Optimal Trend-Following in a Markov Switching Model

Valeriy Zakamulin* and Javier Giner†

April 25, 2022

Abstract

This paper assumes that the market returns follow a two-state Markov process that randomly switches between bull and bear states. We show that in this case, the exponential moving average (EMA) represents the optimal trend-following rule. The paper provides the analytical solution to the optimal window size (decay constant) in the EMA rule. We estimate the optimal window size for timing the S&P 500 stock market index using real-world data. A comparative statics analysis finds that the optimal window size depends mainly on the signal-to-noise ratio of returns and the state transition probabilities.

Key words: Markov switching model, bull-bear markets, optimal trend-following, moving averages

JEL classification: G11, G17

*School of Business and Law, University of Agder, Service Box 422, 4604 Kristiansand, Norway, E-mail: Valeri.Zakamouline@uia.no

†Department of Economics, Accounting and Finance, University of La Laguna, Camino La Hornera s/n, 38071, Santa Cruz de Tenerife, Spain, E-mail: jginer@ull.edu.es

1 Introduction

The econometric literature documents a number of stylized facts about financial asset returns. These facts include negative skewness, positive excess kurtosis, serial dependence, and volatility clustering. A two-state Markov Switching Model (MSM) reproduces most of these stylized facts. In this model, the returns follow a process that randomly switches between bull and bear states. The market states are identifiable and persistent. Consequently, this model justifies trend-following trading.

There are many alternative trend-following indicators (rules) that are designed to identify the bull and bear states of the market. Most of these indicators are based on moving averages of past prices. The most popular is the Simple Moving Average (SMA). Less commonly used types of moving averages are the Linear Moving Average (LMA) and Exponential Moving Average (EMA). Each moving average is computed using an averaging window of a particular size. Besides, a trend-following indicator can be computed using either a single moving average or a combination of moving averages.[1] In this regard, the natural question is: What trend-following indicator is optimal when the returns follow a two-state MSM?

To the best knowledge of the authors, there is only a series of papers that answer this question using a continuous-time model and assuming the existence of trading costs (see Dai, Zhang, and Zhu (2010), Nguyen, Tie, and Zhang (2014a), Nguyen, Yin, and Zhang (2014b), Dai, Yang, Zhang, and Zhu (2016), and Jingzhi Tie (2016)). A typical goal in such a paper is to maximize a specific reward function, for example, the expected return to the trading strategy. The optimal trading strategy is represented by two time-dependent boundaries on the stock price. When the stock price is above (below) the upper (lower) boundary, a Buy (Sell) signal is generated. Finding these optimal boundaries turns out to be a difficult numerical task.

In contrast to the aforementioned papers, we consider a discrete-time model without transaction costs. We demonstrate that, in this case, the model is analytically tractable. We show that the EMA indicator represents the optimal trend-following rule and find the solution to the optimal window size (decay constant) in this indicator. We investigate how the optimal window size depends on the model's parameters through a comparative statics analysis. In a simple empirical application, we fit the model to real-world data and estimate the optimal

[1] For example, a moving average "crossover" is an indicator constructed using two moving averages: one with a short window size and another with a long window size.

window size for timing the S&P 500 stock market index.

2 Markov Switching Model for Returns

We denote by x_t the period-t return on a financial asset and assume that x_t is a stochastic process that randomly switches between two states: A and B. Formally, the state space of the process is $S_t \in \{A, B\}$ and the return distribution depends on the state S_t as follows

$$x_t = \begin{cases} \mu_A + \sigma_A z_t & \text{if } S_t = A, \\ \mu_B + \sigma_B z_t & \text{if } S_t = B, \end{cases} \tag{2.1}$$

where μ_A and σ_A are the mean and standard deviation of returns in state A, μ_B and σ_B are the mean and standard deviation of returns in state B, and z_t is an identically and independently distributed random variable with zero mean and unit variance. We assume that state A is a bull state of the market, while state B is a bear state of the market.

