



Kiel

Working Papers

Kiel Institute
for the World Economy



Sentiment Dynamics and Stock Returns: The Case of the German Stock Market

by Thomas Lux

No. 1470 | December 2008

Web: www.ifw-kiel.de

Kiel Working Paper 1470 | December 2008

Sentiment Dynamics and Stock Returns: The Case of the German Stock Market

Thomas Lux

Abstract:

We use weekly survey data on short-term and medium-term sentiment of German investors in order to study the causal relationship between investors' mood and subsequent stock price changes. In contrast to extant literature for other countries, a tri-variate vector autoregression for short-run sentiment, medium-run sentiment and stock index returns allows to reject exogeneity of returns. Depending on the chosen VAR specification, returns are found to either follow a feedback process caused by medium-run sentiment, or returns form a simultaneous systems together with the two sentiment measures. An out-of-sample forecasting experiment on the base of estimated VAR models shows significant exploitable linear structure for the richer VAR(5) model. Out-of-sample trading experiments underscore the potential for excess profits from a VAR-based strategy compared to the buy-and-hold benchmark.

Keywords: investor sentiment, opinion dynamics, return predictability

JEL classification: G12, G14, C22

Thomas Lux

Kiel Institute for the World Economy

24100 Kiel, Germany

Phone: +49 431-8814 278

E-mail: thomas.lux@ifw-kiel.de

E-mail: lux@bwl.uni-kiel.de

The responsibility for the contents of the working papers rests with the author, not the Institute. Since working papers are of a preliminary nature, it may be useful to contact the author of a particular working paper about results or caveats before referring to, or quoting, a paper. Any comments on working papers should be sent directly to the author.

Coverphoto: uni_com on photocase.com

SENTIMENT DYNAMICS AND STOCK RETURNS: THE CASE OF THE GERMAN STOCK MARKET

THOMAS LUX* †

December 11, 2008

Department of Economics, University of Kiel
and
Kiel Institute for the World Economy

Abstract: We use weekly survey data on short-term and medium-term sentiment of German investors in order to study the causal relationship between investors' mood and subsequent stock price changes. In contrast to extant literature for other countries, a tri-variate vector autoregression for short-run sentiment, medium-run sentiment and stock index returns allows to reject exogeneity of returns. Depending on the chosen VAR specification, returns are found to either follow a feedback process caused by medium-run sentiment, or returns form a simultaneous systems together with the two sentiment measures. An out-of-sample forecasting experiment on the base of estimated VAR models shows significant exploitable linear structure for the richer VAR(5) model. Out-of-sample trading experiments underscore the potential for excess profits from a VAR-based strategy compared to the buy-and-hold benchmark.

JEL classification: G12, G14, C22

Keywords: *investor sentiment, opinion dynamics, return predictability*

*Contact adress: Thomas Lux, Department of Economics,
University of Kiel, Olshausen Str. 40, 24118 Kiel, Germany,
E-Mail: lux@bwl.uni-kiel.de

†The author is grateful for helpful comments to Jonas Dovern, Carsten-Patrick Meier, Adriano Rampini and participants at the staff seminar at the Kiel Institute. I also thank Thomas Theuerzeit of *animusX* for providing the sentiment data and offering useful insights into the structure of the survey. This research is supported by the Volkswagen Foundation through their grant for a research project on *Financial Markets as Complex Networks*.

1 Introduction

Many market participants and financial practitioners seem to believe in the informational content of various measures of “sentiment”. Under an efficient market perspective, however, any publicly available information such as the balances of surveys of traders’ expectations should already be incorporated in current prices. Excess profits on the base of measures of sentiment would contradict this view. Such an argument neglects the additional source of risk introduced by non-fundamental traders via their potentially erratic changes of mood and sentiment-based trading. This source of “noise trader risk” has been investigated by Black (1986) and DeLong et al. (1990) who demonstrate that asset prices can deviate from their intrinsic value, because noise trader risk imposes limits to the arbitrage activities of rational traders. As a consequence of this additional form of market risk non-rational traders would “create their own space”, i.e. a kind of ecological niche in which they can survive despite their erratic beliefs. Under certain conditions, noise traders would even drive smart money out of the market.

The limits of arbitrage, of course, depend on what information traders have. In DeLong et al. (1990) the mood of noise traders is a stochastic variable with independent realizations in each trading period. Noise trader risk is, therefore, unpredictable, in its direction and the extent of its changing moods of euphoria or pessimism. If, on the other hand, rational traders could get reliable signals on the contemporaneous disposition of noise traders, they should find it profitable to built up contrarian positions. It is, therefore, not clear whether the noise trading paradigm would be consistent with return predictability on the base of past or current sentiment information. The timing of information about the noise component is an obvious issue. For instance, if the realisation of the noise component is known prior to the opening of the market, noise traders would not represent an additional risk factor and should, therefore, not be able to survive. From this perspective, measures of noise traders’ mood should be at least synchronous with their measured impact on prices. Otherwise, noise traders could be all too easily exploited by more rational arbitrageurs. Since measures of sentiment are harder to collect than market statistics, we might even expect that survey measures of noise traders’ mood should lag behind

their market impact.

The proximity of certain popular measures of “sentiment” to noise traders’ mood in theoretical models has spawned a sizeable empirical literature on the predictive performance of sentiment. Most studies in this area have used one of the widely publicized indices of institutional or individual investors’ sentiment in the U.S. or indirect measures like the ratio of equity put to call trading volume, number of advancing issues to declining issue and others. Examples of this literature include Neal and Wheatley (1998) who find that indirect measures of sentiment predict the size premium and the difference between small firms and large firms, but have little predictive power for returns. Lee et al. (2002) find that a direct measure of sentiment enters significantly in both the mean and variance part of estimated GARCH-in-mean models for three major U.S. stock indices. Sentiment, therefore, appears to be able to contribute to the explanation of excess returns and also appears to be priced as a systematic component of market risk, which is in harmony with the noise trader story of De Long et al. (1990). Similar influences of sentiment on volatility are emphasized in Verma and Verma (2007). However, Wang et al. (2006) note that the forecasting power of sentiment for volatility vanishes if one adds lagged returns as an additional explanatory variable. The apparent predictive success of sentiment comes from causal influence of both returns and volatility on sentiment itself. According to their findings, sentiment only proxies for the leverage effect and autocorrelation in volatility, but does not have explanatory power for volatility beyond those well-known effects.

