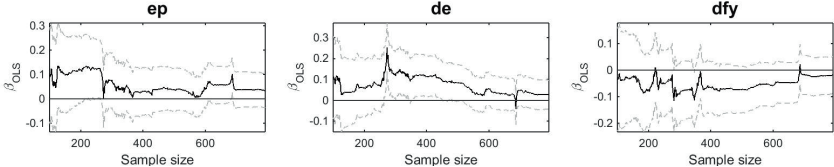
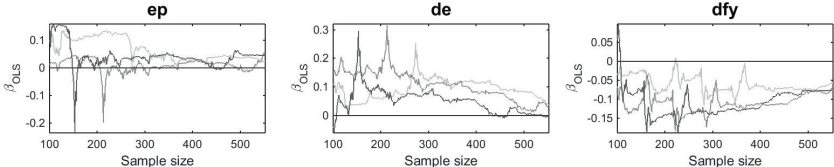


Figure B2: Extending window analysis of variables with no predictive power on small samples



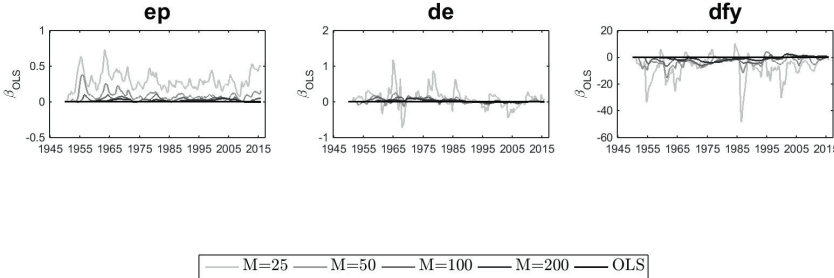
Notes: The figures show the estimated slope coefficients of the univariate predictive regressions as a function of the sample size. All samples start in January 1952. The explanatory variables are the earnings–price ratio (**ep**), dividend payout ratio (**de**), and default yield spread (**dfy**), respectively.

Figure B3: Sensitivity analysis for variables with no predictive power on small samples



Notes: The lines represent the estimated slope coefficients of the univariate predictive regressions as a function of the sample size. Different starting points are used: the lightest grey line represents January 1952, and darker shades mean later starting points (1962 and 1972, respectively). The variable descriptions are given in Figure B2.

Figure B4: Rolling window slope coefficients for variables with no predictive power on small samples



Notes: The lines represent smoothed rolling window estimates of the standard univariate predictive regression using different subsample sizes. That is, $\beta_t^{Smooth} = \omega\beta_{t-1}^{Smooth} + (1 - \omega)\beta_t^{Rolling}$, where $\beta_t^{Rolling}$ is estimated on the rolling window $[t, t - M]$ and $\omega = 0.95$. The variable descriptions are given in Figure B2.

Previous doctoral theses in the Department of Economics, Gothenburg

Avhandlingar publicerade innan serien Ekonomiska Studier startades
(Theses published before the series Ekonomiska Studier was started):

- Östman, Hugo** (1911), Norrlands ekonomiska utveckling
Moritz, Marcus (1911), Den svenska tobaksindustrien
Sundbom, I. (1933), Prusbildning och ändamålsenlighet
Gerhard, I. (1948), Problem rörande Sveriges utrikeshandel 1936/38
Hegeland, Hugo (1951), The Quantity Theory of Money
Mattsson, Bengt (1970), Cost-Benefit analys
Rosengren, Björn (1975), Valutareglering och nationell ekonomisk politik
Hjalmarsson, Lennart (1975), Studies in a Dynamic Theory of Production and its Applications
Örtendahl, Per-Anders (1975), Substitutionsaspekter på produktionsprocessen vid massaframställning
Anderson, Arne M. (1976), Produktion, kapacitet och kostnader vid ett helautomatiskt emballageglasbruk
Ohlsson, Olle (1976), Substitution och odelbarheter i produktionsprocessen vid massaframställning
Gunnarsson, Jan (1976), Produktionssystem och tätortshierarki – om sambandet mellan rumslig och ekonomisk struktur
Köstner, Evert (1976), Optimal allokering av tid mellan utbildning och arbete
Wigren, Rune (1976), Analys av regionala effektivitetsskillnader inom industribranscher
Wästlund, Jan (1976), Skattning och analys av regionala effektivitetsskillnader inom industribranscher
Flöjstads, Gunnar (1976), Studies in Distortions, Trade and Allocation Problems
Sandelin, Bo (1977), Prisutveckling och kapitalvinster på bostadsfastigheter
Dahlberg, Lars (1977), Empirical Studies in Public Planning
Lönnroth, Johan (1977), Marxism som matematisk ekonomi
Johansson, Börje (1978), Contributions to Sequential Analysis of Oligopolistic Competition

