Nonlinear Error Correction and the Efficient Market Hypothesis:

The Case of German Dual-Class Shares

Jörg Breitung and Christian Wulff
Humboldt University Berlin
Spandauer Strasse 1
D-10178 Berlin, Germany

August 1999
JEL classification: *C22, C51*

Abstract

The efficient market hypothesis implies that asset prices cannot be cointegrated. On the other hand, arbitrage processes prevent prices of fundamentally related assets from drifting far away. An attractive model that reconciles these two conflicting facts is the nonlinear error correction mechanism (ECM). Such a process tolerates small deviations from the long run relationship. For more substantial deviations, an effective adjustment process pushes the diverging prices towards their fundamental relationship. In this paper parametric and nonparametric techniques are employed to investigate the ECM between prices of voting and non-voting stocks. Despite its intuitive appeal, we find little evidence for a nonlinear relationship between German dual-class shares. Only in 4 out of 12 cases, the threshold ECM yields a substantial improvement of fit. In other cases, the evidence for nonlinearity is rather weak and the threshold ECM fails to outperform the linear model.

The research for this paper was carried out within "Sonderforschungsbereich 373" at the Humboldt University Berlin. We thank Rolf Tschernig for helful comments and suggestions.

1 Introduction

In an early paper on cointegration Granger (1986, p. 218) states that "[i]f x_t, y_t are a pair of prices from a jointly efficient, speculative market, they cannot be cointegrated." The reason is that whenever two variables are cointegrated, there exists an error correction representation so that at least one variable can be forecasted by using the lagged error correction term.

Recently, however, this statement was called into question in a number of studies. For example, Kasa (1992) finds evidence for common stochastic trends (and thus of cointegration) in international stock markets. Kehr (1997) shows that stock prices traded on different regional markets in Germany are cointegrated and Krämer (1997) and Dittmann (1998) find (fractional) cointegration between different classes of stocks of the same or very similar German companies.

In this paper we consider the possibility that these conflicting views of the efficient market hypothesis can be resolved by assuming a *nonlinear* error correction mechanism (ECM). The idea is that small deviations from the long run relationship are not predictable as claimed by Granger (1986). If the deviations become large, however, an effective adjustment process prevents stock prices of fundamentally related assets from drifting too far away. The economic reason behind such a nonlinear adjustment process is that transaction costs make it unprofitable to exploit small deviations from the fundamental relationship. When undervaluation (or overvaluation) becomes more substantial, agents will buy (or sell) the respective assets until the fundamental relationship is re-established. This reasoning naturally leads to a nonlinear version of the error correction model introduced by Engle and Granger (1987). In our empirical application, the nonlinearities among dual-class shares from six German companies are investigated using nonparametric and parametric techniques.

The paper is organized as follows. In section 2 we formalize Granger's (1986) statement cited above and introduce the class of nonlinear ECMs. Using a simple Monte Carlo experiment we demonstrate that if the data is generated by a threshold ECM, usual cointegration tests may have serious difficulties to detect the underlying long run relationship. In section 3 a variety of cointegration tests is applied to the full sample. Since these tests produce mixed results we analyse in section 4 the structural stability of the cointegration relationship. It turns out that in four of six cases the cointegration relationship appears to change during the sample so that in appropriate subsamples, the evidence for cointegration is much higher. In section 5 we apply nonlinearity tests and nonparametric kernel estimators to investigate possible nonlinearities in the short-run adjustment process. For the five cases that reveal substantial nonlinearities a threshold ECM is estimated in section 6. Using a bootstrap procedure we find that the excess returns from a trading rule based on the threshold ECM are economically and statistically significant. Some conclusion from our empirical analysis are offered in section 7.

2 Predictability in Cointegrated Models

Let (x_t, y_t) be a vector of two time series with a triangular representation (cf Phillips 1991)

$$\begin{aligned} y_t &= \beta x_t + \varepsilon_t && (1) \\ \Delta x_t &= v_t \ , && (2) \end{aligned}$$

where Δ is the difference operator such that $\Delta x_t = x_t - x_{t-1}$. The error terms ε_t and v_t are assumed to have "short memory". That is, they are stationary or admit some similar requirements like the mixing conditions of Phillips (1987). This framework is more general than the more familiar assumption of normally distributed stationary errors. It is widely acknowledged that financial time series are often heteroskedastic with a leptocurtic distribution function. Furthermore, our setup is able to cope for a wide class of nonlinear processes.

Subtracting y_{t-1} from both sides of (1) gives

$$\begin{aligned} \Delta y_t &= (y_{t-1} - \beta x_{t-1}) + \beta v_t + \varepsilon_t && (3) \\ &= \beta v_t + \varepsilon_t - \varepsilon_{t-1} \ , \end{aligned}$$

which shows that Δy_t can be forecasted by using the "error correction term" $\varepsilon_{t-1} = y_{t-1} - \beta x_{t-1}$. If the errors v_t or ε_t are correlated with ε_{t-1} we write (3) in form of a regression model:

$$\Delta y_t = \gamma(y_{t-1} - \beta x_{t-1}) + u_t \ , \qquad (4)$$

where $\gamma = E(\Delta y_t \varepsilon_{t-1})/E(\varepsilon_{t-1}^2)$ and u_t is uncorrelated with $(y_{t-1} - \beta x_{t-1})$. This model will be referred to as the *linear ECM.*

On capital markets there is reason to expect that the ECM is nonlinear (e.g. Martens et al. 1998). Due to transaction costs arbitragers will not completely exploit all differences in prices of the same asset. That means that small deviations are "tolerated", but a reversion to the fundamental relationship occurs if the deviation becomes too large. Such a relationship can be translated into a nonlinear ECM of the form

$$\Delta y_t = f(y_{t-1} - \beta x_{t-1}) + u_t \tag{5}$$

by assuming that the function $f(\cdot)$ is close to zero for small values of $\varepsilon_{t-1} = y_{t-1} - \beta x_{t-1}$, whereas for larger deviations from the long run relation, the function takes large negative values. If $f(\cdot)$ is a step function, then y_t is called a "threshold ECM". Such models are considered in Balke and Fomby (1997), Martens et al. (1998) and Enders and Siklos (1999). For this kind of models the nonlinear function is specified as

$$f(z) = \begin{cases} -\gamma_1 z & \text{for } z < \tau_1 \\ 0 & \text{for } \tau_1 \le z \le \tau_2 \\ -\gamma_2 z & \text{for } z > \tau_2 \end{cases}$$

This function implies that if the equilibrium error $\varepsilon_t = y_t - \beta x_t$ is in the interval $[\tau_1, \tau_2]$ then there is no adjustment towards the long run relationship. If however the process exceeds the limits τ_1 or τ_2 then the process pushes the variables towards the long-run relationship.