Brown and Cliff (2004, 2005) examine the predictability of returns over short and long horizons based on sentiment measures. While Brown and Cliff (2005) find that returns at 1 to 3 year horizons are negatively related to sentiment, neither institutional nor individual investors’ sentiment has predictive power for near-term returns (Brown and Cliff, 2004). The later study considers a simultaneous system composed of sentiment measures and stock returns, but finds that returns act as an exogenous variable. As it turns out, stock returns do Granger-cause sentiment, while sentiment is not a significant predictor of returns. Sentiment, therefore, appears to react as a feedback variable on market developments, but contains no useful information in itself for short-run prediction. A trading experiment using

sentiment as a signal shows no abnormal performance. Verma et al. (2008), in contrast, find some predictive power after decomposing both individual and institutional sentiment into rational and irrational components. Rational components are identified by a regression of the raw sentiment data on a set of fundamental factors while the residuals of this regression are interpreted as the irrational part. Both components have a positive impact on near-term returns with a larger effect of the ‘rational’ term in both individual and institutional sentiment. Apart from the U.S. market, the direction of causation between sentiment and returns has also been investigated for the Shanghai stock market by Kling and Gao (2008). Their results closely resemble those of Brown and Cliff in that they identify causation from returns on sentiment but not *vice versa*. Given the evidence from both the mature U.S. market and the relatively young emerging market in Shanghai, it seems that the use of sentiment data in short-run trading strategies should be quite limited. With the exception of the study by Verma et al., the above results indicate that sentiment does not provide any additional information beyond the informational content of past and contemporaneous stock prices and returns themselves.

This is also in harmony with the theoretical arguments laid out above as predictability of the noise component would evoke straightforward arbitrage arguments. Using sentiment data and returns for the German stock market, however, our study finds almost the opposite of the results of Brown and Cliff (2004) and Kling and Gao (2008). Considering two sentiment indices (short-run and medium run sentiment) and stock returns as a system, we cannot reject a causal influence from sentiment to returns. If we confine the analysis to a parsimonious VAR model with one or two lags only (supported by some information criteria), we indeed find that sentiment is exogenous and Granger-causes returns at high levels of significance. If we allow for more lags (also supported by some information criteria and specification tests), we see additional channels of interaction between our three variables. Under this richer specification, none of the variables would be exogenous, and their development in time would have to be understood as that of a simultaneous system. Still, this is an outcome very different from exogeneity of returns in the VAR analyses of the U.S. and Shanghai markets.

Out-of-sample forecasting experiments and an out-of-sample ‘market

timing' experiment indicate significant gains in accuracy and profitability of a VAR-based strategy. Overall, we find a surprising level of informational content of our sentiment measures. While, on a first view, the better part of the return predictability seems to come from medium-run sentiment, it is indeed a more subtle interplay between short-run sentiment, medium-run sentiment and returns that drives the positive out-of-sample results for the richer lag structure. While this is at odds with the majority of previous results for other markets, it is in line with another recent study for the German stock market. Using sentiment data from another source, Schmeling (2007) performed single predictive regressions and trading experiments. He reports significant predictability in long-horizon regressions plus some indication of profitability of sentiment-based trading. Our study goes beyond his contribution by studying the bi-directional causality between sentiment and returns within a simultaneous system and by exploring the profitability of a sentiment-based market timing strategy and the predictability of near-term returns on the base of estimated VAR models.

Our study proceeds as follows: Sec. 2 provides information on our data set of sentiment measures. While these seem to be quite popular among practitioners the data we use has apparently not been the subject of academic studies so far. Sec. 3 investigates sentiment and returns as a simultaneous system, exploring the causality structure, model specification issues and forecasting performance of VAR models. Sec. 4 continues with a 'market timing' experiment and sec. 5 concludes.

2 Data

Since mid 2004, *animusX-Investors sentiment*³, a provider of technical services and information for German investors, has published results of weekly surveys on the prospects of the German stock market over short-run and medium-run horizon. These surveys are conducted via email among

³Information on the services of *animusX* is available at <http://www.animusx.de>. The company offers a whole range of technical tools as well as various sentiment data for stock, bond and foreign exchange markets collected via weekly surveys among its subscribers.

about 2,000 registered private and institutional investors. According to the organizer of the survey, response rates are about 25 percent. The incoming categorial responses are recorded from Thursday, 6 p.m. to Saturday, 12 p.m. each week and the result is communicated in the form of a diffusion index (balance between optimistic and pessimistic responses) on the following Sunday at about 8 p.m. Our data set covers these diffusion indices for short-run and medium-run sentiment for the time horizon from the 29th calendar week of 2004 to the 22nd week of 2008 (a total of 202 observations)⁴. Short-run sentiment (S-Sent) concerns expectations for the following week while medium-run sentiment (M-Sent) covers respondents' expectations over the next three months. While the animusX sentiment data receive relatively wide coverage in the financial press, I am not aware of any previous scientific study using these data.

Fig. 1 exhibits the time development of the two indices together with the weekly returns (defined as continuously compounded returns, i.e. differences between the logs of the index) of the German share price index DAX. Returns are obtained from *Datastream* and are defined as log differences between weekly closing notations of the DAX⁵. As one might have expected, the short-run index exhibits much higher volatility than the medium-run index. While the short-run index flips often very dramatically from predominant optimism to pessimism and *vice versa*, the medium-run index rather seems to perform long swings between more moderate perceptions.

⁴Survey participants actually have the choice between five categorial answers. These are essentially a strongly pessimistic (expected price drop), mildly pessimistic (bottom formation), neutral, mildly optimistic (ceiling formation) and strongly optimistic view (expected price increase). The weights attached to the different categories are not made public by the owners. However, the company assures that these weights have been kept constant throughout the history of this survey. As is typical of such diffusion indices, their admissible range is $[-1, 1]$.

⁵In our trading experiment below we will also consider weekly opening quotations as the information received on Sunday evening could only be exploited by investment or withdrawal from the stock market on Monday morning with prices differing from last week's closing prices.