Ekonomiska Studier, utgivna av Nationalekonomiska institutionen vid Göteborgs Universitet. Nr 1 och 4 var inte doktorsavhandlingar. (The contributions to the department series 'Ekonomiska Studier' where no. 1 and 4 were no doctoral theses):

2. **Ambjörn, Erik** (1959), Svenskt importberoende 1926-1956: en ekonomisk-statistisk kartläggning med kommentarer
3. **Landgren, K-G.** (1960), Den "Nya ekonomien" i Sverige: J.M. Keynes, E. Wigfors och utvecklingen 1927-39
5. **Bigsten, Arne** (1979), Regional Inequality and Development: A Case Study of Kenya
6. **Andersson, Lars** (1979), Statens styrning av de kommunala budgetarnas struktur (Central Government Influence on the Structure of the Municipal Budget)
7. **Gustafsson, Björn** (1979), Inkomst- och uppväxtförhållanden (Income and Family Background)

8. **Granholt, Arne** (1981), Interregional Planning Models for the Allocation of Private and Public Investments
9. **Lundborg, Per** (1982), Trade Policy and Development: Income Distributional Effects in the Less Developed Countries of the US and EEC Policies for Agricultural Commodities
10. **Juås, Birgitta** (1982), Värdering av risken för personskador. En jämförande studie av implicita och explicita värden. (Valuation of Personal Injuries. A comparison of Explicit and Implicit Values)
11. **Bergendahl, Per-Anders** (1982), Energi och ekonomi - tillämpningar av input-output analys (Energy and the Economy - Applications of Input-Output Analysis)
12. **Blomström, Magnus** (1983), Foreign Investment, Technical Efficiency and Structural Change - Evidence from the Mexican Manufacturing Industry
13. **Larsson, Lars-Göran** (1983), Comparative Statics on the Basis of Optimization Methods
14. **Persson, Håkan** (1983), Theory and Applications of Multisectoral Growth Models
15. **Sterner, Thomas** (1986), Energy Use in Mexican Industry.
16. **Flood, Lennart** (1986), On the Application of Time Use and Expenditure Allocation Models.
17. **Schuller, Bernd-Joachim** (1986), Ekonomi och kriminalitet - en empirisk undersökning av brottsligheten i Sverige (Economics of crime - an empirical analysis of crime in Sweden)
18. **Walfridson, Bo** (1987), Dynamic Models of Factor Demand. An Application to Swedish Industry.
19. **Stålhammar, Nils-Olov** (1987), Strukturomvandling, företagsbeteende och förväntningsbildning inom den svenska tillverkningsindustrin (Structural Change, Firm Behaviour and Expectation Formation in Swedish Manufactory)
20. **Anxo, Dominique** (1988), Sysselsättningseffekter av en allmän arbetstidsförkortning (Employment effects of a general shortage of the working time)
21. **Mbelle, Ammon** (1988), Foreign Exchange and Industrial Development: A Study of Tanzania.
22. **Ongaro, Wilfred** (1988), Adoption of New Farming Technology: A Case Study of Maize Production in Western Kenya.
23. **Zejan, Mario** (1988), Studies in the Behavior of Swedish Multinationals.
24. **Görling, Anders** (1988), Ekonomisk tillväxt och miljö. Föreningens-struktur och ekonomiska effekter av olika miljövårdsprogram. (Economic Growth and Environment. Pollution Structure and Economic Effects of Some Environmental Programs).
25. **Aguiar, Renato** (1988), Efficiency in Production: Theory and an Application on Kenyan Smallholders.
26. **Kayizzi-Mugerwa, Steve** (1988), External Shocks and Adjustment in Zambia.
27. **Bornmalm-Jardelöw, Gunilla** (1988), Högre utbildning och arbetsmarknad (Higher Education and the Labour Market)
28. **Tansini, Ruben** (1989), Technology Transfer: Dairy Industries in Sweden and Uruguay.
29. **Andersson, Irene** (1989), Familjebeskattnings, konsumtion och arbetsutbud - En ekonometrisk analys av löne- och inkomstelasticiteter samt policy simuleringar för