The conditional probabilities $Prob(S_{t+1} = J|S_t = I) = p_{IJ}$ are called one-period transition probabilities. For example, p_{AB} is the probability that the process transits from state A to state B over a single period. We assume the following transition probability matrix:

$$\mathbf{P} = \begin{pmatrix} p_{AA} & p_{AB} \\ p_{BA} & p_{BB} \end{pmatrix} = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}. \tag{2.2}$$

Formally, in this model the return distribution is given by a two-component mixture model. Even if z_t is a standard normal independent random variable, the distribution of x_t is an asymmetric distribution with fat tails, volatility clustering, and positive autocorrelations. Specifically, the lag-k return autocorrelation is given by (see Timmermann (2000) and Frühwirth-Schnatter (2006, Chapter 10)):

$$\rho_k = \frac{\pi_A \pi_B (\mu_A - \mu_B)^2}{\sigma^2}(1-\alpha-\beta)^k, \tag{2.3}$$

where $\boldsymbol{\pi}' = [\pi_A, \pi_B]$ is the vector of the steady-state (stationary or ergodic) probabilities

$$\pi_A = Prob(S_t = A) = \frac{\beta}{\alpha+\beta}, \quad \pi_B = Prob(S_t = B) = \frac{\alpha}{\alpha+\beta}, \tag{2.4}$$

and σ^2 is the variance of x_t

$$\sigma^2 = \pi_A\sigma_A^2 + \pi_B\sigma_A^2 + \pi_A\pi_B(\mu_A - \mu_B)^2. \tag{2.5}$$

In short notation, the lag-k return autocorrelation can be written as

$$\rho_k = c\delta^k, \tag{2.6}$$

where $\delta = 1 - \alpha - \beta$ and the variable c is given by

$$c = \frac{\pi_A\pi_B(\mu_A - \mu_B)^2}{\sigma^2}. \tag{2.7}$$

3 Exponential Moving Average Indicator

The EMA of prices is a trend-following indicator that is computed as

$$EMA_t(n) = (1-\lambda)\sum_{k=0}^{\infty}\lambda^k P_{t-k}, \quad \lambda = \frac{n-1}{n+1}, \tag{3.1}$$

where t is the time of computation, λ is a decay constant, $0 < \lambda < 1$, and P_{t-k} is the closing price at time $t-k$. In contrast to all other types of moving averages, the EMA is computed using the averaging window of an infinite size. The parameter n in the EMA denotes the size of the averaging window in the SMA that has the same average lag time as the EMA (see Zakamulin (2017, Chapter 3)). This convention (to quote n instead of λ) is used to unify the notation for all types of moving averages. The trading signal is generated depending on whether the last closing price is above or below the EMA of prices. In particular, the Buy (Sell) signal is generated when the last closing price is above (below) the EMA.

Alternatively, any trading indicator based on moving averages of prices can be computed using a moving average of returns (see Acar (1998), Lequeux (2005), Beekhuizen and Hallerbach (2017), and Zakamulin (2017)). Generally, the weighting function for returns differs from the weighting function for prices. However, in the case of the EMA, both weighting functions are identical. That is, the same EMA can be computed using returns instead of prices

$$EMA_t(n) = (1-\lambda)\sum_{k=0}^{\infty}\lambda^k x_{t-k}. \tag{3.2}$$

When the EMA is computed using returns, the trading signal is generated depending on the sign of the indicator. Specifically, the Buy (Sell) signal is generated when the indicator is positive (negative). Note that the same trading signal can be generated when the EMA is computed as

$$EMA_t(n) = \sum_{k=0}^{\infty}\lambda^k x_{t-k}. \tag{3.3}$$

This is because the sign of the product $a \times EMA_t(n)$ equals the sign of the $EMA_t(n)$, where a is any positive real number.[2]

4 Analytical Results

There exists a close relationship between Markov models and ARMA models. Specifically, Poskitt and Chung (1996) prove that the process in a p-state MSM can be represented by an $ARMA(p-1, p-1)$ process. Consequently, the observations of the return process in a two-state MSM are indistinguishable from the observations of the return process that follows an ARMA(1,1) model. Put differently, the return process defined by equation (2.1) admits an ARMA(1,1) representation. In particular, the return process can alternatively be specified by

$$x_t = \varphi x_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1}, \tag{4.1}$$

where φ and θ are some constants and ε_t is a homogeneous zero-mean white noise process. We assume that φ and θ satisfy the stationarity conditions.