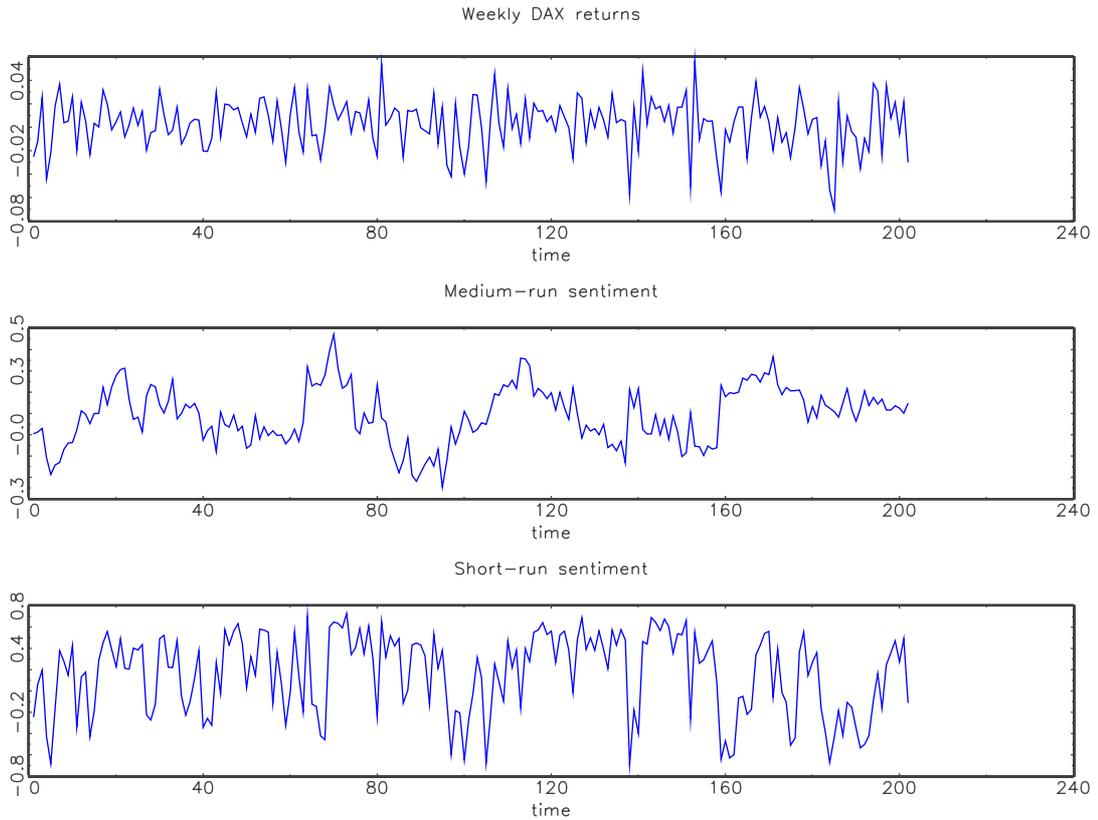


Fig 1: Sentiment and stock market returns. The time horizon is from the 29th calendar week of 2006 to the 22nd week of 2008.

Table 1 provides some sample statistics of our data. Since we will split the entire record into an in-sample of 150 observations and an out-of-sample part of the remaining 52 observations in our subsequent analysis, we give results for both the full sample and the reduced in-sample. The statistics confirm our impression that the short-run index has more variability and less temporal dependence than its medium-run counterpart. The ADF statistics indicate that we have little reason to doubt the stationarity of all three time series. Since both sentiment series are bounded between -1 and 1, strict non-stationarity seems practically impossible anyway. Our subsequent VAR analyses confirm stationarity of the tri-variate system since the estimated coefficient matrices have only stable eigenvalues throughout.

Table 1: Summary Statistics

Panel A: Full sample (202 observations)						
	Mean	S.D.	Skewness	Kurtosis	ρ_1	ADF
S-Sent	0.163	0.376	-0.546	-0.848	0.516	-5.649
M-Sent	0.092	0.132	-0.035	-0.363	0.790	-2.953
Returns	0.003	0.022	-0.503	0.488	-0.054	-10.153
Panel B: In-sample (150 observations)						
S-Sent	0.222	0.354	-0.732	-0.464	0.436	-4.632
M-Sent	0.073	0.136	0.204	-0.223	0.777	-2.856
Returns:	0.004	0.019	-0.414	0.287	-0.095	-9.561

Notes: S-Sent and M-Sent denote the short-run and medium-run sentiment index, respectively. ρ_1 is the autocorrelation coefficient at lag 1. The ADF test statistics have been computed with one lag, but inclusion of further lags did not change the results qualitatively. The one-sided 5 percent and 1 percent critical values are -1.991 (-1.942) and -2.652 (-2.602), for samples with 202 (150) entries, respectively. The qualitative results of the ADF tests also remain unchanged if we allow for a constant mean unequal zero in the alternative hypothesis.

3 VAR Analysis of Sentiment and Returns

Since the two sentiment indices as well as returns appear to be stationary, we can explore their dynamic relationship by estimating tri-variate VAR models⁶. A similar approach has been pursued, for example, by Brown and Cliff (2004) using a composite sentiment measure for the U.S. stock market and Kling and Gao (2008) who investigate daily sentiment data from a survey among institutional investors for the Shanghai stock market. Denoting the triples of observations on S-Sent, M-Sent and returns at time t by a vector $Y_t = (Y_{1t}, Y_{2t}, Y_{3t})'$, we assume a driving process of the form

$$Y_t = V + A_1 Y_{t-1} + \dots + A_p Y_{t-p} + U_t. \quad (1)$$

In (1), V is a 3×1 vector of intercept terms, the A_i are 3×3 coefficient matrices with entries $a_{i,j}^k$ (i the row, j the column number and k the lag order) and U_t is a vector of disturbances. We start by examining the

⁶Lütkepohl (2005) is a good source for the econometrics of simultaneous systems.

appropriate order of the VAR models to be estimated. As can be seen from Table 2, different selection criteria favor different lag orders: the Bayesian criterion (BIC) opts for only one lag to be included while the Hannan-Quinn criterion (HQ) favors two lags and the Akaike Criterion (AIC) assumes its minimum at five lags. It is impossible to judge on a priori grounds which of the three models would have to be preferred: although BIC and HQ are consistent and AIC is not, consistency might not really be that important in small samples like the present one. Also, AIC might produce superior forecasts despite inconsistency. We, therefore, leave the issue of model selection undecided and proceed with our subsequent analyses using all three candidate models. Also shown in Table 2 is a sequence of likelihood ratio tests (LR) that yields even more diverse behaviour: proceeding from larger to smaller lags, we find that the null hypothesis $A_{(4)} = 0$ (i.e. no significant entries in the coefficient matrices of order four or higher) is the first one rejected. From this perspective, a VAR(4) model would appear to be appropriate. However, subsequently, we see that the null hypotheses $A_{(3)} = 0$ and $A_{(2)} = 0$ can both not be rejected and disregarding the test results for higher lags, the results for lower dimensions would favor a parsimonious VAR(1) model.

Because of this ambiguity, we proceed by estimating VAR models up to order 5 and investigate their implied causal structures and forecasting capabilities. Table 3 shows the estimated parameters for the VAR(2) model as a benchmark case. Quite surprisingly in view of earlier literature for other countries, we find a strongly significant effect from sentiment on returns, but not the other way around. This indicates that investors' sentiment causes returns and should be exploitable to predict future market movements. Note that the effect is restricted to medium-run sentiment while short-run sentiment itself also appears to be caused uni-directionally by medium-run sentiment. Since only the second coefficient is significant in all three equations, our estimated VAR(2) model identifies M-Sent as an exogenous variable whose dynamics drives both S-Sent and returns. This is in total contrast to previous VARs of sentiment and stock market returns for the U.S. (Brown and Cliff, 2004) and China (Kling and Gao, 2008) that identify causality from returns on sentiment, but not the other way around. Also shown in Table 2 is the estimated covariance matrix that indicates

significant correlation of innovations of short-run sentiment and returns as well as similarly significant correlation between both sentiment measures. In contrast, medium-run sentiment and returns appear not to be instantaneously correlated.⁷

Specification tests show that residuals only suffer from significant skewness, but do not display excess kurtosis or temporal dependence. Specification tests for unrestricted VAR(1) and VAR(5) models yield almost identical results.