- svenska hushåll (Family Taxation, Consumption and Labour Supply - An Econometric Analysis of Wage and Income Elasticities and Policy Simulations for Swedish Households)
30. **Henrekson, Magnus** (1990), An Economic Analysis of Swedish Government Expenditure
 31. **Sjöo, Boo** (1990), Monetary Policy in a Continuous Time Dynamic Model for Sweden
 32. **Rosén, Åsa** (1991), Contributions to the Theory of Labour Contracts.
 33. **Loureiro, Joao M. de Matos** (1992), Foreign Exchange Intervention, Sterilization and Credibility in the EMS: An Empirical Study
 34. **Irandoust, Manuchehr** (1993), Essays on the Behavior and Performance of the Car Industry
 35. **Tasiran, Ali Cevat** (1993), Wage and Income Effects on the Timing and Spacing of Births in Sweden and the United States
 36. **Milopoulos, Christos** (1993), Investment Behaviour under Uncertainty: An Econometric Analysis of Swedish Panel Data
 37. **Andersson, Per-Åke** (1993), Labour Market Structure in a Controlled Economy: The Case of Zambia
 38. **Storrie, Donald W.** (1993), The Anatomy of a Large Swedish Plant Closure
 39. **Semboja, Haji Hatibu Haji** (1993), Energy and Development in Kenya
 40. **Makonnen, Negatu** (1993), Labor Supply and the Distribution of Economic Well-Being: A Case Study of Lesotho
 41. **Julin, Eva** (1993), Structural Change in Rural Kenya
 42. **Durevall, Dick** (1993), Essays on Chronic Inflation: The Brazilian Experience
 43. **Veiderpass, Ann** (1993), Swedish Retail Electricity Distribution: A Non-Parametric Approach to Efficiency and Productivity Change
 44. **Odeck, James** (1993), Measuring Productivity Growth and Efficiency with Data Envelopment Analysis: An Application on the Norwegian Road Sector
 45. **Mwenda, Abraham** (1993), Credit Rationing and Investment Behaviour under Market Imperfections: Evidence from Commercial Agriculture in Zambia
 46. **Mlambo, Kupukile** (1993), Total Factor Productivity Growth: An Empirical Analysis of Zimbabwe's Manufacturing Sector Based on Factor Demand Modelling
 47. **Ndung'u, Njuguna** (1993), Dynamics of the Inflationary Process in Kenya
 48. **Modén, Karl-Markus** (1993), Tax Incentives of Corporate Mergers and Foreign Direct Investments
 49. **Franzén, Mikael** (1994), Gasoline Demand - A Comparison of Models
 50. **Heshmati, Almas** (1994), Estimating Technical Efficiency, Productivity Growth And Selectivity Bias Using Rotating Panel Data: An Application to Swedish Agriculture
 51. **Salas, Osvaldo** (1994), Efficiency and Productivity Change: A Micro Data Case Study of the Colombian Cement Industry
 52. **Bjurek, Hans** (1994), Essays on Efficiency and Productivity Change with Applications to Public Service Production
 53. **Cabezas Vega, Luis** (1994), Factor Substitution, Capacity Utilization and Total Factor Productivity Growth in the Peruvian Manufacturing Industry
 54. **Katz, Katarina** (1994), Gender Differentiation and Discrimination. A Study of

