Comput Econ (2007) 29:369–381
DOI 10.1007/s10614-006-9064-0

Strategic asset allocation and market timing: a reinforcement learning approach

Thorsten Hens · Peter Wöhrmann

Received: 5 October 2006 / Accepted: 12 October 2006 / Published online: 23 March 2007

Abstract We apply the recurrent reinforcement learning method of Moody, Wu, Liao, and Saffell (1998) in the context of the strategic asset allocation computed for sample data from US, UK, Germany, and Japan. It is found that the optimal asset allocation deviates substantially from the fixed-mix rule. The investor actively times the market and he is able to outperform it consistently over the almost two decades we analyze.

Keywords Dynamic asset allocation · Bond/equity ratio · Reinforcement Learning

1 Introduction

It is well known that a rationally planning investor with constant relative risk aversion (CRRA) will choose a fixed-mix asset allocation if the investment opportuni-

T. Hens (✉)
Swiss Banking Institute, University of Zürich,
Plattenstrasse 32, 8032
Zürich, Switzerland
e-mail: thens@isb.unizh.ch

T. Hens
Norwegian School of Economics and Business Administration,
Helleveien 30, 5045
Bergen, Norway

P. Wöhrmann
Department of Management Science and Engineering,
Stanford University,
Stanford, CA, USA
e-mail: pwoehrma@isb.unizh.ch

ties do not change over time (see, e.g., Merton, 1969; Samuelson, 1969; Campbell & Viceira, 2002). Whether financial markets do however offer constant or rather changing investment opportunities to which a tactically planning investor can react is an ongoing debate that is of central importance to any asset manager. While Malkiel (2003) and Fama (1998), for example, strongly object to the view that markets over- and under-react in a predictable way, researchers like Campbell (2000) and Shleifer (2000) and many others, in particular those in the field of behavioral finance, point out many anomalies giving rise to predictability of market returns. De Bondt and Thaler (1985), Daniel, Hirshleifer, and Subrahmanyam (1998) and Campbell, Andrew, and Craig Mc Kinley (1999), for example, suggest that stock markets are mean-reverting so that a rationally planning investor should deviate from the fixed-mix rule and try to actively time the market. Campbell and Robert (1988) claim that the dividend price ratio is a predictor for future stock returns. The intuition behind this approach stems from the present value relation and says that when the price dividend ratio is high, either expected returns must be low or expected dividend growth rates must be high.

Stock market predictability does not necessarily contradict with market efficiency when agents discount future dividends at time varying rates, i.e., when the intertemporal rate of substitution is time-varying. The most obvious implication of predictability for portfolio advice is that market timing strategies can be exploited. Gallant, Hansen, and Tauchen (1990) measure potential benefits of market timing without considering concrete timing strategies by assessing the volatility of the marginal rate of intertemporal substitution using conditional moment tests. Cochrane (1991) conjectures that market timing should raise average returns by about two-fifths at an annual horizon and it should almost double average returns at a 5-year horizon. Campbell and Viceira (2005) suggest that strategic investors like pension funds should benefit from market timing. Brandt (1999) estimates a market timing rule within a dynamic optimization framework based on the GMM Euler equation approach without making distributional assumptions on the underlying variables. Analytical and numerical solutions to the dymamic planning problem usually impose assumptions on the underlying variables. For example, Campbell and Viceira (1999) assume log-normality of stock returns and Brennan, Schwartz, and Lagnado (1997) require stock returns to follow a Markov process.

In this work, we aim at determining the timing rule by explicitly solving the dynamic optimization problem without distributional assumptions employing the direct version of the Reinforcement Learning technique which has been introduced to economists by Moody, Wu, Liao, and Saffell (1998). Therefore, we solve the problem with a non-linear feedback rule numerically. In contrast to traditional numerical solution techniques such as value iteration, this approach allows for simpler problem representation, avoids Bellman's curse of dimensionality and is more computationally efficient. The approach is called "direct" since it does not attempt to estimate the value function, but it rather recovers the optimal (policy) feedback rule.

We consider actual returns of bonds and stock market indices in the US, UK Germany, and Japan, and compute the optimal asset allocation of a stra-

tegic investor with constant relative risk aversion. Applying the recurrent reinforcement learning method of Moody et al. (1998) we find that the optimal asset allocation on that data is far from being fixed-mix but it actively times the market. This result is not only of theoretical importance since it contributes to the above-mentioned debate but it also has important practical applications. In countries like the US and Switzerland, for example, a substantial part of the society's wealth is held in pension funds because the pension fund system is capital based. The most important issue of such pension funds is to find the optimal bond/equity ratio for the strategic asset allocation. Our paper thus encourages these pension funds to deviate from the passive fixed-mix asset allocation rules that are applied in most of them.