In practice, such a nonlinear ECM may be difficult to detect. As demonstrated by Balke and Fomby (1997), for large absolute values of τ_1 and τ_2 the process stays in the nonstationary region for a long time and thus, cointegration tests have serious difficulties to indicate a cointegration relationship.

To illustrate the loss of power against threshold alternatives, a Monte Carlo experiment is performed using the following symmetric threshold ECM:

$$\Delta y_t = f(y_{t-1} - x_{t-1}) + u_t \tag{6}$$

where

$$f(\varepsilon_{t-1}) = f(y_{t-1} - x_{t-1}) = \begin{cases} -\gamma \varepsilon_{t-1} & \text{for } |\varepsilon_{t-1}| > \tau \\ 0 & \text{for } |\varepsilon_{t-1}| \le \tau \end{cases}$$

and u_t is distributed as $N(0,1)$. Furthermore x_t is a random walk with $\Delta x_t \sim N(0,1)$. To make the threshold value comparable for different values of γ,

Table 1: Power against linear and nonlinear alternatives

τ^*	$\gamma = 0.02$	$\gamma = 0.05$	$\gamma = 0.1$
0.0	0.292	0.963	1.000
0.5	0.265	0.959	1.000
1.0	0.172	0.893	0.999
1.5	0.096	0.634	0.997
2.0	0.066	0.303	0.889
2.5	0.064	0.151	0.458
3.0	0.048	0.104	0.259

Note: The entries of the table report the empirical power of the Dickey-Fuller cointegration test for a model given in (6). The 0.05 critical value 2.76 is used. The rejection frequencies are computed from 1000 replications of the model with $T = 500$.

the threshold is measured relative to the variance of the equilibrium errors $\varepsilon_t = y_t - x_t$ in a linear specification, that is, $\tau = \tau^* \sigma_\varepsilon = \tau^* \sqrt{2/[1 - (1 - \gamma^2)]}$.

The results of the Monte Calo experiment with $T = 500$ and 1000 replications of the model are presented in Table 1. To test the null hypothesis that there is no cointegration between y_t and x_t, the Dickey-Fuller test is applied to the residuals of the cointegration regression. The significance level of the test is 0.05. For $\tau = 0$ the data generating process is linear and thus this specification can be used as a benchmark to asses the loss of power due to the nonlinear form of ECM.

The results reported in Table 1 suggest that the loss of power is small for small values of τ but can be dramatical for large values of τ. For example, for the linear ECM with $\gamma = 0.05$ and $\tau^* = 0$ the power of the test is close to one, whereas the power of the threshold ECM with $\tau^* = 2$ is 0.3 and for $\tau^* = 3$ the power is only slighly higher than the significance level. This outcome demonstrates that the power of a (linear) cointegration test against nonlinear alternatives may be poor for high absolute values of the threshold parameter.

3 Cointegration tests for voting and non-voting stocks

In this section we apply a variety of cointegration tests to time series data of pairs of German voting and non-voting shares issued by the same firm (dual-

Table 2: Details of the dual-class shares

Firm	Abbrev.	No. obs.	Sample range	Index	Div. adv.
RWE	RWE	5894	1/2/74 – 7/31/97	DAX	5 / –
MAN	MAN	5588	4/24/75 – 7/31/97	DAX	4 / –
BMW	BMW	1983	8/25/89 – 7/31/97	DAX	2 / –
Volkswagen	VW	2706	10/6/86 – 7/31/97	DAX	4 / 2
Rheinmetall	RHM	3182	10/31/84 – 7/31/97	MDAX	6 / 2
Boss	BOSS	2052	5/22/89 – 7/31/97	MDAX	3 / 3

Note: Dividend advantage ("Div. adv.") is expressed in percent of par value, whereby the first figure indicates the minimum dividend and the excess dividend is given after the slash.

class firm). Daily stock price data adjusted for stock splits, dividends and other corporate events are from the *"Deutsche Finanzdatenbank"* (DFDB) in Karlsruhe. We use logarithms of stock prices in our statistical analysis. According to the criteria liquidity and availability of long time series a sample of 6 dual-class firms is chosen for examination. Among those four are contained in the index of the 30 largest German blue-chip stocks (DAX), the remaining two are in the German mid-cap index MDAX. In Table 2 the sample is further described. German corporate law requires that holders of non-voting shares must be compensated for the lack of corporate control by a dividend advantage. This usually takes the form of a minimum preferred dividend (stated as percentage of par value) which will be carried forward in particular years of dividend omissions (cumulative preferred dividend). Both the cumulative (past) preferred dividends and the current preferred dividends have to be paid out before the common shareholders can receive anything. In addition some firms commit themselves to pay the non-voting shareholders a certain (non-cumulative) amount in excess of the common stock dividend.