- Soviet Wages
55. **Asal, Maher** (1995), Real Exchange Rate Determination and the Adjustment Process: An Empirical Study in the Cases of Sweden and Egypt
 56. **Kjulin, Urban** (1995), Economic Perspectives on Child Care
 57. **Andersson, Göran** (1995), Volatility Forecasting and Efficiency of the Swedish Call Options Market
 58. **Forteza, Alvaro** (1996), Credibility, Inflation and Incentive Distortions in the Welfare State
 59. **Locking, Håkan** (1996), Essays on Swedish Wage Formation
 60. **Välilä, Timo** (1996), Essays on the Credibility of Central Bank Independence
 61. **Yilma, Mulugeta** (1996), Measuring Smallholder Efficiency: Ugandan Coffee and Food-Crop Production
 62. **Mabugu, Ramos E.** (1996), Tax Policy Analysis in Zimbabwe Applying General Equilibrium Models
 63. **Johansson, Olof** (1996), Welfare, Externalities, and Taxation; Theory and Some Road Transport Applications.
 64. **Chitiga, Margaret** (1996), Computable General Equilibrium Analysis of Income Distribution Policies in Zimbabwe
 65. **Leander, Per** (1996), Foreign Exchange Market Behavior Expectations and Chaos
 66. **Hansen, Jörgen** (1997), Essays on Earnings and Labor Supply
 67. **Cotfas, Mihai** (1997), Essays on Productivity and Efficiency in the Romanian Cement Industry
 68. **Horghy, Per-Johan** (1997), Essays on Sharing, Management and Evaluation of Health Risks
 69. **Nafar, Nosratollah** (1997), Efficiency and Productivity in Iranian Manufacturing Industries
 70. **Zheng, Jinghai** (1997), Essays on Industrial Structure, Technical Change, Employment Adjustment, and Technical Efficiency
 71. **Isaksson, Anders** (1997), Essays on Financial Liberalisation in Developing Countries: Capital mobility, price stability, and savings
 72. **Gerdin, Anders** (1997), On Productivity and Growth in Kenya, 1964-94
 73. **Sharifi, Alimorad** (1998), The Electricity Supply Industry in Iran: Organization, performance and future development
 74. **Zamanian, Max** (1997), Methods for Mutual Fund Portfolio Evaluation: An application to the Swedish market
 75. **Manda, Damiano Kulundu** (1997), Labour Supply, Returns to Education, and the Effect of Firm Size on Wages: The case of Kenya
 76. **Holmén, Martin** (1998), Essays on Corporate Acquisitions and Stock Market Introductions
 77. **Pan, Kelvin** (1998), Essays on Enforcement in Money and Banking
 78. **Rogat, Jorge** (1998), The Value of Improved Air Quality in Santiago de Chile
 79. **Peterson, Stefan** (1998), Essays on Large Shareholders and Corporate Control
 80. **Belhaj, Mohammed** (1998), Energy, Transportation and Urban Environment in Africa: The Case of Rabat-Salé, Morocco
 81. **Mekonnen, Alemu** (1998), Rural Energy and Afforestation: Case Studies from Ethiopia

