

Working Paper 9014

COINTEGRATION AND TRANSFORMED SERIES

by Jeffrey J. Hallman

Jeffrey J. Hallman is an economist at the Federal Reserve Bank of Cleveland.

Working papers of the Federal Reserve Bank of Cleveland are preliminary materials circulated to stimulate discussion and critical comment. The views stated herein are those of the author and not necessarily those of the Federal Reserve Bank of Cleveland or of the Board of Governors of the Federal Reserve System.

December 1990

I. Introduction

A large and growing literature is concerned with the theory, estimation, and applications of cointegrating vectors and associated error correction models. A cointegrated system is a set of time series that individually follow difference-stationary linear processes, but one or more linear combinations of the series do not require differencing to appear stationary. The stationary linear combinations indicate stable long-run relationships. Engle and Granger (1987) demonstrate the correspondence between cointegrated time series and error correction models: generating processes for cointegrated systems have error correction representations, and error correction models generate cointegrated series.

Nearly all of the work in the unit root literature thus far is applicable only to series generated by a linear process. The exceptions are two papers by Granger and Hallman (1988, 1990). The first of these considers properties of nonlinearly transformed integrated series and the effect of such transformations on unit root tests. The second introduces the concept of "attractor sets," a nonlinear generalization of cointegration. Roughly speaking, if \mathbf{x}_t is an n -dimensional time series with all components having long memory (defined below), then a subset \mathbf{A} of \mathbf{R}^n is an attractor set if z_t , the Euclidean distance from \mathbf{x}_t to \mathbf{A} , is a short-memory process with bounded variance. Linear cointegration is a special case in which \mathbf{A} is a hyperplane, and the components $\{x_{it}\}$ of \mathbf{x}_t are not only long memory but difference stationary as well.

This paper can be regarded as falling between the studies described above, focusing on series that are (linearly) cointegrated only after they are individually nonlinearly transformed. Such series may be thought of as having an attractor that is the kernel of an additively separable function of

$$\mathbf{x}_t = (x_{1t}, x_{2t}, \dots); \text{ that is,}$$

$$\mathbf{A} = \{ \mathbf{x} : f(\mathbf{x}) = 0 \}$$

where

$$f(\mathbf{x}) = \phi_1(x_1) + \phi_2(x_2) + \dots + \phi_n(x_n),$$

but this is not always true. Nonlinear cointegration is more general than the notion of an attractor set and may be more useful to economists as well. The relationship between attractor sets and nonlinear cointegration is explored in section II.

If two series are cointegrated and the second series is also cointegrated with a third, then it is well known that the first and third series are also cointegrated. Granger and Hallman (1988) show that an integrated series is not cointegrated with a nonaffine transformation of itself. From this it follows that if $f(\mathbf{x}_t)$ and $g(\mathbf{y}_t)$ are cointegrated, $f(\mathbf{x}_t)$ and $h(\mathbf{y}_t)$ are not, making it important to get the transformations right. By allowing for nonparametric transformation of the series as part of the estimation procedure, the two algorithms outlined in section III increase the odds of finding long-run relationships if they do exist.

Section IV discusses testing for cointegration among transformed series and is followed by some illustrative examples in the fifth section. Section VI concludes.

II. Attractor Sets and Nonlinear Cointegration

Start with some definitions from Granger and Hallman (1990). Let the information set I_t be defined as $I_t = \{ x_{t-j}, Q_{t-j} : j = 0, 1, 2, \dots \}$, where Q_t is a vector of other explanatory variables. Then the series x_t is said to be short memory in distribution (SMD) with respect to I_t if

$$|\text{Prob}(x_{t+h} \in A \mid I_t \in B) - \text{Prob}(x_{t+h} \in A)| \rightarrow 0 \quad (1)$$

as $h \rightarrow \infty$ for all appropriate sets A and B . If equation (1) does not hold, x_t is called long memory in distribution (LMD). Denoting the conditional expectation as

$$f_{t,h} = E(x_{t+h} \mid I_t),$$

x_t is said to be short memory in mean (SMM) if $\lim_{h \rightarrow \infty} f_{t,h} = F$, where

F is a random variable with a distribution that does not depend on I_t . If $f_{t,h}$ depends on I_t for all h , then x_t is long memory in mean (LMM).