First we compute the augmented Dickey-Fuller (ADF) and Phillips-Perron (PP) test applied to the residuals of a OLS cointegrating regression (cf. Phillips and Oularis 1991). For the ADF test ten lagged differences and a constant are included in the regression. The truncation lag for the Phillips-Perron test on the residuals is set to 20. Using these tests, only a weak evidence for cointegration is found for RWE, and RHM, whereas in the other cases, the null hypothesis of no cointegration is rejected at a significance level of 0.05. Applying Johansen's (1988) likelihood-ratio test procedure based on a VAR(10) model with a constant restricted to the cointegration relationship gives a slightly different picture. This test finds a cointegration relation-

Table 3: Cointegration tests for voting and non-voting stocks

Test	RWE	MAN	BMW	VW	RHM	BOSS
ADF	–3.197	–5.638**	–6.641**	–3.551*	–2.824	–3.812*
PP	–3.744*	–8.707**	–11.19**	–4.139**	–3.517*	–5.701**
LR	21.42*	36.21**	47.14**	15.87	17.78	18.52
NP	94.15	236.6	501.38**	83.53	340.91*	302.6
Rob(0)	–16.17**	–8.828**	–2.473*	–6.789**	–8.596**	–5.931**
Rob(1)	5.856**	5.540**	4.089**	5.394**	4.821**	7.874**
$\hat{d}$	0.7837	0.7218	0.8149	0.8516	0.8499	0.7106
RDF	–22.47**	–15.43**	–10.74**	–6.859**	–7.219**	–9.380**
Rdiff	0.0005**	0.0011**	0.0024**	0.0058*	0.0047**	0.0030

Note: "ADF" denotes the augmented Dickey-Fuller test including 10 lagged differences and a constant. "PP" is the unit root test of Phillips and Perron (1988) applied to the residuals of the cointegration regression. "LR" is Johansen's trace statistic for the hypothesis that the cointegration rank is zero. "NP" denotes Breitung's (1999) variant of Bierens' (1997) nonparametric test for cointegration. "Rob(0)" and "Rob(1)" is the test for the hypothesis that the residuals of the cointegration regression are $I(0)$ and $I(1)$ against fractional alternatives (Gil-Alana and Robinson 1997). "$\hat{d}$" is the estimate of the fractional parameter using Robinson's (1995) semiparametric approach. "RDF" and "Rdiff" are rank tests for cointegration as suggested by Granger and Hallman (1991) and Breitung (1998), respectively. * and ** indicate significance with respect to the 0.05 and 0.01 signicance level, repsectively.

ship for RWE, MAN and BMW. In the remaining cases, the LR statistic is insignificant.

As argued in the previous section, these test procedures are based on fairly restrictive assumptions that are problematical in applications using financial time series data. To overcome the potential problems with parametric tests, a robust cointegration tests is applied. Here we use a variant of Bierens' (1997) nonparametric test, which is is based on the work of Vogelsang (1998). This test does not require a correction for possible short run dynamics in the system (Breitung 1999). Using this test, cointegration is found for BMW and RHM only and, thus, the evidence for cointegration is even weaker than using the parametric tests.

Following Krämer (1997) and Dittmann (1998) we next test the hypothesis that the errors of the cointegration regression are fractionally integrated. For this purpose we use the score statistics for the hypotheses $d = 0$ and $d = 1$ against fractional alternatives of the form $0 < d < 1$. The respective test statistics applied by Gil-Alana and Robinson (1997) are denoted as Rob(0) and Rob(1). Under the null hypothesis the test statistics are asymptotically standard normally distributed.[1] The results presented in Table 3 indicate that in all cases the fractional parameter is significantly larger than zero and smaller than one, that is, the tests suggest that the errors are indeed fractionally integrated. Using Robinson's (1995) semiparametric estimator for the fractional parameter we obtain estimates in the interval [0.71,0.86].

It is important to note that a rejection of a hypothesis does not imply that the alternative hypothesis is correct. Thus, we cannot conclude that the errors are in fact best described by a fractionally integrated process. Rather, the results suggest that the usual linear framework of integer-valued integrated processes is too restrictive to model the relationship between stock prices. The findings may alternatively be explained by nonlinearities in the data generating process. To investigate this possibility we apply the ranked Dickey-Fuller test (RDF) suggested by Granger and Hallman (1991) and a test based on the squared differences between the ranked series (Rdiff) as suggested by Breitung (1998). The results presented in Table 3 show that the evidence for integration increase substantially when using the ranks of the series instead of the original data. All test statics are well beyond their 1% critical values.

Summing up, standard cointegration tests do not reveal an overwhelming evidence for cointegration. Regarding the large sample sizes, borderline significance with respect to a significance level of 0.05 implies that the cointegration relationship between pairs of voting and non-voting stocks is quite weak. A notable exception is perhaps BMW, where all tests suggest a cointegration relationship even at a significance level of 0.01. Still, if we allow for a fractional difference operator or if nonlinear transformations are accommodated by applying the rank transformation, the evidence for cointegration (better: "comovement") is even more pronounced.

4 Structural Changes

A possible explanation for the mixed results reported in the previous section is that the cointegration relationship is subject to structural changes. The main reasons for different prices of dual-class shares are differences in the dividend payment, liquidity and voting rights (e.g. Kunz and Angel 1996,

[1] The test procedures and the estimator of the fractional parameter are implemented in the Software XploRe. See Härdle et al. (1999) for more details.

Table 4: Subsample cointegration statistics

Test	RWE	MAN	BMW	BOSS
ADF	–4.278**	–3.436*	–4.309**	–4.002*
PP	–6.040**	–5.715**	–10.71**	–6.798**
LR	25.17**	19.55	25.00**	26.06**
NP	362.3*	235.6	338.81*	638.4**
Rob(0)	–3.748**	–5.142**	–1.467	–1.022
Rob(1)	5.596**	7.287**	6.147**	4.381**
$\widehat{d}$	0.779	0.636	0.636	0.6253
RDF	–10.43**	–8.613**	–11.84**	–7.741**
Rdiff	0.0026**	0.0034**	0.0020**	0.0045**
Start	4000	3000	800	1200

Note: See Table 3 for an explanation of the test statistics. "Start" indicates the beginning observation number of the respective subsamples. * and ** indicate significance with respect to the 0.05 and 0.01 signicance level, repsectively.

and Gardiol et al. 1997). Accordingly, a structural change in these variables may affect the long-run relationship between dual-class shares.

To investigate the structural stability between prices of dual-class stocks, the cointegration regression is estimated sequentially so that the prices of non-voting stocks are first regressed on the prices of voting stocks by starting with the first 30 observations and then one by one a new case enters into the estimation sample until the full sample is used for parameter estimation. The resulting time paths of the sequential estimates are presented in the left panel of Figure 1. In the right panel of Figure 1, we present Hansen's sequential F-statistic for a structural change in the cointegration regression along with the critical value according to a significance level of 0.05.