82. **Johansson, Anders** (1998), Empirical Essays on Financial and Real Investment Behavior
83. **Köhlin, Gunnar** (1998), The Value of Social Forestry in Orissa, India
84. **Levin, Jörgen** (1998), Structural Adjustment and Poverty: The Case of Kenya
85. **Ncube, Mkhululi** (1998), Analysis of Employment Behaviour in Zimbabwe
86. **Mwansa, Ladslous** (1998), Determinants of Inflation in Zambia
87. **Agnarsson, Sveinn** (1998), Of Men and Machines: Essays in Applied Labour and Production Economics
88. **Kadenge, Phineas** (1998), Essays on Macroeconomic Adjustment in Zimbabwe: Inflation, Money Demand, and the Real Exchange Rate
89. **Nyman, Håkan** (1998), An Economic Analysis of Lone Motherhood in Sweden
90. **Carlsson, Fredrik** (1999), Essays on Externalities and Transport
91. **Johansson, Mats** (1999), Empirical Studies of Income Distribution
92. **Alemu, Tekie** (1999), Land Tenure and Soil Conservation: Evidence from Ethiopia
93. **Lundvall, Karl** (1999), Essays on Manufacturing Production in a Developing Economy: Kenya 1992-94
94. **Zhang, Jianhua** (1999), Essays on Emerging Market Finance
95. **Mlima, Aziz Ponary** (1999), Four Essays on Efficiency and Productivity in Swedish Banking
96. **Dauidsen, Björn-Ivar** (2000), Bidrag til den økonomisk-metodologiske tenkningen (Contributions to the Economic Methodological Thinking)
97. **Ericson, Peter** (2000), Essays on Labor Supply
98. **Söderbom, Måns** (2000), Investment in African Manufacturing: A Microeconomic Analysis
99. **Höglund, Lena** (2000), Essays on Environmental Regulation with Applications to Sweden
100. **Olsson, Ola** (2000), Perspectives on Knowledge and Growth
101. **Meuller, Lars** (2000), Essays on Money and Credit
102. **Österberg, Torun** (2000), Economic Perspectives on Immigrants and Intergenerational Transmissions
103. **Kalinda Mkenda, Beatrice** (2001), Essays on Purchasing Power Parity, Real Exchange Rate, and Optimum Currency Areas
104. **Nerhagen, Lena** (2001), Travel Demand and Value of Time - Towards an Understanding of Individuals Choice Behavior
105. **Mkenda, Adolf** (2001), Fishery Resources and Welfare in Rural Zanzibar
106. **Eggert, Håkan** (2001), Essays on Fisheries Economics
107. **Andrén, Daniela** (2001), Work, Sickness, Earnings, and Early Exits from the Labor Market. An Empirical Analysis Using Swedish Longitudinal Data
108. **Nivorozhkin, Eugene** (2001), Essays on Capital Structure
109. **Hammar, Henrik** (2001), Essays on Policy Instruments: Applications to Smoking and the Environment
110. **Nannyonjo, Justine** (2002), Financial Sector Reforms in Uganda (1990-2000): Interest Rate Spreads, Market Structure, Bank Performance and Monetary Policy
111. **Wu, Hong** (2002), Essays on Insurance Economics

112. **Linde-Rahr, Martin** (2002), Household Economics of Agriculture and Forestry in Rural Vietnam
113. **Maneschiöld, Per-Ola** (2002), Essays on Exchange Rates and Central Bank Credibility
114. **Andrén, Thomas** (2002), Essays on Training, Welfare and Labor Supply
115. **Granér, Mats** (2002), Essays on Trade and Productivity: Case Studies of Manufacturing in Chile and Kenya
116. **Jaldell, Henrik** (2002), Essays on the Performance of Fire and Rescue Services
117. **Alpizar, Francisco, R.** (2002), Essays on Environmental Policy-Making in Developing Countries: Applications to Costa Rica
118. **Wahlberg, Roger** (2002), Essays on Discrimination, Welfare and Labor Supply
119. **Piculescu, Violeta** (2002), Studies on the Post-Communist Transition
120. **Pylkkänen, Elina** (2003), Studies on Household Labor Supply and Home Production
121. **Löfgren, Åsa** (2003), Environmental Taxation – Empirical and Theoretical Applications
122. **Ivaschenko, Oleksiy** (2003), Essays on Poverty, Income Inequality and Health in Transition Economies
123. **Lundström, Susanna** (2003), On Institutions, Economic Growth and the Environment
124. **Wambugu, Anthony** (2003), Essays on Earnings and Human Capital in Kenya
125. **Adler, Johan** (2003), Aspects of Macroeconomic Saving
126. **Erlandsson, Mattias** (2003), On Monetary Integration and Macroeconomic Policy
127. **Brink, Anna** (2003), On the Political Economy of Municipality Break-Ups
128. **Ljungwall, Christer** (2003), Essays on China's Economic Performance During the Reform Period
129. **Chifamba, Ronald** (2003), Analysis of Mining Investments in Zimbabwe
130. **Muchapondwa, Edwin** (2003), The Economics of Community-Based Wildlife Conservation in Zimbabwe
131. **Hammes, Klaus** (2003), Essays on Capital Structure and Trade Financing
132. **Abou-Ali, Hala** (2003), Water and Health in Egypt: An Empirical Analysis
133. **Simatele, Munacinga** (2004), Financial Sector Reforms and Monetary Policy in Zambia
134. **Tezic, Kerem** (2004), Essays on Immigrants' Economic Integration
135. INSTÄLLD
136. **Gjirja, Matilda** (2004), Efficiency and Productivity in Swedish Banking
137. **Andersson, Jessica** (2004), Welfare Environment and Tourism in Developing Countries
138. **Chen, Yinghong** (2004), Essays on Voting Power, Corporate Governance and Capital Structure
139. **Yesuf, Mahmud** (2004), Risk, Time and Land Management under Market Imperfections: Applications to Ethiopia
140. **Kateregga, Eseza** (2005), Essays on the Infestation of Lake Victoria by the Water Hyacinth
141. **Edvardsen, Dag Fjeld** (2004), Four Essays on the Measurement of Productive Efficiency