In the next section, we describe our model which will then be solved numerically in Sect. 3 before Sect. 4 concludes.

2 Strategic asset allocation with market timing

The investor has a discrete time planning horizon $t = 0, 1, \ldots, T$. The investable universe of assets is restricted to long-term bonds and equity. Bond and equity returns are denoted by R_t^{B}, and R_t^{E}, $t = 0, 1, \ldots, T$, respectively. The investor has to choose the weight of bonds in any period t, $\omega_t \in [0; 1]$, which implies that the weight of equity is $1 - \omega_t$. We allow the investor to believe that he can predict the yield spread between bonds and equity based on historical returns on bonds and equity. This predictability is modeled in the strategic asset allocation problems (1)–(4) by an adaptive policy function parameterized in θ determining time-varying weights conditional on the actual yields,

$$\omega_t = \left(\theta/1.15 + \frac{1}{\exp(F_3 + F_2\theta(R_{t-1}^{\mathrm{B}} - R_{t-1}^{\mathrm{E}}))} - 0.5 \right) F_1^{-1},\ F_1, F_2, F_3 \in \mathbb{R}$$
$$t = 0, 1, \ldots, T. \qquad (1)$$

Weights ω_t are restricted to the interval $[0; 1]$. Note that the degree of predictability is endogenously determined by the numerical value of θ that is learnt from the data by application of the reinforcement learning algorithm. While the particular form of the policy function is certainly ad hoc, it has some important qualitative properties. The sigmoid chosen in this paper restricts the bond weight to the interval [0,1]. Moreover, the sign of the slope is not restricted and the degree of concavity can be varied up to linearity of the function (see Fig. 1 for alternative shapes of the timing function).

The portfolio return, $R_t^P \in \mathbb{R}$, is given by

$$R_t^P = \omega_t R_t^{\mathrm{B}} + (1 - \omega_t) R_t^{\mathrm{E}}. \qquad (2)$$

The wealth of the investor, $W_t \in \mathbb{R}^+$, $t = 0, 1, \ldots, T$, measured in monetary units, starts with $W_0 \in \mathbb{R}$. Wealth evolves along the equation

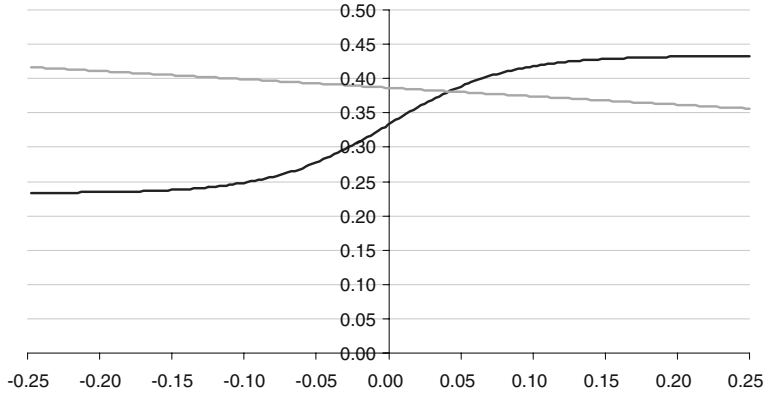

Fig. 1 Sigmoid timing function (1). The graph represents the amount of the bond weight (vertical axis) depending on the return spread of bonds and equity (horizontal axis) as determined by the market timing function (1). The black and grey line refer to $\theta = -0.5, F_1 = 5, F_2 = 50, F_3 = 1$, and $\theta = 0.25, F_1 = -5, F_2 = -10$, respectively.