It turns out that for RWE, MAN, BMW, and BOSS there is a clear indication of structural changes in the cointegrating relationship. For these cases we construct subsamples for which the sequential estimates and tests suggest a stable relation. We also experimented with including step dummies but it seems that the regimes changes gradually and, thus, cannot be appropriatly represented by dummy variables. In what follows we therefore use the most recent regimes with a stable relationship between the variables. The starting points of the subsamples can be found in Table 4. All sub-sample run up to the last observation.

For these four subsamples we perform the same estimates and tests pre-

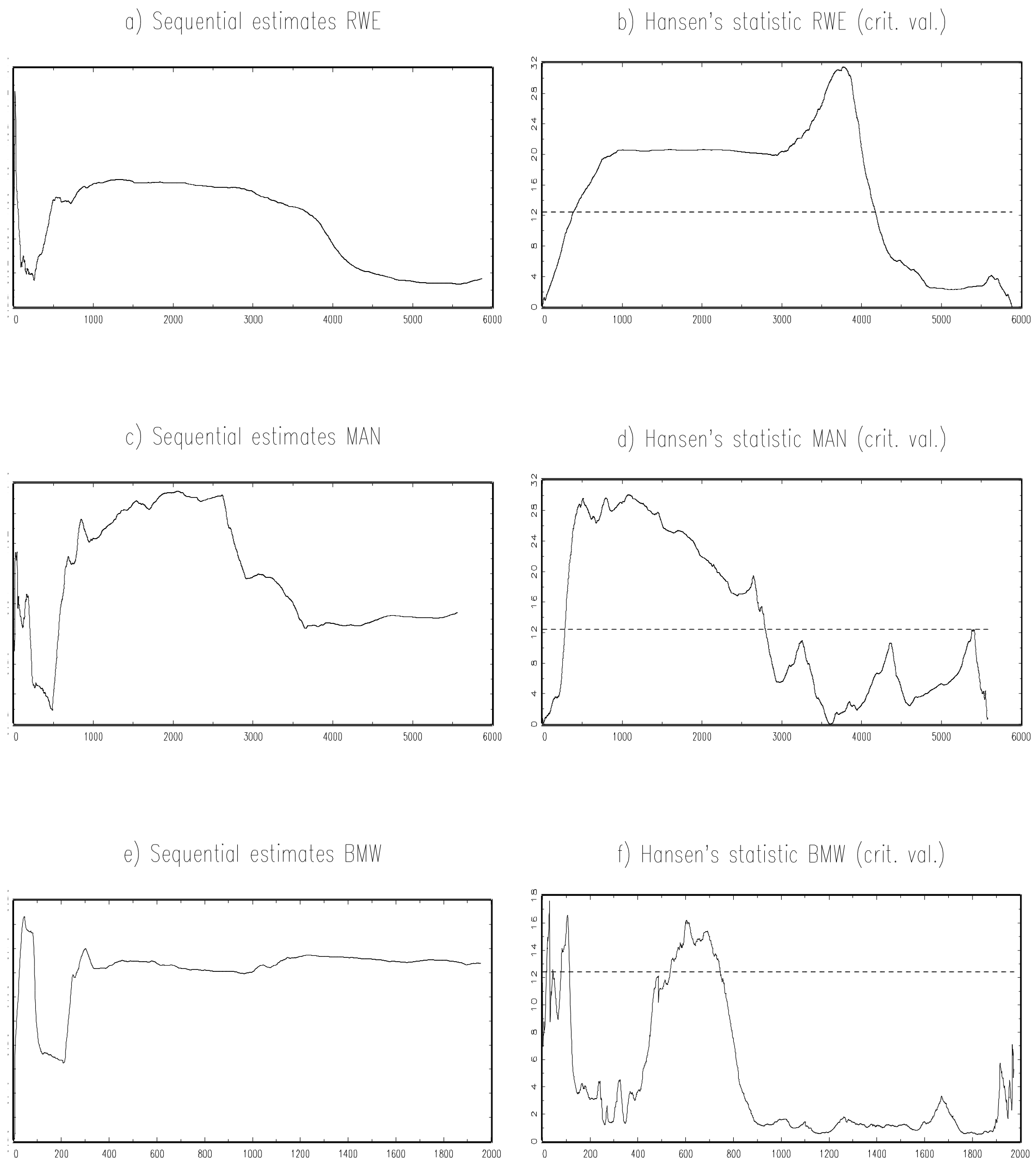
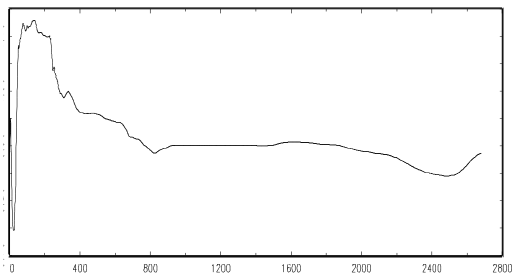
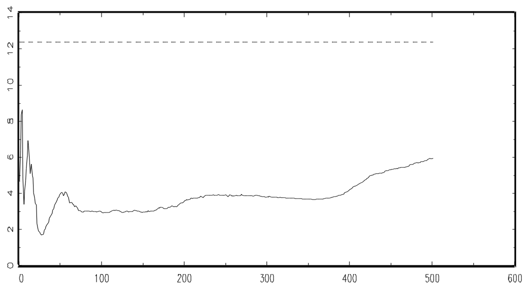


Figure 1: Structural stability of the cointegration coefficient

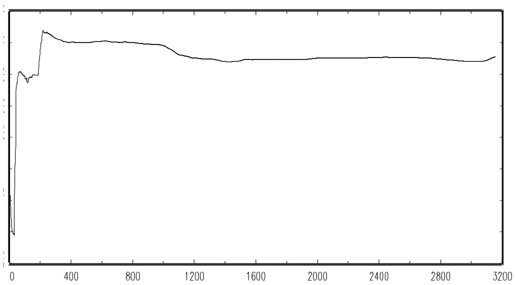
g) Sequential estimates VW



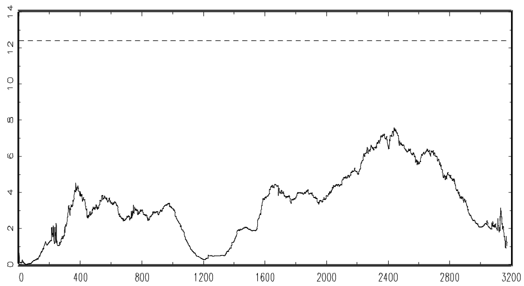
h) Hansen's statistic VW (crit. val.)