142. **Lidén, Erik** (2005), Essays on Information and Conflicts of Interest in Stock Recommendations
143. **Dieden, Sten** (2005), Income Generation in the African and Coloured Population – Three Essays on the Origins of Household Incomes in South Africa
144. **Eliasson, Marcus** (2005), Individual and Family Consequences of Involuntary Job Loss
145. **Mahmud, Minhaj** (2005), Measuring Trust and the Value of Statistical Lives: Evidence from Bangladesh
146. **Lokina, Razack Bakari** (2005), Efficiency, Risk and Regulation Compliance: Applications to Lake Victoria Fisheries in Tanzania
147. **Jussila Hammes, Johanna** (2005), Essays on the Political Economy of Land Use Change
148. **Nyangena, Wilfred** (2006), Essays on Soil Conservation, Social Capital and Technology Adoption
149. **Nivorozhkin, Anton** (2006), Essays on Unemployment Duration and Programme Evaluation
150. **Sandén, Klas** (2006), Essays on the Skill Premium
151. **Deng, Daniel** (2006), Three Essays on Electricity Spot and Financial Derivative Prices at the Nordic Power Exchange
152. **Gebreeyesus, Mulu** (2006), Essays on Firm Turnover, Growth, and Investment Behavior in Ethiopian Manufacturing
153. **Islam, Nizamul Md.** (2006), Essays on Labor Supply and Poverty: A Microeconomic Application
154. **Kjaer, Mats** (2006), Pricing of Some Path-Dependent Options on Equities and Commodities
155. **Shimeles, Abebe** (2006), Essays on Poverty, Risk and Consumption Dynamics in Ethiopia
156. **Larsson, Jan** (2006), Four Essays on Technology, Productivity and Environment
157. **Congdon Fors, Heather** (2006), Essays in Institutional and Development Economics
158. **Akpalu, Wisdom** (2006), Essays on Economics of Natural Resource Management and Experiments
159. **Daruvala, Dinky** (2006), Experimental Studies on Risk, Inequality and Relative Standing
160. **García, Jorge** (2007), Essays on Asymmetric Information and Environmental Regulation through Disclosure
161. **Bezabih, Mintewab** (2007), Essays on Land Lease Markets, Productivity, Biodiversity, and Environmental Variability
162. **Visser, Martine** (2007), Fairness, Reciprocity and Inequality: Experimental Evidence from South Africa
163. **Holm, Louise** (2007), A Non-Stationary Perspective on the European and Swedish Business Cycle
164. **Herbertsson, Alexander** (2007), Pricing Portfolio Credit Derivatives
165. **Johansson, Anders C.** (2007), Essays in Empirical Finance: Volatility, Interdependencies, and Risk in Emerging Markets
166. **Ibáñez Díaz, Marcela** (2007), Social Dilemmas: The Role of Incentives, Norms and Institutions