It is well-known that any $ARMA(p,q)$ process admits an $AR(\infty)$ representation. The $AR(\infty)$ representation of the process specified by equation (4.1) is given by

$$x_t = (\varphi - \theta)\sum_{k=1}^{\infty}\theta^{k-1}x_{t-k} + \varepsilon_t.$$

Therefore, the best linear predictor of the next period return is given by (see Box, Jenkins,

[2] In our case, $a = (1-\lambda)^{-1}$.

Reinsel, and Ljung (2016, Chapter 5))

$$\hat{x}_{t+1} = (\varphi - \theta) \sum_{k=0}^{\infty} \theta^k x_{t-k}.$$

Note that the goal of a trend-following indicator is to predict the sign of the future return, not the return *per se*. That is, the goal is to predict whether the price will increase or decrease in the future. Therefore, the sign of the next period return in an $ARMA(1,1)$ model can be predicted by

$$\widehat{\text{sign}}(x_{t+1}) = \sum_{k=0}^{\infty} \theta^k x_{t-k}. \tag{4.2}$$

The similarity of the right-hand sides of equations (3.3) and (4.2) allows us to conclude that the EMA indicator represents the optimal predictor of the sign of the next period return in an $ARMA(1,1)$ model and, therefore, in a two-state MSM model as well. The optimal decay constant in the EMA rule $\lambda = \theta$. Our next goal is to find the analytical expressions for φ and θ.

The lag-1 autocorrelation of the $ARMA(1,1)$ process is given by (see Box et al. (2016, Chapter 3))

$$\rho_1 = \frac{(\varphi - \theta)(1 - \varphi\theta)}{1 - 2\varphi\theta + \theta^2}. \tag{4.3}$$

Then for every ρ_k, $k > 1$, we have

$$\rho_k = \rho_1 \varphi^{k-1}. \tag{4.4}$$

The functional form of the lag-k return autocorrelation given by equation (4.4) is similar to that of the lag-k return autocorrelation specified by equation (2.6). This similarity is nothing else than a direct consequence of the duality between a 2-state MSM process and an ARMA(1,1) process. A comparison of (4.4) and (2.6) suggests that

$$\varphi = \delta = 1 - \alpha - \beta. \tag{4.5}$$

This expression is natural from the viewpoint that φ is interpreted as the persistence of the $ARMA(1,1)$ process. The smaller the values of α and β, the higher the value of φ, and the process is more likely to stay in the same state than to transit to another state.

Remains to find the analytical expression for θ. We find it by matching the expressions

(4.3) and (2.6) for $k = 1$. This gives us the following equation:

$$\frac{(\varphi - \theta)(1 - \varphi\theta)}{1 - 2\varphi\theta + \theta^2} = c\varphi. \tag{4.6}$$

The result is a quadratic equation in θ that has two roots:

$$\theta_1 = d - \sqrt{d^2 - 1}, \quad \theta_2 = d + \sqrt{d^2 - 1},$$

where d is given by

$$d = \frac{1 + \varphi^2(1 - 2c)}{2\varphi(1 - c)}. \tag{4.7}$$

Since $d > 1$ (otherwise, both roots are complex numbers), we conclude that $\theta_2 > 1$ and, therefore, in this case the $ARMA(1, 1)$ process is non-stationary. Hence, the only sensible solution to equation (4.6) is provided by θ_1. Consequently, the analytical solution to both the constant θ and the optimal decay constant λ in the EMA indicator is given by

$$\theta = \lambda = d - \sqrt{d^2 - 1}. \tag{4.8}$$

5 Empirical Application

In our empirical application, we use the monthly capital gain returns[3] on the S&P 500 stock market index over the period from January 1926 to December 2020. This index is commonly used as a proxy for the US stock market. The data are provided by the Center for Research in Security Prices (CRSP). We reconstruct the stock index values from returns and identify the bull and bear markets using the method proposed by Pagan and Sossounov (2003). This method seems to be the most widely accepted method among researchers for such purposes. In brief, this method adopts, with minor modifications, the dating algorithm developed by Bry and Boschan (1971) to identify the US business cycle turning points using the GDP data.