⁷Note that instantaneous correlation would not be problematic for the viewpoint of the efficient market theory. Such comovements of sentiment and returns could be due to anticipated changes of fundamental factors.

Table 2: Lag Selection for Tri-variate VAR Model

Panel A: Information criteria			
Lags	BIC	AIC	HQ
0	-14.808	-14.808	-14.808
1	-16.686	-16.867	-16.794
2	-16.602	-16.966	-16.818
3	-16.394	-16.943	-16.720
4	-16.263	-16.998	-16.699
5	-16.094	-17.018	-16.643
6	-15.822	-16.936	-16.483
7	-15.540	-16.845	-16.315
8	-15.214	-16.712	-16.103
9	-14.960	-16.654	-15.965
10	-14.670	-16.561	-15.793
Panel B: Sequence of LR tests			
VAR under H_0	LR	prob	
9	7.286	0.607	
8	12.513	0.186	
7	2.015	0.991	
6	5.265	0.811	
5	6.803	0.658	
4	17.653	<i>0.039</i>	
3	11.319	0.254	
2	16.825	0.052	
1	23.870	<i>0.005</i>	
0	304.046	<i>0.000</i>	

The results for the VAR(1) and VAR(5) models preferred by the BIC and AIC criteria are very similar. In order to save space we only list the significant parameter estimates in the following. We denote estimated parameters by $a_{i,j}^{(k)}$ where i is the row, j the column number and k refers to the lag order. For VAR(1) we obtain the following significant parameters: $a_{11}^{(1)} = 0.437^{**}$, $a_{12}^{(1)} = 0.490^*$, $a_{22}^{(1)} = 0.787^{**}$, $a_{32}^{(1)} = 0.029^*$. Again, we see the causal influence of M-Sent on both S-Sent and returns plus a strong relation of S-Sent to its own past which was absent in the

Table 3: Coefficient Estimates of VAR(2) Model

Panel A: Autoregressive Parameters				
	indep. var.			
dep. var	lag	S-Sent	M-Sent	Ret.
S-Sent	1	0.341	0.813*	2.137
M-Sent	1	-0.028	0.642**	0.306
Ret	1	0.004	0.056**	-0.118
S-Sent	2	0.211	-0.394	-3.025
M-Sent	2	-0.006	0.184*	0.092
Ret	2	0.000	-0.036	-0.206
Panel B: Covariance Matrix				
	indep. var.			
dep. var		S-Sent	M-Sent	Ret.
S-Sent		0.306**	0	0
M-Sent		-0.022**	0.080**	0
Ret		0.017**	-0.001	0.007**

Notes: The table shows parameter estimates obtained via maximum likelihood. * and ** denote significant parameters at the 95 and 99 percent significance level, respectively. Tests for model adequacy yield no indication of residual autocorrelation and excess kurtosis, but indicate significant skewness in residuals. The detailed results of standard tests are (probabilities in parenthesis):

Portmanteau test for autocorrelation with 12 lags: 89.776 (0.487)

Test for Normality:

Skewness statistic: 13.188 (0.004)

Kurtosis statistic: 5.260 (0.154)

Skew+kurt statistic: 18.448 (0.005)

VAR(2) model (in fact, $a_{11}^{(1)}$ was marginally insignificant at a p-value of 0.067 in the previous model). We also note that the estimates for the covariances and specification tests are qualitatively identical to those of the VAR(2) model. Although the VAR(5) model allows for a total of 45 autoregressive parameters, not too many of them are significant. These are: $a_{12}^{(1)} = 0.727^*$, $a_{22}^{(1)} = 0.558^{**}$, $a_{32}^{(1)} = 0.049^*$, $a_{12}^{(2)} = -0.947^*$, $a_{32}^{(2)} = -0.056^*$, $a_{33}^{(2)} = -0.408^*$, $a_{21}^{(5)} = -0.089^*$, $a_{23}^{(5)} = 1.302^*$. All other results are similar except for the fact that the covariance between the innovations of M-Sent and returns now becomes also significant. The VAR(5) model adds two

major types of dependencies. First, the lag 2 entries are all negative and, therefore, weaken the positive feedback at lag 1. More importantly, while all coefficients are insignificant at lags 3 and 4, lag 5 provides us with two significant parameters which entail feedback effects from both S-Sent and returns on M-Sent which consequently is not exogenous anymore in the VAR(5) model. To highlight the differences in the economic interpretation of the estimated models, we report in Table 4 Wald test statistics for Granger causality tests between returns and sentiment. Since we are mainly interested in the endogeneity or exogeneity of returns, we test jointly for Granger causality from both indices on returns and *vice versa*. Also shown in Table 4 is the result of a Wald test for instantaneous causality between sentiment and returns which is highly significant for all three models.

Table 4: Causality Tests

	sent \rightarrow ret	ret \rightarrow sent	inst
Var(1)			
Wald	7.372	1.234	204.208
p-value	0.025	0.540	0.000
Var(2)			
Wald	10.058	3.179	213.073
p-value	0.040	0.528	0.000
Var(5)			
Wald	16.672	15.944	214.100
p-value	0.082	0.101	0.000

Note: Granger causality tests report the p-values of the null hypotheses that all coefficients of the first variable(s) at all lags are jointly insignificant in the equation(s) of the second variable(s). The test for contemporaneous causation tests the hypothesis of no correlation between the innovation of returns and the innovations of sentiment variables. p-values are computed from the underlying Chi-square distribution under the null.

There is a striking difference between the more parsimonious VAR(1) and VAR(2) models on the one hand, and the richer VAR(5) on the other hand: From the former we would infer that the sentiment variables are exogenous to the system and returns develop as a endogeneous feedback process caused by sentiment. That sentiment itself is independent from returns (though it might have an interacting, intrinsic term structure in the interplay between R-Sent and M-Sent) suggests a kind of pure noise trading

interpretation of our findings. Sentiment would itself be decoupled from the price dynamics but would allow prediction of future returns.