167. **Ekbom, Anders** (2007), Economic Analysis of Soil Capital, Land Use and Agricultural Production in Kenya
168. **Sjöberg, Pål** (2007), Essays on Performance and Growth in Swedish Banking
169. **Palma Aguirre, Grisha Alexis** (2008), Explaining Earnings and Income Inequality in Chile
170. **Akay, Alpaslan** (2008), Essays on Microeconometrics and Immigrant Assimilation
171. **Carlsson, Evert** (2008), After Work – Investing for Retirement
172. **Munshi, Farzana** (2008), Essays on Globalization and Occupational Wages
173. **Tsakas, Elias** (2008), Essays on Epistemology and Evolutionary Game Theory
174. **Erlandzon, Karl** (2008), Retirement Planning: Portfolio Choice for Long-Term Investors
175. **Lampi, Elina** (2008), Individual Preferences, Choices, and Risk Perceptions – Survey Based Evidence
176. **Mitrut, Andreea** (2008), Four Essays on Interhousehold Transfers and Institutions in Post-Communist Romania
177. **Hansson, Gustav** (2008), Essays on Social Distance, Institutions, and Economic Growth
178. **Zikhali, Precious** (2008), Land Reform, Trust and Natural Resource Management in Africa
179. **Tengstam, Sven** (2008), Essays on Smallholder Diversification, Industry Location, Debt Relief, and Disability and Utility
180. **Boman, Anders** (2009), Geographic Labour Mobility – Causes and Consequences
181. **Qin, Ping** (2009), Risk, Relative Standing and Property Rights: Rural Household Decision-Making in China
182. **Wei, Jiegen** (2009), Essays in Climate Change and Forest Management
183. **Belu, Constantin** (2009), Essays on Efficiency Measurement and Corporate Social Responsibility
184. **Ahlerup, Pelle** (2009), Essays on Conflict, Institutions, and Ethnic Diversity
185. **Quiroga, Miguel** (2009), Microeconomic Policy for Development: Essays on Trade and Environment, Poverty and Education
186. **Zerfu, Daniel** (2010), Essays on Institutions and Economic Outcomes
187. **Wollbrant, Conny** (2010), Self-Control and Altruism
188. **Villegas Palacio, Clara** (2010), Formal and Informal Regulations: Enforcement and Compliance
189. **Maican, Florin** (2010), Essays in Industry Dynamics on Imperfectly Competitive Markets
190. **Jakobsson, Niklas** (2010), Laws, Attitudes and Public Policy
191. **Manescu, Cristiana** (2010), Economic Implications of Corporate Social Responsibility and Responsible Investments
192. **He, Haoran** (2010), Environmental and Behavioral Economics – Applications to China
193. **Andersson, Fredrik W.** (2011), Essays on Social Comparison
194. **Isaksson, Ann-Sofie** (2011), Essays on Institutions, Inequality and Development
195. **Pham, Khanh Nam** (2011), Prosocial Behavior, Social Interaction and Development: Experimental Evidence from Vietnam