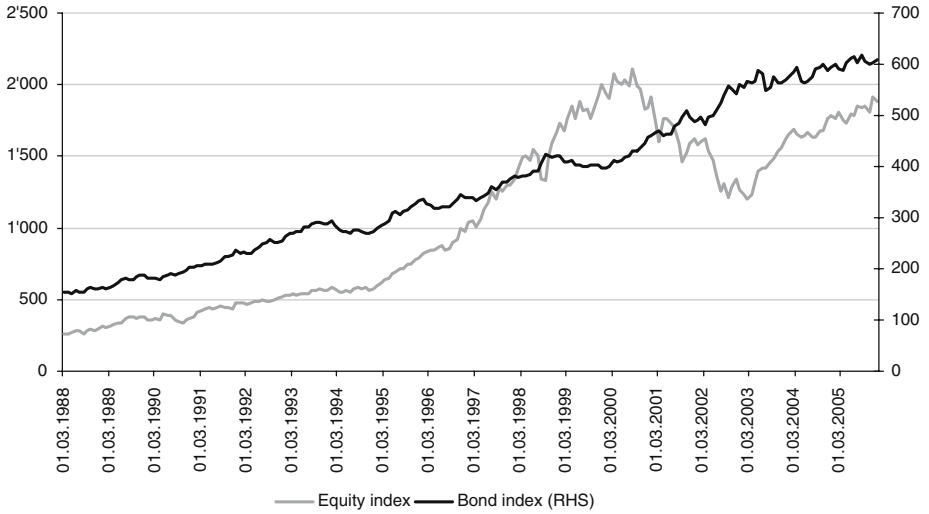

Fig. 2 Bond and equity index in US

$$W_{t+1} = W_t(1 + R_t^P). \tag{3}$$

We consider a totally rational investor optimizing a power utility U arising from terminal wealth,

$$U_T(\theta) = \frac{W_T^{1-\gamma}}{1-\gamma}.^1 \tag{4}$$

[1] Performing a second-order Taylor series approximation to U, the utility depends on expected returns and squared returns viewed as risk.

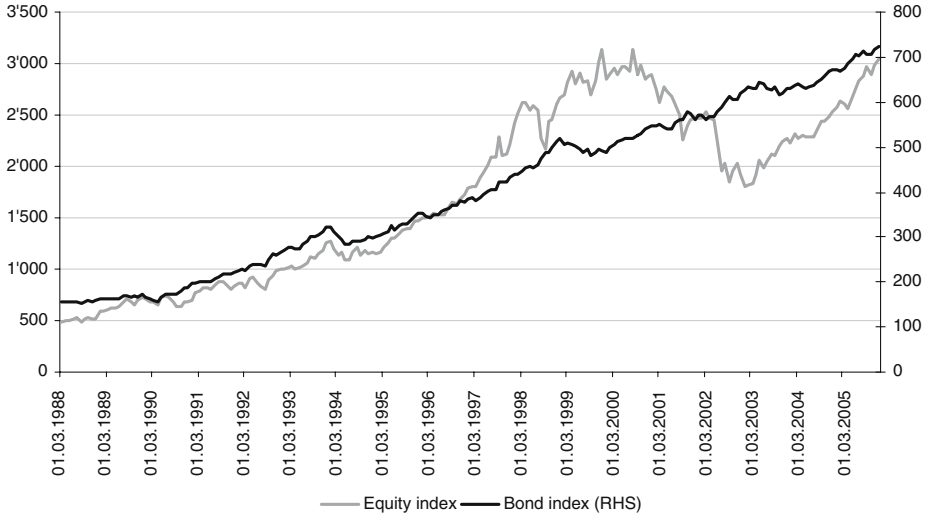

Fig. 3 Bond and equity index in UK

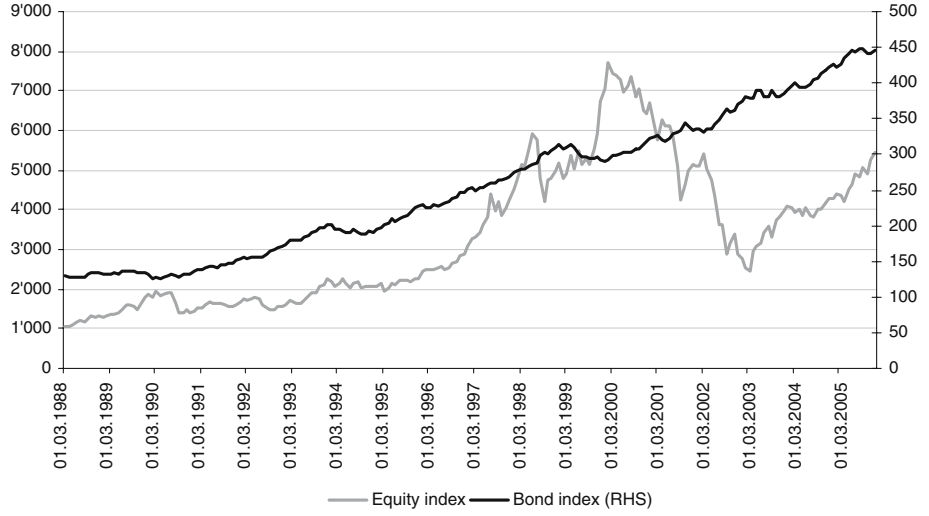

Fig. 4 Bond and equity index in Germany

Note that on a constant opportunity set, as it is for example delivered by a random walk, such an investor would choose a fixed-mix asset allocation. That is, he would hold the mix of bonds and equities fixed over time.