Table 1 presents the summary statistics of the bull and bear markets. The values of the descriptive statistics agree closely with those reported in the papers by Pagan and Sossounov (2003) and Gonzalez, Powell, Shi, and Wilson (2005). From 1926 to 2020, there were 26 bull

[3] In practice, a moving average of prices is computed using closing prices not adjusted for dividends. That is, one uses prices directly observable in the market. Since we compute the EMA using returns, we use returns not adjusted for dividends.

Statistic	**Bull markets**	**Bear markets**
Number of states	26	25
Average duration	28.54	13.48
Stationary probability, %	67.92	32.08
Transition probability, %	3.50	7.42
Mean return, %	23.79	-28.69
Standard deviation, %	16.19	19.70

Table 1: Summary statistics of the bull and bear market states. Duration is measured in months. Mean returns and standard deviations are annualized and reported in percentages. Transition probability is the probability to transit to another state over a month.

and 25 bear states in the US stock market. The average bull (bear) market duration is 28.5 (13.5) months. The average bull market duration is approximately twice as long as the average bear market duration. The values of the stationary state probabilities confirm this observation. In particular, the probability that the US stock market is in the bull (bear) state amounts to 67.9% (32.1%). Similarly, the transition probability from a bear to a bull state, β, is about twice as large as the transition probability from a bull to a bear state, α. Finally, the mean return is equal to 23.8% (-28.7%) in a bull (bear) state of the market, while the standard deviation of returns amounts to 16.2% (19.7%) in a bull (bear) market.

The estimated optimal decay constant in the EMA indicator is $\lambda = 0.797$. Consequently, the optimal n in the EMA indicator is $n = 8.85$. Since n must be an integer number, the optimal EMA indicator is computed using the "averaging window" of 9 months. This number corresponds very well with the trading practice and can be justified as follows. The most popular trend-following indicator among traders is the 10-month SMA (see, for example, Faber (2007)). The paper by Zakamulin and Giner (2020) provides a number of results on the similarity of various trading indicators based on moving averages. The illustrations provided in Section 5 of this paper suggest that the trading signal of the 9-month EMA has the highest correlation (among all $EMA(n)$ indicators) with the trading signal of the 10-month SMA, and this correlation is close to 100%.

6 Comparative Statics Analysis

In the preceding section, we estimate the optimal size of the averaging window in the EMA indicator. Yet, we still lack a fundamental understanding of how the optimal window size depends on the model parameters. We fill this gap by conducting a comparative statics analysis.

There are totally six parameters in our two-state MSM for returns: α, β, μ_A, μ_B, σ_A, and σ_B. Note that the market states differ mainly in the value of mean return, not in the value of standard deviation. Roughly, the standard deviation is the same in both states, while the state absolute mean returns are approximately equal. Thus, for simplicity, we assume that $\mu_A = -\mu_B = \mu$ and $\sigma_A = \sigma_B = \sigma$. Further note that the transition probability β is roughly twice as large as the transition probability α. Therefore, for simplicity, we assume that $\beta = 2\alpha$. Consequently, our task reduces to investigating how the optimal n in the $EMA(n)$ indicator depends on three parameters: μ, σ, and α.

First of all, we fix $\alpha = 0.05$ and $\sigma \in \{20\%, 30\%\}$ and compute the optimal n for various $\mu \in [10\%, 40\%]$. The results are depicted in Figure 1, Panel A. Second, we fix $\alpha = 0.05$ and $\mu \in \{20\%, 30\%\}$ and compute the optimal n for various $\sigma \in [10\%, 40\%]$. The results are shown in Figure 1, Panel B. The clear-cut conclusion that can be drawn from the curves in these panels is that the optimal window size increases when either the state volatility increases or the mean state return decreases. In this regard, the fraction $\frac{|\mu_A - \mu_B|}{\sigma}$ can be interpreted as the *signal-to-noise ratio* of the market returns.[4] The higher the signal-to-noise ratio, the smaller the optimal n and vice versa. Finally, we investigate how the optimal n depends on the transition probability α. For this purpose, we fix $\frac{|\mu_A - \mu_B|}{\sigma} \in \{2, 3\}$ (in annual terms) and vary $\alpha \in [0.02, 0.20]$. Panel C in Figure 1 plots the results. We conclude that the optimal n decreases as the state transition probability increases.