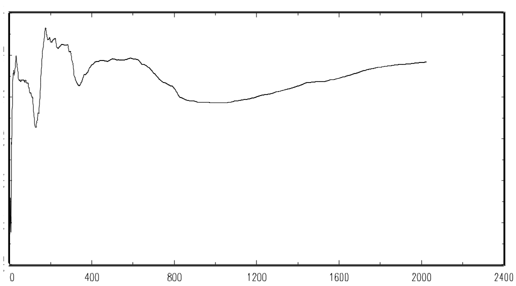
i) Sequential estimates RHM



j) Hansen's statistic RHM (crit. val.)



k) Sequential estimates BOSS



l) Hansen's statistic BOSS (crit. val.)

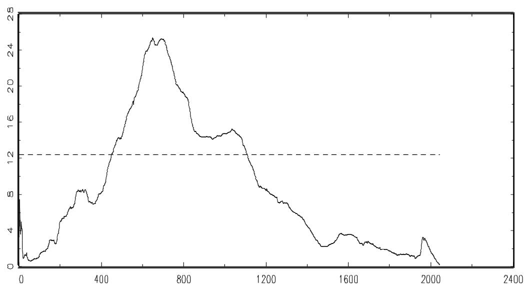


Table 5: Estimation of the cointegration relationship

stock	$\widehat{\alpha}$	$s.e.(\widehat{\alpha})$	$\widehat{\beta}$	$s.e.(\widehat{\beta})$
RWE	–0.3073	0.0439	1.0180	0.0117
MAN	0.1801	0.0678	0.9265	0.0120
BMW	–0.1046	0.0288	0.9633	0.0043
VW	0.4765	0.1085	0.8843	0.0183
RHM	0.5754	0.9672	0.9558	0.1856
BOSS	-1.2380	0.1097	1.1674	0.0147

Note: The table reports the estimates and standard errors (s.e.) of a fully-modified estimator (Phillips and Hansen 1990) for the cointegration regression $y_t = \alpha + \beta x_t + \varepsilon_t$, where y_t indicates the non-voting stock and x_t is the (log) price of the voting stock.

sented for the full sample in Table 3. As can be seen from the results in Table 4 the evidence for cointegration is much higher in the subsamples than in the full samples. With the exception of MAN, the tests reject the hypothesis of no cointegration at a significance level of 0.01.

5 Evidence for Nonlinearity

In this section we investigate the functional form between Δy_t and $(y_{t-1} - \beta x_{t-1})$. For this purpose we first estimate the parameter β in a regression of the non-voting stock prices on voting stock prices and a constant term using the "fully-modified" approach of Phillips and Hansen (1990). This estimator employs a nonparametric correction to the least-squares estimator in order to obtain an efficient asymptotically normally distributed estimator. For the correction a Parzen kernel with a truncation lag of 20 is applied. The resulting estimates are presented along with their asymptotic standard errors in Table 5.

If voting and non-voting stocks possess a constant price ratio, for example if there exists a "voting premium" of a fixed percentage, then $\beta = 1$. From the results presented in Table 5, such a constant voting premium can be rejected for all but RWE and RHM, as the estimates of β are significantly different from unity. For BOSS the relative voting premium increases with stock prices whereas for MAN, BMW, VW the voting premium decreases with increasing stock prices. Similar results were found by Dittmann (1998) by using an ordinary least-squares estimator.

With the estimated cointegration parameter we construct the lagged error correction term according to

$$z_{t-1} = y_{t-1} - \widehat{\alpha} - \widehat{\beta} x_{t-1} \ ,$$

where the first set of regressions uses voting stock prices for y_t and non-voting stock prices for x_t. The second set of regression interchanges the variables so that y_t are non-voting stock prices and x_t are the respective prices of the voting stock. For the second set of regressions we use $\widehat{\beta}^* = 1/\widehat{\beta}$ as the cointegration parameter.

Following Tsay (1989) we re-arrange the data according to the threshold error correction term such that $z_{t^*+1} \geq z_{t^*}$, where t^* is the new "time index". Using the re-arranged sample we are able to test the hypothesis that the relationship between Δy_t and z_{t-1} is linear. The idea is, that by ordering the observations with respect to z_{t-1}, a nonlinear functional form will induce structural breaks in the relationship between Δy_t and z_{t-1}. Hence, the usual tests for a structural break can be used. Specifically, we apply the Chow test procedure with an unknown break point (see Andrews 1993). This procedure is based on the maximum of Chow's LR-statistic in a given range of time periods. Here we compute the LR-statistics for the interval [0.05T,0.95T], which is the widest range considered in Andrews (1993). The test statistic is the supremum of the sequence of LR-statistics and will be denoted by "sup-LR" in Table 6.

Approximating the nonlinear relationship by a fourth order polynomial in z_{t-1}, we can use a Wald statistic to test the higher order terms in the polynomial against zero. The results of this parametric test are presented in the column "Wald(poly4)" in Table 6. Another test for nonlinearity is the neural network test proposed by Lee, White and Granger (1993). This test is derived from the squared multiple correlation of an auxiliary regression of the residuals from a linear regression on the regressors and the 3 principal components of a nonlinear transformation of the regressors.[2] Under the null hypothesis of a linear relationship, the test statistic is χ^2 distributed with 3 degrees of freedom.