In the subsequent section, we solve the optimization problems (1)–(4) numerically based on monthly data on bond and equity returns in US, UK, Japan, and Germany.

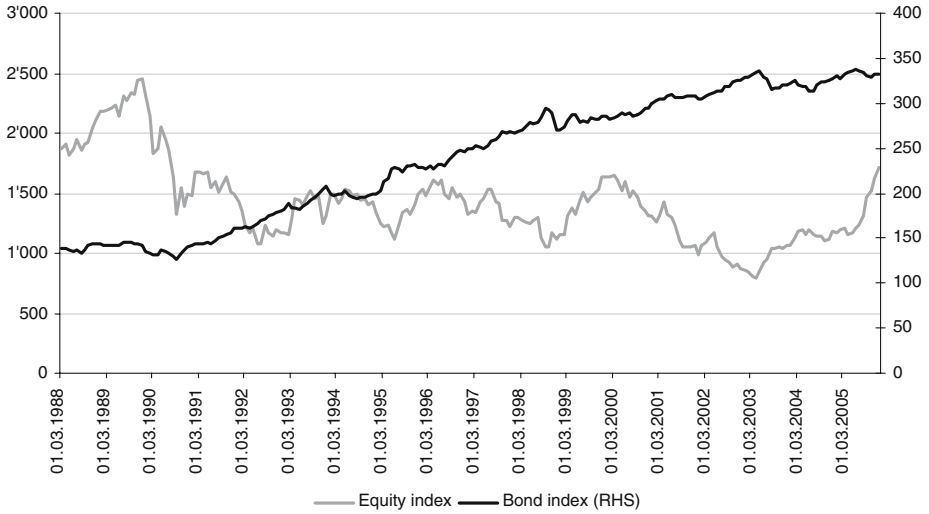

Fig. 5 Bond and equity index in Japan

3 Numerical solution based on data

To obtain the optimal policy parameter θ in the dynamic optimization problems (1)–(4) of the investor we apply the Recurrent Reinforcement Learning method of Moody et al. (1998). In contrast to traditional numerical solution techniques such as value iteration, this approach allows for a simpler problem representation, avoids Bellman's curse of dimensionality and is computationally efficient. The approach is called "direct" since it does not attempt to estimate the value function, but it rather recovers the optimal (policy) feedback rule.[2]

The gradient of U with respect to θ reads

$$\frac{\mathrm{d}U_T(\theta)}{\mathrm{d}\theta} = -\sum_{t=0}^{T} \frac{\mathrm{d}U_T(\theta)}{\mathrm{d}R_t^P} \left(\frac{\mathrm{d}R_t^P}{\mathrm{d}\omega_t} \frac{\mathrm{d}\omega_t}{\mathrm{d}\theta} + \frac{\mathrm{d}R_t^P}{\mathrm{d}\omega_{t-1}} \frac{\mathrm{d}\omega_{t-1}}{\mathrm{d}\theta} \right). \tag{5}$$

Derivatives of the portfolio return with respect to the bond weight are fully accounted for by partial derivatives. Quantities $\frac{\mathrm{d}\omega_t}{\mathrm{d}\theta}$ are total derivatives that depend upon the entire sequence of previous time periods. The temporal dependencies in a sequence of decisions are accounted for through a recursive update equation for the parameter gradients

[2] Note, that Brandt, Goyal, Santa-Clara, and Stroud (2006) and van Binsbergen and Brandt (2006) suggest an alternative approach to avoid computing the value function.

Table 1 Source of bond index data

Country	Total return bond index	Currency
United States	Citigroup WGBI U.S. 7–10Y	USD
United Kingdom	Citigroup WGBI UK 7–10Y	GBP
Germany	Citigroup WGBI Germany 7–10Y	EUR
Japan	Citigroup WGBI Japan 7–10Y	JPY

Table 2 Source of equity index data

Country	Equity Performance index	Currency
United States	S&P 500 Composite	USD
United Kingdom	FTSE 100	GBP
Germany	DAX 30	EUR
Japan	NIKKEI 225	JPY

$$\frac{\mathrm{d}\omega_t}{\mathrm{d}\theta} = \frac{\partial \omega_t}{\partial \theta} + \frac{\partial \omega_t}{\partial \omega_{t-1}} \frac{\mathrm{d}\omega_{t-1}}{\mathrm{d}\theta}. \tag{6}$$

This is a chain rule for ordered derivatives. They represent the total, i.e., the direct and indirect impact of the variable in the denominator on the variable in the nominator, while the ordinary partial derivatives only account for the direct impact. In our model, the feedback rule drives a wedge between both types of derivatives.