In concluding this section, we suggest an economic interpretation of the results of our comparative statics analysis. Any trend-following indicator must simultaneously target two goals: to achieve the best accuracy in determining the trend direction and to identify turning points in the trend as early as possible. In this respect, first, it is worth mentioning that any trend-following indicator identifies the turning points in the trend with some delay. Second, it is pretty evident that increasing the averaging window size increases both the accuracy in forecasting the trend direction and the lag time in determining the trend change.[5] Therefore, one cannot achieve the two goals simultaneously; it is only possible to find the optimal trade-off between the accuracy and the lag time.

The higher the signal-to-noise ratio of returns, the easier to identify the trend direction.

[4]This fraction is motivated by the expression for the lag-k return autocorrelation specified by equation (2.3).
[5]Specifically, the lag time of the $EMA(n)$ indicator equals $(n-1)/2$, see Zakamulin (2017, Chapter 3).

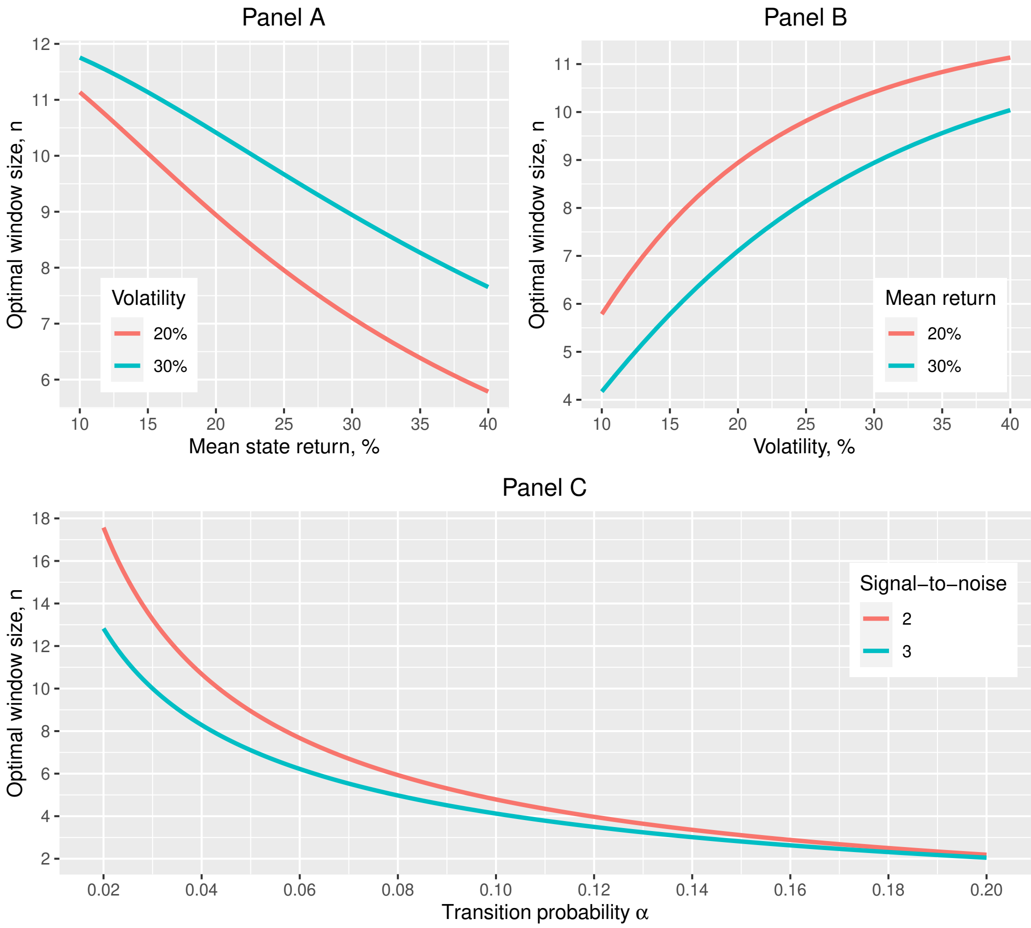

Figure 1: Comparative statics analysis.