From the results presented in Table 6 it turns out that in some cases there is clear evidence of a nonlinear relationship. In particular, for the two stock classes of MAN and BOSS as well as the voting share of RHM, all tests

[2]The test statistic is calculated using XploRe macro `annlintest`. See Härdle et al. (1999) for more details

Table 6: Tests for nonlinearity

stock	sup-LR	Wald(poly4)	Neural Net
RWE(vote)	6.124	0.125	0.124
RWE(nonv)	2.408	0.796	0.816
MAN(vote)	11.92*	9.388*	9.376*
MAN(nonv)	16.94**	17.57**	17.48**
BMW(vote)	5.594	2.313	2.317
BMW(nonv)	7.883	5.615	5.607
VW(vote)	10.86	3.024	3.103
VW(nonv)	11.29	2.196	2.208
RHM(vote)	19.42**	19.53**	19.18**
RHM(nonv)	6.260	2.995	2.996
BOSS(vote)	15.69**	11.59**	11.59**
BOSS(nonv)	18.50**	11.53**	11.41**

Note: The column "sup-LR" presents the reports the supLR statistic of Andrews (1993). The column indicated by "LR(poly4)" presents the Wald statistic for linearity against a fourth order polynomial. The statistic is χ^2 distributed with 3 degrees of freedom. The column labelled as "rank test" presents the t-statistic for the addition of the ranks of z_{t-1}. "*" and "**" indicate significance at the levels 0.05 and 0.01, respectively.

indicate departures from the linear model. For VW, Andrews' test against structural breaks indicates a nonlinear relationship. However, since the test statistic only slightly exceeds the 0.05 critical value and the other two tests are well below their critical values, we conclude that for VW the evidence for nonlinearity is rather weak.

For the five cases, where the nonlinearity tests yield a substantial indication of nonlinearity, we employ the Nadaraya-Watson kernel estimator to obtain a nonparametric estimate of the function $f(\cdot)$ in

$$\Delta y_t = f(z_{t-1}) + u_t \ . \tag{7}$$

The bandwith is set to $h = 0.1$ which appears to yield an appropriate compromise between smoothness and fit of the regression function.[3] From the estimates presented in Figure 2 it appears that only MAN and RHM exhibit a nonlinear relationship that resembles to some extent a threshold process. In any case the approximation by a threshold ECM appears to be of limited value. In the next section we neverless estimate a threshold ECM for all these five cases.

[3] To compute the estimates we use the XploRe macro `regestp`. See Härdle et al. (1999) for more details.

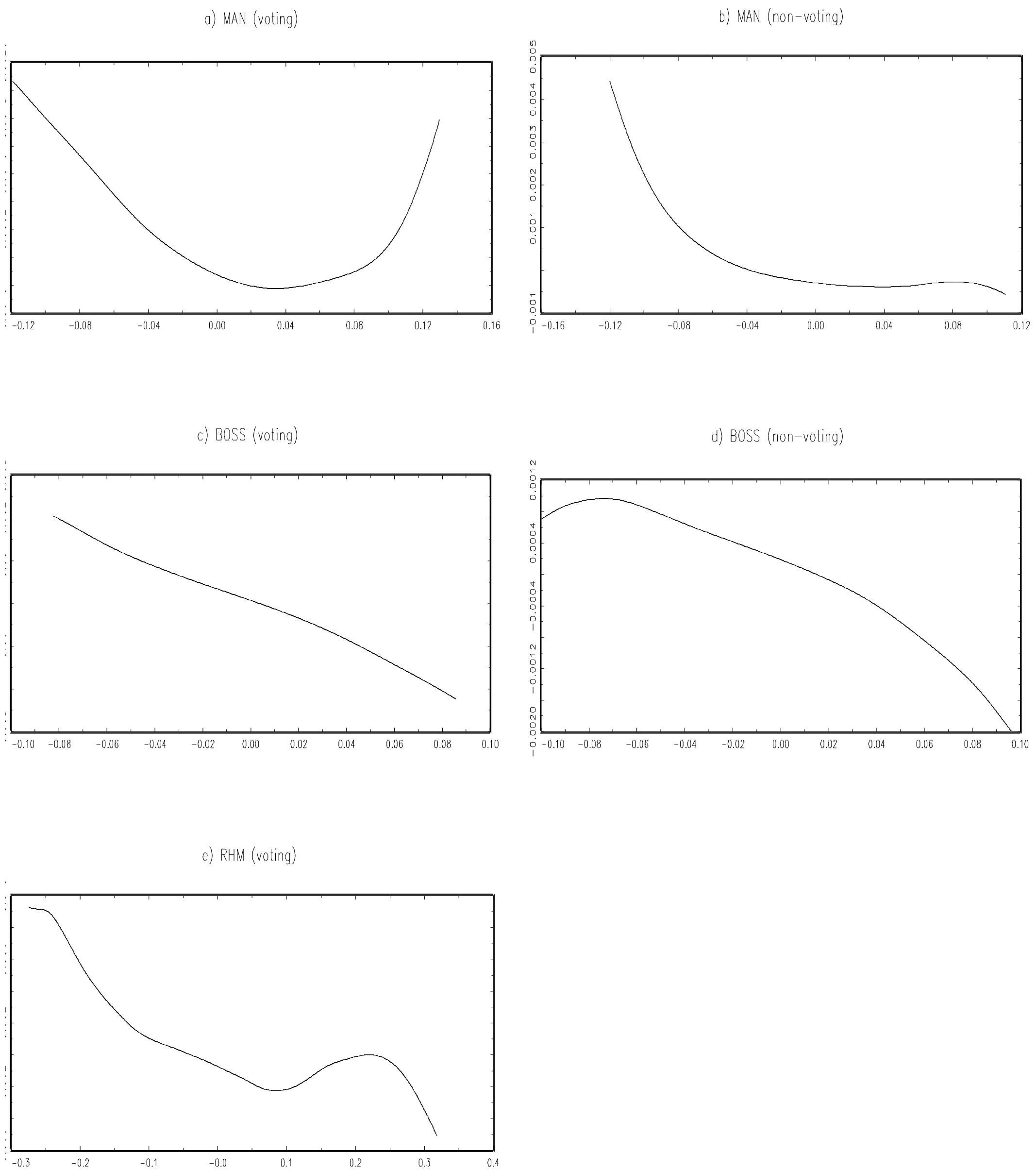


Figure 2: Nonparametric estimates of the functional form

6 Threshold ECM

In this section we present ML estimates of the threshold ECM and assess the economic relevance of such models. If the price of the voting stock is too low relative to the price of the non-voting stock, arbitragers will buy the voting stocks and sell short the non-voting stocks. If the voting stock is too high, they will do the opposite. In this case there are two threshold levels but an effective adjustment process is also possible with a single threshold for each stock price. A single threshold implies that investors only take one-sided (unhedged) positions. A full arbitrage strategy requires the possibility of short-selling dual-class shares, (which exists, if at all, for only one of the dual-class shares). Indeed, in all cases we found only a single threshold level. We therefore present the results of the single threshold model given by