However, note that the price impact of this noise component could be predicted prior to its realisation on the base of the information contained in the survey. This contradicts the standard timing conventions in the noise trader literature: If the realisation of noise traders' mood is known before the market opens, arbitrageurs should be able to trade against it and noise traders would be unable to 'create their own space'. While we can certainly not exclude that sentiment proxies for some other factors, the uni-directional Granger causality of the VAR(1) and VAR(2) models constitutes surprisingly strong evidence against informational efficiency of the German stock market. Somewhat in contrast, the VAR(5) model does not allow rejection of any of the nulls for causation between sentiment and returns. Here we can not identify a hierarchical structure of exogenous and endogenous variables. The dynamic evolution appears more complex and all three variables seem to coevolve as a system with feedback in all directions. In particular, sentiment appears to be influenced also by past returns, albeit with a relatively long lag of more than one month. Note, however, that this bi-lateral causation has the same implications with respect to lack of efficiency since endogeneity of returns is not rejected.⁸

Playing around with impulse responses and variance decomposition we found that the major part of the forecast error variance of returns is accounted for by either innovations in returns themselves or innovations in

⁸Note also that M-Sent reacts negatively on optimistic short-run sentiment with a lag of five weeks. Such an inverse relationship between different sentiment measures has also been found elsewhere in the literature: Both Brown and Cliff (2005) for the U.S. and Schmeling (2007) for a different set of German sentiment data found a negative effect from individual investors' sentiment on that of institutional investors. Identifying short-run sentiment with individual investors' disposition and medium-run sentiment with that of institutional investors, our finding of a negative effect at lag 5 is similar to their results. Nevertheless, the interpretation is cumbersome: Following the above authors, individual (short-term) euphoria would drive prices up which makes institutional (medium-term) investors expect a reversal of the trend in the medium run. Besides the problem that these considerations would coincide in the simultaneous assessment of the near-term and medium-term prospects by the same participants in our survey, we also lack an indication of the direct influence of S-Sent on returns (all pertinent coefficients are insignificant).

S-Sent, depending on the order of the variables in our system. In our estimated VAR models this factor typically accounts for about 80-90 percent of the error variance. Apparently, since it depends on the ordering of the simultaneous system, this contribution is due to the strong correlation of the innovations of both variables - which at least with our present VAR approach can hardly be exploited for forecasting. If S-Sent appears in the ordering of the variables behind returns its contribution is tiny while M-Sent has a more sizable influence. In addition, we have also computed variance decompositions replacing our original models by reestimated ones in which the insignificant parameters have been replaced by zero constraints to reduce estimation noise. This exercise yielded somewhat more variability in the composition of error components as can be seen in the selected examples given in Table 5. Note, however, that despite the different assignment of the major component we see a relatively consistent small error component of about 6 percent associated with M-Sent which might reflect the *dynamic* causality identified above. In order to see in how far this helps to predict stock prices in practice we turn to an explicit out-of-sample forecasting exercise. Following a general-to-specific philosophy, we use restricted versions of our VAR(1), VAR(2) and VAR(5) models in which insignificant parameters have been set equal to zero in order to reduce estimation noise from the large number of insignificant parameters in the general VAR specifications. Since we have used 150 of our 202 observations for estimation, we have a full year of out-of-sample observations left.

As it turned out, restricted models had residuals that strongly violated all specification tests: the portmanteau as well as the Normality tests (for both skewness and kurtosis) always yield strong rejections of either null hypothesis for the VAR(2) and VAR(5) models with constrained parameter coefficients. Because of the violation of the distributional assumptions of the ML estimator we also estimated these models via (constrained) least squares. While the LS estimates show almost exactly the same pattern of qualitative results (in terms of significant parameters and Granger causality) some of their numerical estimates exhibit quite some variation⁹.

⁹For example, the estimated parameters of the VAR(5) model obtained by constrained LS are: $a_{12}^{(1)} = 0.351$, $a_{22}^{(1)} = 0.703$, $a_{32}^{(1)} = 0.063$, $a_{12}^{(2)} = 0.862$, $a_{32}^{(2)} = -0.005$, $a_{21}^{(5)} = -0.088$, $a_{23}^{(5)} = 1.048$. The major difference appears to be in the dynamic feedback between both sentiment measures, namely strongly positive versus strongly

Table 5: Some Examples of Error Variance Decompositions for Returns

VAR(5) model: unrestricted			
periods	S-Sent innovations	M-Sent innovations	ret. innovations
1	83.6971	0.5963	15.7066
2	80.2474	4.8440	14.9086
3	78.1227	5.6280	16.2493
4	77.2252	5.5497	17.2251
5	76.6964	5.6232	17.6804
6	75.8860	6.5444	17.5696
7	75.9921	6.5241	17.4837
8	76.0038	6.5220	17.4743
VAR(5) model: unrestricted, different order			
periods	ret. innovations	S-Sent innovations	M-Sent innovations
1	100.0000	0.0000	0.0000
2	96.0860	0.0324	3.8815
3	94.3096	0.4204	5.2699
4	92.4099	2.4320	5.1581
5	92.0069	2.6374	5.3558
6	91.1769	2.7017	6.1215
7	91.1305	2.7659	6.1036
8	91.1323	2.7668	6.1010
VAR(5) model: restricted			
periods	S-Sent innovations	M-Sent innovations	ret. innovations
1	64.8130	1.4030	33.7839
2	62.4474	5.2427	32.3099
3	62.5441	5.3986	32.0572
4	62.5282	5.4243	32.0474
5	62.5215	5.4087	32.0698
6	62.4903	5.4592	32.0505
7	62.5621	5.4494	31.9885
8	62.5546	5.4651	31.9803

Note: The order of the variables in the respective VARs corresponds to the ordering of the forecast error components.

negative entices of $a_{12}^{(2)}$.

Table 6: RMSEs of Out-of-Sample Forecasts

Panel A: Forecasts from models estimated via constrained ML						
Forecasts of single returns						
horizon	VAR(1)	DM _{adj}	VAR(2)	DM _{adj}	VAR(5)	DM _{adj}
1	1.027	-0.184	0.974	1.247	1.187	-1.125
2	1.031	-0.610	0.977	1.017	1.195	-1.372
3	1.018	-0.320	0.973	1.017	1.018	-0.392
4	1.012	-0.285	0.956	1.242	1.016	-0.260
5	1.009	-0.294	0.957	1.420	1.019	-0.542
6	1.003	-0.014	0.953	1.821*	1.020	-0.625
7	1.004	-0.217	0.957	1.893*	1.012	-1.129
8	0.999	0.183	0.953	1.518	1.011	-1.504
Forecasts of cumulative returns						
1	1.027	-0.184	0.974	1.247	1.187	-1.125
2	1.056	-0.392	0.958	1.248	1.266	-0.977
3	1.061	-0.367	0.942	1.216	1.165	-0.881
4	1.062	-0.353	0.919	1.264	1.110	-0.696
5	1.064	-0.375	0.898	1.354	1.111	-0.682
6	1.064	-0.405	0.873	1.557	1.124	-0.784
7	1.068	-0.500	0.855	1.652*	1.127	-0.880
8	1.066	-0.489	0.835	1.722*	1.129	-0.951

cont'd

Using both sets of estimates, we launch our forecasting exercise using time horizons of 1 up to 8 periods ahead. We consider both single returns as well as cumulative returns over longer horizons. Results can be found in Table 6. All mean-squared errors in this Table are standardized by dividing by the MSE of a random walk with drift as the benchmark model. While the performance of ML estimates is rather dismal, the LS-VAR models do consistently better than the benchmark of the random walk with drift. Judged by the adjusted Diebold-Mariano Test (Clark and West, 2007), their performance is significantly better in many cases than a white noise prediction. The performance of estimated models also appears to improve with the number of included lags in the VAR model as well as the forecast horizon. Since systematic effects at longer horizons are captured better by VAR models with a richer lag structure, these two findings taken together sug-