196. **Lindskog, Annika** (2011), Essays on Economic Behaviour: HIV/AIDS, Schooling, and Inequality
197. **Kotsadam, Andreas** (2011), Gender, Work, and Attitudes
198. **Alem, Yonas** (2011), Essays on Shocks, Welfare, and Poverty Dynamics: Microeconomic Evidence from Ethiopia
199. **Köksal-Ayhan, Miyase Yesim** (2011), Parallel Trade, Reference Pricing and Competition in the Pharmaceutical Market: Theory and Evidence
200. **Vondolia, Godwin Kofi** (2011), Essays on Natural Resource Economics
201. **Widerberg, Anna** (2011), Essays on Energy and Climate Policy – Green Certificates, Emissions Trading and Electricity Prices
202. **Siba, Eyerusalem** (2011), Essays on Industrial Development and Political Economy of Africa
203. **Orth, Matilda** (2012), Entry, Competition and Productivity in Retail
204. **Nerman, Måns** (2012), Essays on Development: Household Income, Education, and Female Participation and Representation
205. **Wicks, Rick** (2012), The Place of Conventional Economics in a World with Communities and Social Goods
206. **Sato, Yoshihiro** (2012), Dynamic Investment Models, Employment Generation and Productivity – Evidence from Swedish Data
207. **Valsecchi, Michele** (2012), Essays in Political Economy of Development
208. **Teklewold Belayneh, Hailemariam** (2012), Essays on the Economics of Sustainable Agricultural Technologies in Ethiopia
209. **Wagura Ndiritu, Simon** (2013), Essays on Gender Issues, Food Security, and Technology Adoption in East Africa
210. **Ruist, Joakim** (2013), Immigration, Work, and Welfare
211. **Nordén, Anna** (2013), Essays on Behavioral Economics and Policies for Provision of Ecosystem Services
212. **Yang, Xiaojun** (2013), Household Decision Making, Time Preferences, and Positional Concern: Experimental Evidence from Rural China
213. **Bonilla Londoño, Jorge Alexander** (2013), Essays on the Economics of Air Quality Control
214. **Mohlin, Kristina** (2013), Essays on Environmental Taxation and Climate Policy
215. **Medhin, Haileselassie** (2013), The Poor and Their Neighbors: Essays on Behavioral and Experimental Economics
216. **Andersson, Lisa** (2013), Essays on Development and Experimental Economics: Migration, Discrimination and Positional Concerns
217. **Weng, Qian** (2014), Essays on Team Cooperation and Firm Performance
218. **Zhang, Xiao-Bing** (2015), Cooperation and Paradoxes in Climate Economics
219. **Jaime Torres, Monica Marcela** (2015) Essays on Behavioral Economics and Policy Design
220. **Bejenariu, Simona** (2015) Determinants of Health Capital at Birth: Evidence from Policy Interventions
221. **Nguyen, Van Diem** (2015) Essays on Takeovers and Executive Compensation
222. **Tolonen, Anja** (2015) Mining Booms in Africa and Local Welfare Effects: Labor Markets, Women’s Empowerment and Criminality
223. **Hassen, Sied** (2015) On the Adoption and Dis-adoption of Household Energy and

Farm Technologies

224. **Moursli, Mohamed-Reda** (2015) Corporate Governance and the Design of Board of Directors
225. **Borcan, Oana** (2015) Economic Determinants and Consequences of Political Institutions
226. **Ruhinduka, Remidius Denis** (2015) Essays on Field Experiments and Impact Evaluation
227. **Persson, Emil** (2016) Essays on Behavioral and Experimental Economics, Cooperation, Emotions and Health.
228. **Martinangeli, Andrea** (2017) Bitter divisions: inequality, identity and cooperation
229. **Björk, Lisa** (2017) Essays on Behavioral Economics and Fisheries: Coordination and Cooperation
230. **Chegere, Martin Julius** (2017) Post-Harvest Losses, Intimate Partner Violence and Food Security in Tanzania
231. **Villalobos-Fiatt, Laura** (2017) Essays on forest conservation policies, weather and school attendance
232. **Yi, Yuanyuan** (2017) Incentives and Forest Reform: Evidence from China
233. **Kurz, Verena** (2017) Essays on behavioral economics: Nudges, food consumption and procedural fairness
234. **Yashodha, Yashodha** (2017) Contract Choice and Trust in Informal Groundwater Markets
235. **Meles, Tensay Hadus** (2017) Power Outages, Increasing Block Tariffs and Billing Knowledge
236. **Mühlrad, Hanna** (2018) The Impact of Reproductive and Birth Technologies on Health, Fertility and Labor Outcomes
237. **Gakii Gatua, Josephine** (2018) Primary Health Care Interventions and Social Ties in Kenya
238. **Felgendreher, Simon** (2018) Essays on behavioral economics and the effects of the colonial rule on Java
239. **Demeke, Eyoual** (2019) Essays on Environmental and Behavioral Economics
240. **Kiss, Tamás** (2019) Predictability in Equity Markets: Estimation and Inference