$$\Delta y_t = f_{\tau_j}(z_{t-1};\gamma) + u_t \quad j = 1 \text{ or } j = 2 \tag{8}$$

where

$$f_{\tau_1}(z_{t-1};\gamma_1) = \begin{cases} -\gamma_1 z_{t-1} & \text{for } z_{t-1} < \tau_1 \\ 0 & \text{for } z_{t-1} \geq \tau_1 \end{cases}$$

$$f_{\tau_2}(z_{t-1};\gamma_2) = \begin{cases} 0 & \text{for } z_{t-1} < \tau_2 \\ -\gamma_2 z_{t-1} & \text{for } z_{t-1} \geq \tau_2 \end{cases} .$$

To estimate the unknown parameters τ_j and γ_j, a ML procedure is applied. For $j = 1$, the log-likelihood function can be written as

$$\mathcal{L}_1(\gamma_1, \tau_1) = -\frac{T}{2}\log\, \hat{\sigma}^2(\gamma_1, \tau_1)$$

where the constant term is omitted and

$$\hat{\sigma}^2(\gamma_1, \tau_1) = \frac{1}{T}\sum_{t=1}^{T} \mathbb{I}[z_{t-1} < \tau_1](\Delta y_t + \widehat{\gamma}_1 z_{t-1})^2 .$$

In this expression, $\mathbb{I}[z_{t-1} < \tau_1]$ represents the indicator function which is one if the argument is true and zero otherwise. Conditional on τ_1 the likelihood function is easily maximized by the least-squares estimator of the subsample with $z_{t-1} < \tau_1$. Accordingly, a grid search in the interval $\tau_1 \in [min(z_{t-1}), 0]$ may be adopted to obtain the minimum of $\hat{\sigma}^2$.

In Table 7, the ML estimates of τ_j and γ_j are given. The standard errors for γ_j are estimated conditional on the threshold value. Since the likelihood

Table 7: ML estimation of the threshold models

	MAN(vote)	MAN(nonv)	BOSS(vote)	BOSS(nonv)	RHM(vote)
$\widehat{\tau}_1$	–0.041	–0.070	—	—	–0.061
$\widehat{\tau}_2$	—	—	—	0.065	—
$\widehat{\gamma}_j$	0.017	0.022	0.068	0.034	0.014
s.e.($\widehat{\gamma}_j$)	0.0084	0.0093	0.01556	0.0158	0.0032
R^2(lin)	0.0016	0.0021	0.0211	0.0054	0.0060
$R^2(\tau_j)$	0.0048	0.0087	—	0.0309	0.0110
LR	8.254	16.94	—	22.11	15.96
obs. outside	541	60	—	5	1012
ExRet	0.0024	0.0115	—	0.0345	0.0026
p-value	0.0009	0.0000	—	0.0000	0.0000

Note: For the model with the returns of the voting share as dependent variable we write "(vote)", whereas for the non-voting share we write "(nonv)". $\widehat{\tau}_j$ and $\widehat{\gamma}_j$ denote the ML estimator using a grid search for maximizing the likelihood function. $s.e.(\widehat{\gamma}_j)$ is the standard errors conditional on the estimated threshold level. R^2(lin) is the coefficient of determination in a linear ECM and $R^2(\tau_j)$ is the R^2 of a threshold ECM. "LR" denotes the likelihood ratio statistic of the hypothesis that $\tau_1 = \infty$ or $\tau_2 = -\infty$. "obs. outside" indicates the number of observations outside the range determined by the threshold, i.e. $|z_{t-1}| > |\tau_j|$. "ExRet" denotes the mean excess returns implied by the threshold model.

function is not differentiable with respect to τ_j, the usual asymptotic standard errors cannot be computed. For the non-voting share of BOSS the likelihood function of the threshold model is always below the likelihood of the linear model and, thus, the results for the linear ECM are presented in Table 6.

The improvement of fit due to the threshold function can be assessed by comparing R^2(lin) of a linear ECM and $R^2(\tau_j)$ from the threshold ECM. The gain is in particular important for the non-voting shares of BOSS. For this case a positive threshold value (τ_2) performs better than a negative threshold, whereas in all other cases, a negative threshold (τ_1) is preferable.

A statistic that compares the log-likelihood value of a linear and a non-linear specification is the likelihood ratio test statistic given by

$$LR = T \log \frac{1 - R^2(\text{lin})}{1 - R^2(\tau_j)}$$

This test statistic tests the null hypotheses of a linear ECM, that is, $\tau_1 = \infty$ or $\tau_2 = -\infty$, respectively. However, it is not clear whether this test statistic has the usual χ^2 distribution, because the threshold parameter is not identified under the alternative hypothesis (Hansen 1996). Nevertheless, the test statistic may be used as an indicator for the improvement of fit due to the threshold specification.

From the LR statistics presented in Table 7 it can be concluded that for the non-voting shares of MAN, BOSS and RHM the improve of fit is substantial and applying the usual critical values (which may be inappropriate here) the threshold ECM seems to yield a statistically significant improve of fit.

Next we investigate economic significance of the estimated model. To this end we compute the mean return which is realized by applying the arbitrage strategy outlined before, whenever the ECM term is below the lower threshold τ_1 or above the upper threshold τ_2. To obtain the excess returns ("ExRet" in Table 7), we compare the resulting mean return from this trading strategy to the expected return, calculated as the mean of the sample returns of the respective share.