Panel B: Forecasts from models estimated via constrained LS						
Forecasts of single returns						
horizon	VAR(1)	DM _{adj}	VAR(2)	DM _{adj}	VAR(5)	DM _{adj}
1	0.974	1.248	0.974	1.247	0.913	2.042*
2	0.978	1.010	0.978	1.016	0.921	1.898*
3	0.967	1.193	0.972	1.052	0.933	1.593
4	0.947	1.434	0.954	1.281	0.911	2.008*
5	0.947	1.605	0.955	1.463	0.931	1.933*
6	0.941	1.976*	0.950	1.866*	0.933	2.066*
7	0.945	2.166*	0.955	1.968*	0.940	2.009*
8	0.940	1.880*	0.950	1.597	0.941	1.744*
Forecasts of cumulative returns						
1	0.974	1.248	0.974	1.247	0.913	2.042*
2	0.956	1.321	0.958	1.263	0.850	2.175*
3	0.934	1.350	0.941	1.242	0.813	2.134*
4	0.906	1.419	0.917	1.295	0.769	2.151*
5	0.879	1.544	0.894	1.392	0.736	2.266*
6	0.846	1.788*	0.867	1.603	0.699	2.377**
7	0.821	1.950*	0.848	1.716*	0.674	2.485**
8	0.794	2.067*	0.826	1.796*	0.649	2.461**

Note: For each model, the first entry gives the relative MSE of the forecasts from the pertinent VAR (i.e., original MSE divided by that of the random walk model with drift). DM_{adj} is the adjusted Diebold-Mariano test statistic for equal predictive accuracy of nested models. To compute this statistics we used the Newey-West autocorrelation and heteroskedasticity consistent estimator of the standard deviation with automatic lag selection by Andrew's method. * and ** identify cases of significantly better predictive accuracy of the pertinent VAR forecasts compared to those of a random walk with drift using one-sided 5 and 1 percent tests.

gest that the inclusion of additional parameters at lag 5 is not spurious but covers medium-term feedback effects between sentiment and returns as well as feedback effects within the term structure of the sentiment indices that would otherwise have been suppressed.¹⁰

¹⁰Note that the forecast improvement against the random walk is quite expressive: For a 8-week horizon, the forecasts based on the restricted VAR(5)-LS estimates have a MSE that is 35 percent below the benchmark.

4 Profitability of VAR-Based Trading Strategies

The estimated vector autoregressions have been used above to compute forecasts of future returns on the base of contemporaneous and past data for both the sentiment indices and returns. These forecasts could easily be used as input in the design of a straightforward trading strategy. Investigation of the performance of such a strategy allows us to let the data speak about the economic significance of the forecast improvements against the random walk found in Table 6. Since our simultaneous system provides us with forecasts of price changes for the stock market index, our experiment will test the ‘market timing’ potential of our VAR models. We, therefore, allow our ‘VAR trader’ to switch between investment in stocks and investment in safe government bonds. As the alternative risk-free rate we use the appropriately discounted interest rate for German government bonds with 10 years maturity at weekly frequency (downloaded from *Datastream*). Since bond rates are very persistent, we just take the current rate and disregard changes over one week or expectations thereof (the documented rates are averages within one week which is the time horizon of the investor who receives the sentiment data from *animusX* on Sunday evening). Following the pertinent literature, realistic transaction costs might range between 0.1 percent for large institutional investors and 0.5 percent for others (taking into account price impact). We therefore, compute trading profits for transaction costs c of 0.1, 0.25 and 0.5 percent¹¹.

Given the expectations for the next period produced by our VAR models, a risk-neutral myopic investor would switch between the stock and bond market if the alternative provides a higher expected return net of transaction costs. Denoting by r_t as before the continuously discounted expected return of the stock market, and by i_t the discrete discounted yield on government bonds, an investor who is currently invested in stocks, will switch to bonds if

¹¹For large institutional investors, the lower numbers might be realistic while for individual investors the higher end of our spectrum might be more relevant.

$$E[r_{t+1}] < \ln((1 - c)(1 + i_{t+1})). \quad (2)$$

Vice versa, a move from the bond to the stock market will happen if

$$E[r_{t+1}] > \ln\left(\frac{1 + i_{t+1}}{1 - c}\right) \quad (3)$$

holds. If the reverse of one of these inequalities holds, the trader will simply remain invested in either stocks or bonds.

For computing returns of the VAR-based strategy, we have to take into account that a positive or negative signal obtained from the VAR model with the new sentiment information on Sunday evening could only be used for trading once the stock market opens again on Monday morning. We have, therefore, used the weekly opening prices as transaction prices if a buy or sell signal was received during the weekend. However, as a comparison with the closing quotation of the preview week shows, the impact of the weekend return is small. Overall results would be almost completely unaffected if we had used instead the previous closing notation. Note that this shows that incoming information is not immediately incorporated into asset prices during the opening auction. At best, a very small part of it is transmitted into opening prices.

We start out with investment in stocks which is a very likely scenario for the bull market in mid 2007 (at the end of our in-sample, the DAX had been steadily rising for 11 consecutive weeks). Short-run sentiment was highly optimistic and medium-run sentiment had been close to neutral (fluctuating within ± 10 percent about zero) for this time. Starting with a bond portfolio only changes results marginally since the VAR models would have signalled an immediate switch to stocks at the beginning of the experiment anyway.

A buy-and-hold strategy for the one-year period from June 07 to June 08 would have generated a negative return of -10.8 percent. Following our sequence of estimates, we use VAR models estimated via least squares, both with constrained and unconstrained parameter sets. We also repeated this exercise with (constrained and unconstrained) ML estimates (not shown

here). Both the ML-VARs (whether constrained or not) and the unconstrained LS-VARs (exhibited in Table 7) had relatively poor results without any systematic pattern. The more interesting ones were the constrained LS-VARs on which we focus below. Their more interesting performance is in nice agreement with the out-of-sample forecasting experiments reported above in Table 6. Note also that our focus on constrained LS-VARs is quite plausible given the results of the parameter estimates and specifications tests for the in-sample series. Common sense application of the statistical apparatus in the spirit of a general-to-specific philosophy would lead us to prefer this specification, which speaks against a model-mining interpretation of our results. Table 7 exhibits the aggregate returns of actively managed portfolios on the base of different VAR models and for different levels of transaction costs. As we can see, the results depend on both the dimension of the VAR and the level of transaction costs. In order to be able to draw inferences with respect to significance of net or excess profits, we perform two bootstrap tests: We resample the out-of-sample observations either with or without replacement and apply our trading strategy for 1,000 bootstrapped resamples of both types.¹² Table 7 also displays the 90, 95 and 99 percent quantiles of these tests.