The dynamic system is then optimized by repeatedly computing the value of $U_T(\theta)$ on forward passes through the data and adjusting the coefficient θ by employing gradient descent,

$$\Delta\theta = \eta \frac{\mathrm{d}U_T(\theta)}{\mathrm{d}\theta}, \tag{7}$$

where η may be referred to as the adjustment speed or the learning rate. Optimization is carried out in batch mode, i.e., gradient of utility with respect to θ is calculated based on the full data sample.

We use benchmark indices for long–term bonds and equity to calculate associated returns. In particular, we use indices for the US, Germany, UK, and Japan given in Tables 1 and 2 in the period from March 1988 to January 2006 on a monthly basis.

Tables 3 and 4 report the summary statistics of the indices.

Figures 2–5 show the indices graphically.

Table 3 Summary statistics of bond index data

	United States	Germany	United Kingdom	Japan
Mean	0.006477285	0.005890761	0.007407128	0.004272847
Standard deviation	0.01764916	0.01331913	0.01745185	0.01568065
Skewness	−0.310161811	−0.625071902	0.137096437	−0.512137811
Kurtosis	0.3690106	0.722890735	0.682868314	2.776031095

Table 4 Summary statistics of equity index data

	United States	Germany	United Kingdom	Japan
Mean	0.009891415	0.009457494	0.009496691	0.001354506
Standard deviation	0.038921436	0.06101563	0.043256413	0.056704968
Skewness	−0.217538801	−0.552715052	−0.181931235	−0.038269241
Kurtosis	0.450711698	1.134471623	0.897494629	0.751726681

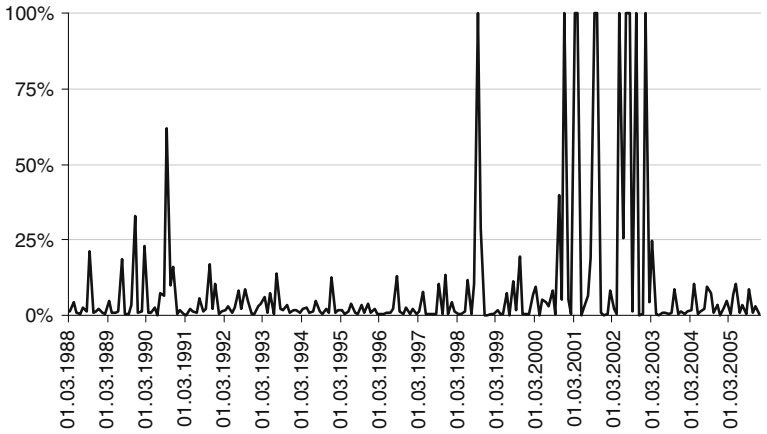

Fig. 6 Time-varying equity weight in the United States portfolio

The discount factor, and the coefficient of constant relative risk aversion are set, $\beta = 0.95$, and $\gamma = 2$, respectively.[3] With speed of adaption of $\eta = 0.0001$ the algorithm is well balanced since it converges reasonably fast and does not overshoot. The choice of $\beta = 0.95$ corresponds to a 5% interest rate and the degree of risk aversion is in the range that many researchers have found plausible (see e.g., Friend & Blume, 1975; Samuelson, 1991; Barsky, Juster, Kimball, & Shapiro, 1997).

To ensure that the convergence solution is statistically significant, we evaluate the results over 1,000 runs.

In the strategic asset allocation setting we get the following optimal bond weights, reported in Table 5. In contrast to western countries, in Japan the

[3] Note that our results are not unstable regarding variation in those parameters.