All excess returns are positive and tend to increase as the number of days outside the thresholds decrease. To test whether these excess returns are significant the following bootstrap procedure is applied. Consider for example the voting share of MAN. Since 541 values of the ECM term are below the threshold of -4.1 percent, we draw by chance and with replacement 10.000 samples of 541 returns from the 2588 observations that are used to estimate the threshold model. Then we compute the relative frequency of mean returns that are higher than the observed excess return of 0.24 percent. Since we only find 9 out of 10000 cases with a higher mean return than 0.24 percent, the $p - value$ results as 0.0009. From the results of this bootstrap procedure we conclude that the positive excess returns are very unlikely to be observed by chance alone.

7 Conclusion

It has been argued that in principle the nonlinear ECM is able to reconcile conflicting facts concerning the efficient market hypothesis. It allows small and unpredictable deviations from the long run relationship between fundamentally related assets such as dual class shares. If the deviations become large the nonlinear error correction mechanism pushes the stock prices towards their long run relationship.

Despite its intuitive appeal, we only find limited evidence for such a nonlinear relationship between German dual-class shares. Only for the non-voting stocks of MAN, BOSS and RHM, the threshold ECM yield a substantial improvement of fit. In other cases, the evidence for nonlinearity is weak and, thus, the threshold ECM fails to outperform the linear model.

Our results indicate that trading rules based on the long-run relationship between dual-class shares appear to have only a limited potential. In the best case (the non-voting share of BOSS) the in-sample R^2 is roughly 0.03, and, therefore, the predictive ability of the nonlinear ECM is quite poor. Taking transaction costs into account, there does not seem to be a reasonable basis for exploiting the deviations from the long-run relationship between dual-class share prices. In other words, our results suggest that the German market for dual-class shares is practically efficient although statistical tests produce some indication for a (possibly nonlinear) cointegrating relationship.

References

Andrews, D.W.K. (1993), Tests for Parameter Instability and Structural Change with Unknown Change Point, Econometrica, 821–856.

Balke, N.S. and T.B. Fomby (1997), Threshold Cointegration, International Economic Review, 38, 627–645.

Bierens, H.J. (1997), Cointegration Analysis, in: C. Heij, H. Schumacher, B. Hanzon, and C. Praagman (eds), System Dynamics in Economic and Financial Models, New York: John Wiley.

Breitung, J. (1998), Rank Tests for Nonlinear Cointegration, Humboldt University Berlin, SFB 373 Discussion Paper No. 98-65.

Breitung, J. (1999), Some Nonparametric Tests for Unit Roots and Cointegration, Humboldt University Berlin, SFB 373 Discussion Paper No. 99-36.

Dittmann, I. (1998), Fractional Cointegration of Voting and Non-Voting Shares, Discussion Paper, Universität Dortmund, Nr. 98-02.

Enders, W. and P.L.Siklos (1999), Cointegration and Threshold Adjustment, Iowa State University, Feb. 1999.

Engle, R.F. and C.W.J. Granger, C.W.J. (1987), Co-integration and Error Correction: Representation, Estimation, and Testing, Econometrica, 55, 251–276.

Gardiol, L., R. Gibson-Asner, and N.S. Tuchschmid (1997), Are Liquidity and Corporate Control Priced by Shareholders? Empirical evidence from Swiss dual class shares, Journal of Corporate Finance 3, 299–323.

Gil-Alana, L.A. and P.M. Robinson (1997), Testing of Unit Root and Other Nonstationary Hypotheses in Macroeconomic Time Series, Journal of Econometrics, 80, 241–268.

Granger, C.W.J. (1986), Developments in the Study of Cointegrated Economic Variables, Oxford Bulletin of Economics and Statistics, 48, 3, 213–228.

Granger, C.W.J. and J. Hallman (1991), Long Memory Processes With Attractors, Oxford Bulletin of Economics and Statistics, 53, 11–26.

Härdle, W., S. Klinke and M.M. Müller (1999), XploRe Learning Guide, Berlin: Springer.

Hansen, B. (1996), Inference When A Nuisance Parameter is not Identified Under the Null Hypothesis, Econometrica, 64, 413–430.

Johansen, S. (1988), Statistical Analysis of Cointegration Vectors, Journal of Economic Dynamics and Control, 12, 231–254.

Kasa, K. (1992), Common Stochastic Trends in International Stock Markets, Journal of Monetary Economics, 29, 95–124.

Kehr, L.-H. (1997), Preisfindung bei verteilter Börsenstruktur, Wiesbaden: DUV.

Kunz, R. and J.J. Angel (1996), Factors Affecting the Value of the Voting Right: Evidence from the Swiss Equity Market, Financial Management, 25, 7-20.

Krämer, W. (1997), Kointegration von Aktienkursen, Universität Dortmund, mimeo.

Lee T.-H., H. White, C.W.J. Granger (1993), Testing for Neglected Nonlinearity in Time Series Models: a comparison of neural network methods and alternative tests, Journal of Econometrics, 56, 269–290.

Martens, M., P. Kofman, and T.C.F. Vorst (1998), A Threshold Error-Correction Model for Intraday Futures and Index Returns, Journal of Applied Econometrics, 13, 245–263.

Phillips, P.C.B. (1987), Time Series Regression With a Unit Root, Econometrica, 55, 277–301.

Phillips, P.C.B. (1991), Optimal Inference in Cointegrated Systems, Econometrica, 59, 282–306.

Phillips, P.C.B. and B.E. Hansen (1990) Statistical Inference in Instrumental Variable Regression with I(1) Processes, Review of Economic Studies, 57, 99–125.

Phillips, P.C.B. and S. Oularis (1990), Asymptotic properties of residual based tests for cointegration, Econometrica, 58, 165–193.

Phillips, P.C.B. and P. Perron (1988), Testing for a Unit Root in Time Series Regression, Biometrika 75, 335–346.

Robinson, P.M. (1995), Gaussian Semiparametric Estimation of Long Range Dependence, Annals of Statistics, 23, 1630–1661.

Tsay, R.S. (1989), Testing and Modeling Threshold Autoregressive Processes, Journal of the American Statistical Association, 84, 231–240.

Vogelsang, T.J. (1998b), Testing for the Shift in Mean Without Having to Estimate Serial-Correlation Parameters, Journal of Business & Economic Statistics, 16, 73–80.