Focusing on the constrained VARs we see quite interesting results: the parsimonious VAR(1) and VAR(2) models reduce the investor's loss to about -1 percent under transaction costs of 0.1 and 0.25 percent. For both models, this performance is highly significant at least at the 95 percent level and, in fact, the trading operations are exactly identical for both models and both specifications of transaction costs (cf. the left panel of Fig. 2). The outcome of the trading strategy and the bootstrapping experiments on the base of the VAR(5) model seems quite bizarre on a first view. While it has even slightly higher profits at low and medium transaction costs than VAR(1) and VAR(2), these are not significant under standard confidence levels. However, when increasing transaction costs to 0.5 percent, this strategy achieves a net return of 5.2 percent (16 percent above the buy-and-hold strategy) which now is significant at the 95 percent confidence level.

¹²To be precise: Bootstrapping means here that we destroy the temporal order of the triples of observations consisting of S-Sent, M-Sent and returns at time t .

Table 7: Profitability of Trading Based on VARs

Panel A: Transaction costs of 0.1%					
Unconstrained VARs					
			bootstrapped quantiles		
VAR	B&H	active	90%	95%	99%
1	-0.108	-0.012	-0.060 (-0.053)	-0.037 (-0.025)	0.003 (0.033)
2	-0.108	-0.094	-0.023 (-0.019)	0.004 (0.019)	0.047 (0.072)
3	-0.108	-0.107	-0.012 (-0.007)	0.017 (0.032)	0.089 (0.095)
Constrained VARs					
1	-0.108	-0.012	-0.035 (-0.032)	-0.006 (-0.006)	0.041 (0.040)
2	-0.108	-0.012	-0.035 (-0.032)	-0.006 (-0.006)	0.041 (0.040)
3	-0.108	0.012	0.039 (0.051)	0.080 (0.088)	0.142 (0.167)
Panel B: Transaction costs of 0.25%					
Unconstrained VARs					
			bootstrapped quantiles		
VAR	B&H	active	90%	95%	99%
1	-0.108	-0.094	-0.098 (-0.094)	-0.076 (-0.078)	-0.067 (-0.028)
2	-0.108	-0.088	-0.057 (-0.043)	-0.031 (-0.013)	0.017 (0.055)
3	-0.108	-0.110	-0.043 (-0.039)	-0.008 (-0.005)	0.052 (0.071)
Constrained VARs					
1	-0.108	-0.018	-0.048 (-0.046)	-0.019 (-0.021)	0.032 (0.025)
2	-0.108	-0.018	-0.048 (-0.046)	-0.019 (-0.021)	0.032 (0.025)
3	-0.108	-0.005	0.016 (0.013)	0.044 (0.048)	0.108 (0.104)

cont'd

Fig. 2 provides some insights into the origin of these divergent results. As can be seen, VAR(1) and VAR(2) achieve their advantage over buy-and-hold via a very short period of absence from the stock market at the beginning of the out-of-sample period. Overall, these models issue very few signals to switch between markets which makes it difficult to earn returns different from the buy-and-hold strategy. Apparently, the pertinent

Panel C: Transaction costs of 0.5%					
Unconstrained VARs					
			bootstrapped quantiles		
VAR	B&H	active	90%	95%	99%
1	-0.108	-0.108	-0.108 (-0.108)	-0.108 (-0.108)	-0.108 (-0.108)
2	-0.108	-0.062	-0.079 (-0.067)	-0.060 (-0.040)	-0.009 (0.023)
3	-0.108	-0.180	-0.063 (-0.053)	-0.030 (-0.021)	0.020 (0.045)
Constrained VARs					
1	-0.108	-0.108	-0.108 (-0.108)	-0.108 (-0.108)	-0.108 (-0.108)
2	-0.108	-0.108	-0.108 (-0.108)	-0.108 (-0.108)	-0.108 (-0.108)
3	-0.108	0.052	0.015 (0.016)	0.052 (0.050)	0.104 (0.106)

Note: The bootstrapped quantiles are obtained from 1,000 resamples by either scrambling the sequence of the trivariate out-of-sample data or by drawing with replacement from this sample (those in brackets).

models would typically not be able to find these few profitable switches in randomized data. Despite the slight difference in activity compared to the buy-and-hold strategy, the gains in performance are, therefore, significant at least at the 95 percent level. With higher transaction costs (0.5 percent), however, the VAR(1) and VAR(2) traders would become completely inactive in both the out-of-sample experiments and all its randomized bootstrap replications which leads to the degenerate distribution of the bootstrapping experiment displayed in the subsequent rows of table 7.

The VAR(5) model generates more activity presumably because of the richer interaction between returns and sentiment leading to a higher number of forecasts exceeding the transaction cost boundary. However, the bootstrapped distribution in the cases of $c = 0.1$ and $c = 0.25$ show that these higher excess profits would not be too untypical for such an active strategy even if the order of the data is randomized. To switch at appropriate times could be simply a matter of luck. With $c = 0.5$, the VAR(5) strategy becomes more selective in the choice of switching times, i.e. it needs higher expected returns to indicate a shift of the portfolio composition. Appar-

ently, its choices are more often right than wrong and to end up with this sequence of events is relatively unlikely in the bootstrap experiments.