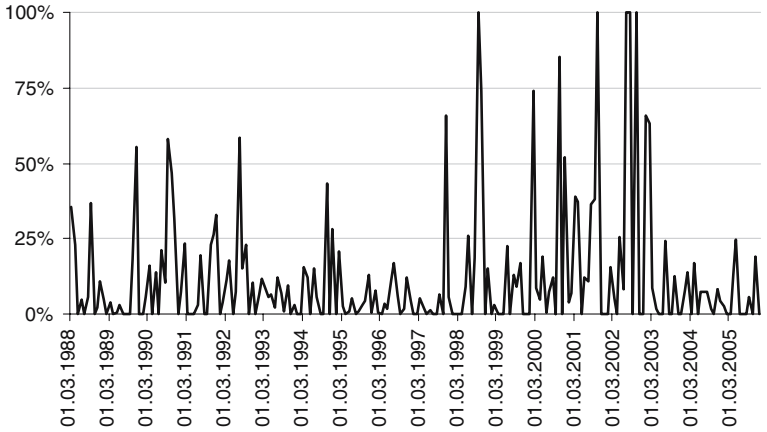

Fig. 7 Time-varying bond weight in the United Kingdom portfolio

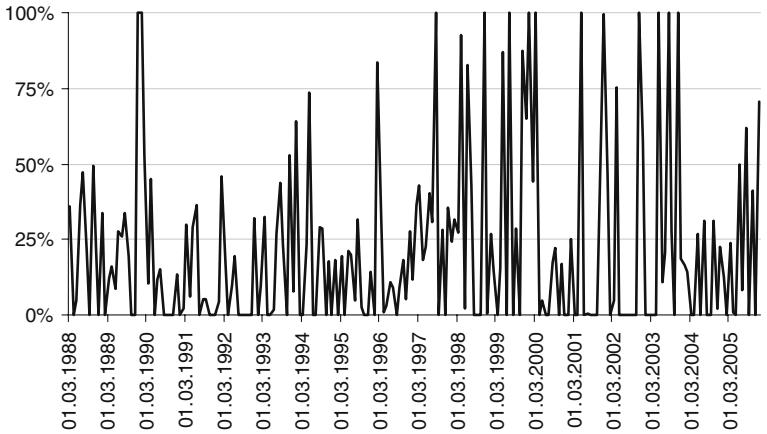

Fig. 8 Time-varying bond weight in the Germany portfolio

Table 5 Optimal bond weights in the strategic asset allocation

	United States	Germany	United Kingdom	Japan
ω	0.02	0.43	0.42	0.99

wealth is allocated more to bonds since equity prices have decreased in the considered sample from 1988 to 2005.

Table 6 shows that expected utility is improved by allowing for market timing based on the policy function introduced in the previous section.

These results are consistent with the autocorrelation coefficients of the yield spread in the time series in US, Germany, UK and Japan of 0.05743,

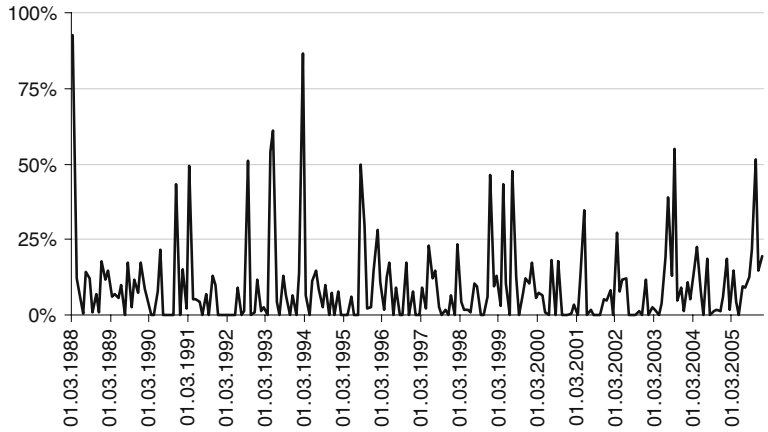

Fig. 9 Time-varying bond weight in the Japan portfolio

Table 6 Expected utility in the strategic asset allocation and market timing

	United States	Germany	United Kingdom	Japan
Strategic asset allocation	−0.352624795	−0.428284801	−0.386965587	−0.620097931
Market timing	−0.314529236	−0.370922543	−0.386184334	−0.596176757

Table 7 Optimal market timing parameters

	United States	Germany	United Kingdom	Japan
θ	−0.45	−0.34	−0.38	−0.36
F_1	50	5	5	5
F_2	100	50	25	25
F_2	0	0	0	0

0.085888, −0.01033, and 0.080646. Non-parametric independence tests indicate non-linear relationships significant on levels of 0.3989, 0.0894, 0.3892, and 0.1453, respectively.[4]

The optimal market timing parameters are given in Table 7.

Figures 6–9 show the weights of bonds resulting from optimal parameters within the market timing approach.

Figures 10–13 show the timing rule (1) resulting from optimal parameters within the market timing approach.

[4] See Wöhrmann (2006) for a description of the independence test.