Hence, a high signal-to-noise ratio allows one to decrease the lag time without considerable sacrifice in the accuracy by shorting the averaging window size. This explains why the optimal window size decreases when either the state mean return increases or the state volatility decreases. The negative relationship between the state transition probability and the optimal window size can be explained along similar lines. In particular, the larger the state transition probability, the shorter time the market stays in the same state. Consequently, if the market state changes frequently, one needs an indicator with a shorter delay time. A shorter delay time is achieved by decreasing the averaging window size.

7 Conclusions

A two-state Markov switching model has become increasingly popular in economic and financial studies because it reproduces most of the stylized facts of financial asset returns. Using a discrete-time model without transaction costs, we show that the EMA indicator represents the optimal trend-following rule and provide the analytical solution to the optimal window size (decay constant) in this rule. Using a comparative statics analysis, we find that the optimal window size depends mainly on the signal-to-noise ratio of returns and the state transition probabilities. Specifically, the optimal window size decreases when either the signal-to-noise ratio or the state transition probability increases. We estimate that the 9-month EMA rule is optimal for timing the S&P 500 stock market index.

References

Acar, E. (1998). "Expected Returns of Directional Forecasters", In Acar, E. and Satchell, S. (Eds.), *Advanced Trading Rules*, pp. 51–76. Butterworth-Heinemann, Oxford, UK.

Beekhuizen, P. and Hallerbach, W. G. (2017). "Uncovering Trend Rules", *Journal of Alternative Investments*, *20*(2), 28–38.

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2016). *Time Series Analysis: Forecasting and Control* (5 edition). John Wiley & Sons Inc., New Jersey.

Bry, G. and Boschan, C. (1971). *Cyclical Analysis of Time Series: Selected Procedures and Computer Programs*. NBER.

Dai, M., Zhang, Q., and Zhu, Q. J. (2010). "Trend Following Trading under a Regime Switching Model", *SIAM Journal on Financial Mathematics*, *1*(1), 780–810.

Dai, M., Yang, Z., Zhang, Q., and Zhu, Q. J. (2016). "Optimal Trend Following Trading Rules", *Mathematics of Operations Research*, *41*(2), 626–642.

Faber, M. T. (2007). "A Quantitative Approach to Tactical Asset Allocation", *Journal of Wealth Management*, *9*(4), 69–79.

Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. Springer, New York.

Gonzalez, L., Powell, J. G., Shi, J., and Wilson, A. (2005). "Two Centuries of Bull and Bear Market Cycles", *International Review of Economics and Finance*, *14*(4), 469 − 486.

Jingzhi Tie, Q. Z. (2016). "An Optimal Mean-Reversion Trading Rule Under a Markov Chain Model", *Mathematical Control & Related Fields*, *6*(3), 467–488.

Lequeux, P. (2005). "Optimal Allocation of Trend-Following Rules: An Application Case of Theoretical Results", In Dunis, C. L., Laws, J., and Naim, P. (Eds.), *Applied Quantitative Methods for Trading and Investment*, pp. 333–347. Wiley-Blackwell, West Sussex, UK.

Nguyen, D., Tie, J., and Zhang, Q. (2014a). "An Optimal Trading Rule Under a Switchable Mean-Reversion Model", *Journal of Optimization Theory and Applications*, *161*, 145–163.

Nguyen, D., Yin, G., and Zhang, Q. (2014b). “A Stochastic Approximation Approach for Trend-Following Trading”, In Mamon, R. S. and Elliott, R. J. (Eds.), *Hidden Markov Models in Finance: Further Developments and Applications, Volume II*, pp. 167–184. Springer US, Boston, MA.

Pagan, A. R. and Sossounov, K. A. (2003). “A Simple Framework for Analysing Bull and Bear Markets”, *Journal of Applied Econometrics*, *18*(1), 23–46.

Poskitt, D. S. and Chung, S.-H. (1996). “Markov Chain Models, Time Series Analysis and Extreme Value Theory”, *Advances in Applied Probability*, *28*(2), 405–425.

Timmermann, A. (2000). “Moments of Markov Switching Models”, *Journal of Econometrics*, *96*(1), 75–111.

Zakamulin, V. (2017). *Market Timing with Moving Averages: The Anatomy and Performance of Trading Rules*. Palgrave Macmillan.

Zakamulin, V. and Giner, J. (2020). “Trend Following With Momentum Versus Moving Averages: A Tale of Differences”, *Quantitative Finance*, *20*(6), 985–1007.