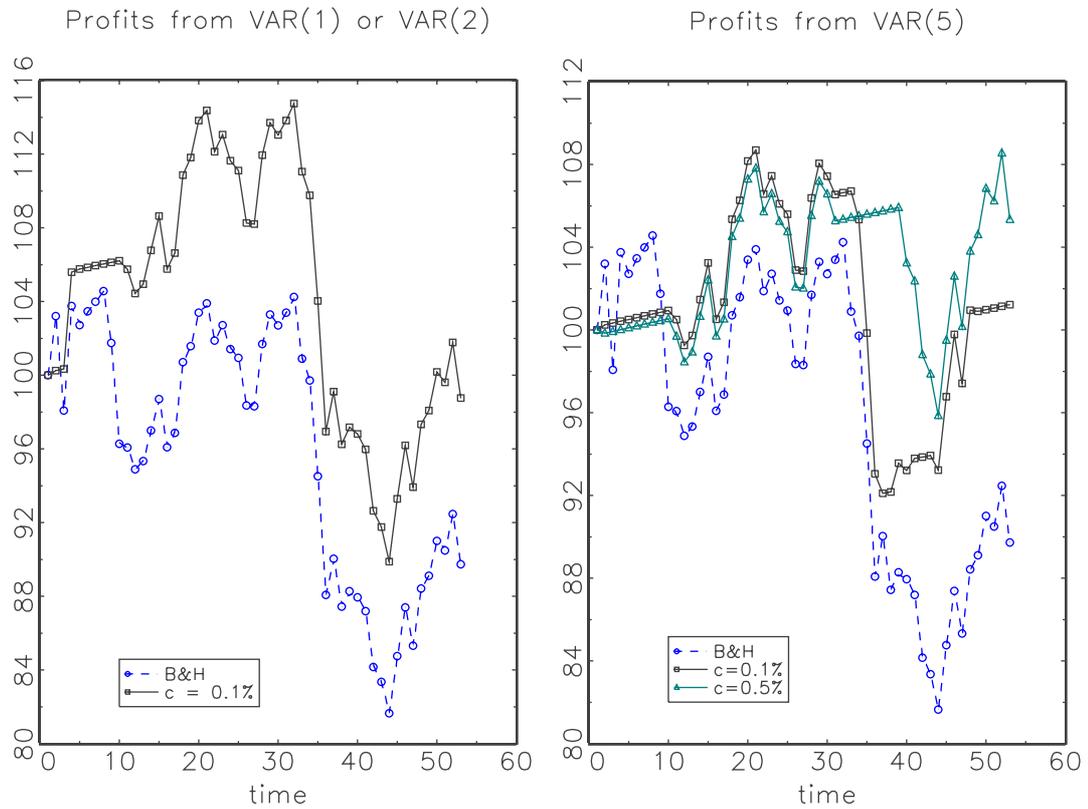


Fig. 2: Development of EUR 100 invested in the German stock market under a buy-and-hold strategy and actively managed VAR portfolios based on restricted least squares estimates under different transaction costs. The returns displayed in Table 7 are the log differences between the final value of the portfolio at the end of the out-of-sample period and the initial capital.

5 Conclusion

Using a new set of short-run and medium-run sentiment data for the German stock market, we investigated the structural properties of VAR models including these two sentiment measures as well as returns of the stock index DAX at the same frequency. In striking contrast to similar studies for the U.S. and Shanghai markets, we found that, depending on the specification of our VAR model, either sentiment is exogenous and drives returns, or returns and sentiment define a simultaneous system with mutual causation. Out-of-sample forecasting and trading exercises confirm that the causality from sentiments to returns could have been used to predict future returns and to design a market-timing strategy based on the sign and size of expected returns. While we recover predictability and profitability only for constrained VAR models estimated via least squares, this is the theoretically preferred structure of our VAR estimation given the many insignificant parameters and the results of specification tests. Our most successful specification -the restricted VAR(5) estimated via least squares- could be seen as the adaptation of a straightforward general-to-specific approach to our setting¹³. However, although the success of our VAR models in the forecasting and trading exercises depends crucially on an appropriate specification, the results on causality are extremely robust. In particular, all model specifications allow rejection of exogeneity of returns at any conventional level of significance.

However one might look at these results, the overall message is that of a surprising degree of informational inefficiency in the German stock market. Apparently, the anonymously collected sentiments of a large number of individual and institutional investors in this survey has produced an overall indicator that has significant predictive power for near-term returns. What is also surprising is that sentiment is exogenous at least up to lag four while in many related studies in the literature strong causation from returns on sentiment is found. Note also that our trading experiments show that the sentiment data can not be interpreted as an inflow of information on Sunday evening whose release is incorporated into prices as soon as trading starts

¹³See Mizon (1995) for the general methodology and Hoover and Perez (1999) for Monte Carlo evidence on successful specification search under this approach.

again on Monday morning. Taking into account Monday opening quotation still indicates profitable trading opportunities by shifting portfolios according to the sentiment signals.

Although the news contained in sentiment is not incorporated immediately into prices, exogeneity of medium-run sentiment suggests that its predictive power partially reflects information on fundamental factors. However, while there might be an element of aggregation of dispersed information in M-Sent, its high auto-correlation makes an interpretation along the informational efficiency paradigm cumbersome. Since M-Sent itself is highly predictable, it can certainly not be interpreted as a measure of *new* fundamental information. It rather appears like a slowly moving basic mood of the market that changes in a nearly autonomous fashion and only has very weak links to returns and short-run sentiment at high lags. However, identification of such a slow-moving basic mood is hardly reconcilable with any notion of efficiency and is also hard to square with the traditional noise trader story.

In contrast, short run sentiment seems more in harmony with noise trader models: it performs wild, short-lived swings between euphoria and depression and, therefore, gets closer to the stochastic noise component in, for example, DeLong et al. (1990). However, it also is a largely autistic component: Although it gets itself a strong impetus from M-Sent, its feedback on M-Sent and returns is tenuous. In the VAR(5) model, we only find a small negative feedback from short-run sentiment on M-Sent at lag 5 ($\alpha_{21}^{(5)} \approx -0.09$) and no direct effect from S-Sent on returns themselves. Despite this fragile feedback, it seems that considering these subtle interactions at long lags helps to forecast returns. Overall, the incompatibility of our findings with standard asset pricing theories calls for a behavioral explanation that goes beyond the existence of noise trading risk.

References

- [1] Clark, T. and K. West, Approximately normal tests for equal predictive accuracy in nested models, *Journal of Econometrics* 138, 2007, 291-311.
- [2] Black, F., Noise, *Journal of Finance* 41, 1986, 529-543.
- [3] Brown, G. and M. Cliff, Investors sentiment and the near-term stock market, *Journal of Empirical Finance* 11, 2004, 1-27.
- [4] DeLong, J., A. Schleifer, L. Summers and R. Waldman, Noise trader risk in financial markets, *Journal of Political Economy* 98, 1990, 703-738.
- [5] Hoover, K. and S. Perez, Data mining reconsidered: Encompassing and the general-to-specific approach to specification search, *Econometrics Journal* 2, 1999, 167-191.
- [6] Kling, G. and L. Gao, Chinese institutional investors' sentiment, *International Financial Markets, Institutions and Money* 18, 2008, 374-387.
- [7] Lee, W., C. Jiang and D. Indro, Stock-market volatility, excess returns, and the role of investor sentiment, *Journal of Banking and Finance* 26, 2002, 2277-2299.
- [8] Lütkepohl, H., *New Introduction to Multiple Time Series*. Springer, Berlin 2005.
- [9] Mizon, G., Progressive modelling of macroeconomic time series: The LSE methodology, in K. Hoover, ed., *Macroeconometrics: Developments, Tensions and Prospects*, pp. 107-170. Kluwer, Boston 1995.
- [10] Neal, R. and S. Wheatley, Do measures of investors sentiment predict returns?, *Journal of Financial and Quantitative Analysis* 33, 1998, 523-547.
- [11] Schmeling, M., Institutional and individual sentiment: Smart money or noise trader risk?, *International Journal of Forecasting* 23, 2007, 127-145.

- [12] Verma, R. and P. Verma, Noise trading and stock market volatility, *Journal of Multinational Financial Management* 17, 2007, 231-243.
- [13] Verma, R., H. Baklaci and G. Soydemir, The impact of rational and irrational sentiments of individual and institutional investors on DIJA and S&P500 index returns, *Applied Financial Economics* 18, 2008, 1303-1317.
- [14] Wang, Y.-H., A. Keswani and s. Taylor, The relationship between sentiment, returns and volatility, *International Journal of Forecasting* 22, 2006, 109-123.