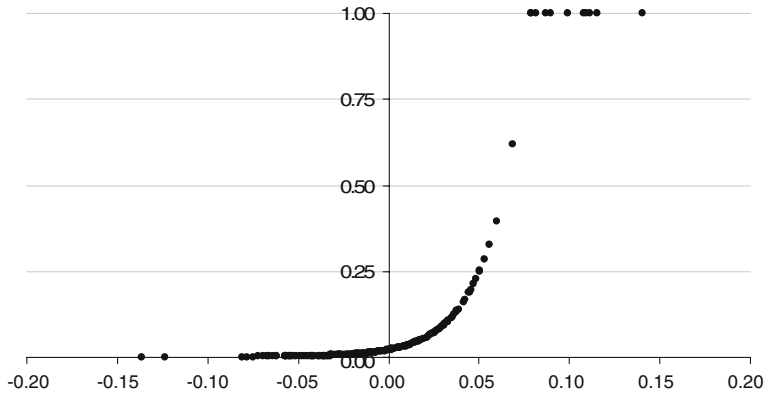

Fig. 10 Timing rule in the United States portfolio

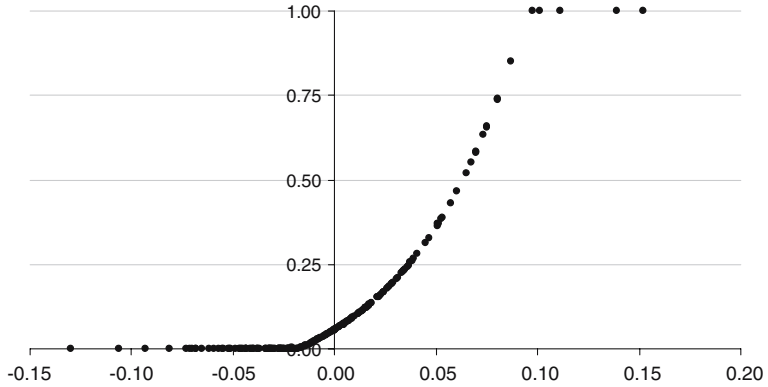

Fig. 11 Timing rule in the United Kingdom portfolio

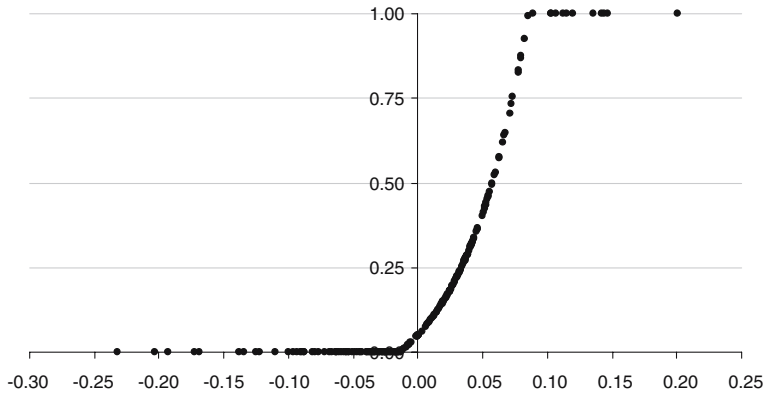

Fig. 12 Timing rule in the Germany portfolio

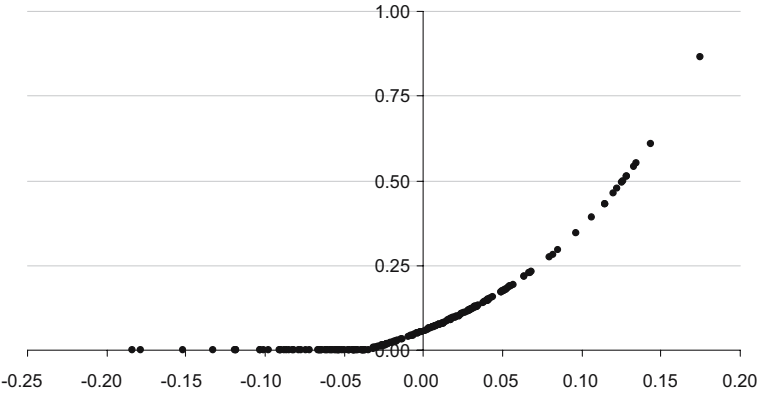

Fig. 13 Timing rule in the Japan portfolio

4 Conclusion

Our analysis shows that on the data we considered a rational investor with constant relative risk aversion will actively time the market and he will be able to outperform the market in terms of risk adjusted returns. Thus the data offers sufficient predictability to contest the fixed mix asset allocation rule. Our methodology can be applied on any financial market data so that a broader view of the efficiency of markets could be provided in the future.

Acknowledgements Financial support by the University Research Priority Programme "Finance and Financial Markets" of the University of Zurich and the national center of competence in research "Financial Valuation and Risk Management" is gratefully acknowledged. The national centers in research are managed by the Swiss National Science Foundation on behalf of the federal authorities.

References

Barsky, R. B., Juster, F. T., Kimball, M. S., & Shapiro, M. D. (1997). Preference parameters and behavioral heterogeneity. An experimental approach in the health and retirement study. *Quarterly Journal of Economics, 111*, 537–579.

van Binsbergen, J. H., & Brandt, M. W. (2006). Solving dynamic portfolio choice problems by recursing on optimized portfolio weights or on the value function. *Computational Economics*, forthcoming.

Brandt, M. W. (1999). Estimating portfolio and consumption choice: A conditional Euler equations approach. *Journal of Finance, 54*, 1609–1646.

Brandt, M. W., Goyal, A., Santa-Clara, P., & Stroud, J. R. (2006). A simulation approach to dynamic portfolio choice with an application to learning about return predictability. *Review of Financial Studies, 18*, 831–873.

Brennan, M. J., Schwartz, E. S., & Lagnado, R. (1997). Strategic asset allocation. *Journal of Economic Dynamics and Control, 21*(7), 1377–1403.

Campbell, J. Y., (2000). Asset pricing at the millenium. *Journal of Finance, 55*(4), 1515–1567.

Campbell, J. Y., Andrew, L. & Craig Mc Kinley, A. (1999). *The econometrics of financial markets*. Princeton NJ: Princeton University Press.

Campbell, J. Y., & Shiller, R. J. (1988). The dividend price ratio and expectations of future dividens and discount factors. *Review of financial studies, 1*, 195–228.
Campbell, J. Y., & Viceira L. M. (2002). *Strategic asset allocation*. Oxford: Oxford University Press.
Campbell, J. Y., & Viceira L. M. (1999). Consumption and portfolio decisions when expected returns are time-varying. *Quarterly Journal of Economics, 114*, 433–495.
Campbell, J. Y., & Viceira, L. M. (2005). Strategic asset allocation for pension plans. forthcoming. In Gordon Clark, Alicia Munnell, and Michael Orszag (Eds.), *Oxford handbook of pensions and retirement income*, Oxford: Oxford University Press.
Cochrane, J. H. (1991). *Portfolio advice for a multifactor world, economic perspectives XXIII (3) Third quarter 1999* (pp. 59–78). Chicago: Federal Reserve Bank of Chicago.
Daniel, K., Hirshleifer, D., & Subrahmanyam A. (1998). Investor psychology and security market under- and overreactions. *Journal of Finance, 53*(6), 1839–1885.
De Bondt, W., & Thaler R. H. (1985). Does the stock market overreact? *Journal of Finance, 40*(3), 793–805.
Gallant, A. R., Hansen, L. P., & Tauchen G. (1990). Using conditional moments of asset payoffs to infer the volatility of intertemporal marginal rates of substitution. *Journal of Econometrics, 45*(1–2), 141–179.
Fama, E. (1998). Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics, 49*, 283–306.
Friend, M. E., & Blume, I. (1975). The Demand for risky assets. *American Economic Review 65*(5), 900–922.
Malkiel, B. G. (2003). The efficient market hypothesis and its critics. *Journal of Economic Perspectives, 17*, 59–82.
Merton, R. (1969). Lifetime portfolio selection under uncertainty: The continuous-time case. *Review of Economics and Statistics, 51*, 247–257.
Moody, J., Wu, L., Liao, Y., & Saffell, M. (1998). Performance functions and reinforcement learning for trading systems and portfolios. *Journal of Forecasting, 17*, 441–470.
Samuelson, P. A. (1969). Lifetime portfolio selection by dynamic stochastic programming. *The Review of Economics and Statistics, 51*(3), 239–246.
Samuelson, P. A. (1991). Long-run risk tolerance when equity returns are mean regressing: Pseudo-paradoxes and vindication of Businessman's Risk, Brainard, Nordhaus, & Watts (Eds.), *Money, macroeconomics, and economic policy*. Essay in Honor of James Tobin, (chap. 7, pp. 181–200). Newyork: MIT Press.
Shleifer, A. (2000). *Inefficient markets: An introduction to behavioral finance*. Oxford: Oxford University Press.
Wöhrmann, P. (2006). *An axiomatic approach to prediction*. University of Zurich.