The univariate series x_t has a point attractor m if x_t is SMM with $\lim_{h \rightarrow \infty} f_{t,h} = m$ and the forecast error $e_{t,h} = (x_{t+h} - m)$ has bounded variance as $h \rightarrow \infty$. Similarly, the n -dimensional series \mathbf{x}_t is said to have an attractor $A \in R^n$ if z_t , the signed Euclidean distance from \mathbf{x}_t to the nearest point in A , is SMM and has finite variance. It is obvious that \mathbf{x}_t has the point attractor $\mathbf{m} = (m_{1t}, m_{2t}, \dots, m_{nt})$ if its components $\{x_{it}\}$ are each SMM with mean m_{it} . Two interesting cases analogous to cointegration arise when the components $\{x_{it}\}$ of

\mathbf{x}_t are LMM and either (i) \mathbf{x}_t has an attractor or (ii) a nontrivial function $f : R^n \rightarrow R^1$ exists such that $f(\mathbf{x}_t)$ is SMM. This second case will be called nonlinear cointegration, and the function f will be referred to as a cointegrating function.

If \mathbf{x}_t is LMM with an attractor, it is also nonlinearly cointegrated, with the Euclidean distance function as one cointegrating function (there may be others). However, the notion of an attractor may be overly restrictive, since it is possible for series to be nonlinearly cointegrated in an economically interesting way without having an attractor. To see in general how this can happen, suppose $f(\mathbf{x}_t)$ is a cointegrating function with mean zero and kernel \mathbf{A} ; that is,

$$\mathbf{A} = \{ \mathbf{x} : f(\mathbf{x}) = 0 \}.$$

If \mathbf{x}_t^A is the closest point in \mathbf{A} to \mathbf{x}_t , then by the Mean Value Theorem there exists a real number $\eta_t \in [0,1]$ and a point

$$\mathbf{x}_t^* = \eta_t \mathbf{x}_t + (1-\eta_t) \mathbf{x}_t^A \text{ such that}$$

$$\begin{aligned} f(\mathbf{x}_t) &= f(\mathbf{x}_t^A) + \nabla f(\mathbf{x}_t^*) \cdot (\mathbf{x}_t - \mathbf{x}_t^A) \\ &= \nabla f(\mathbf{x}_t^*) \cdot (\mathbf{x}_t - \mathbf{x}_t^A), \end{aligned}$$

since $f(\mathbf{x}_t^A) = 0$. Let θ_t be the angle between $\nabla f(\mathbf{x}_t^*)$ and $(\mathbf{x}_t - \mathbf{x}_t^A)$,

and let z_t be the signed Euclidean distance from \mathbf{x}_t to \mathbf{x}_t^A . Then

$$f(\mathbf{x}_t) = |z_t| \cos \theta_t \|\nabla f(\mathbf{x}_t^*)\|,$$

implying that

$$|z_t| = \frac{f(x_t)}{\cos\theta_t \|\nabla f(x_t^*)\|} . \quad (2)$$

If the denominator of equation (2) is bounded away from zero (that is, $\exists \delta > 0$ s.t. $|\cos\theta_t| \|\nabla f(x_t^*)\| > \delta$), then the finite variance property of $f(x_t)$ will carry through to z_t . Given the bound, $|z_t|$ may be thought of as the product of two series, at least one of which ($f[x_t]$) is SMM. Granger and Hallman (1988) show that for linear series, the product of an I(0) series with either another I(0) series or an I(1) series is SMM.¹ This suggests that in many cases the right side of equation (2) will also be SMM, so that the kernel of f will be an attractor. However, if the denominator of equation (2) tends to zero as t gets large, the finite variance property for z_t required by the definition of an attractor may not hold.

Bounding $|\cos\theta_t|$ away from zero seems reasonable enough, since it is zero only if ∇f evaluated at x_t^* is perpendicular to ∇f evaluated at the (nearby) point x_t^A . For economically interesting functions, this seems unlikely to happen. For example, if f is additively separable of the form

¹Actually, they show that the product of an I(0) and a random walk is SMM, but since an I(1) series can always be written as the sum of a pure random walk and an I(0) series, the result follows.

$$f(\mathbf{x}) = \phi_1(x_1) + \phi_2(x_2) + \dots + \phi_{x_n}(x_n) ,$$

then the gradient of f at x_t^* can become perpendicular to the gradient at x_t^A only if some of the slopes of the $\{\phi_i\}$ change sign.

Requiring monotonicity of the $\{\phi_i\}$ is enough to prevent this.

However, going further and also bounding $\|\nabla f(x_t^*)\|$ away from zero is

quite restrictive. For example, the log of the U.S. M2 money supply follows an integrated process with positive drift and is cointegrated with the log of nominal GNP. The cointegrating vector is $(1,-1)$, so that the log of M2 velocity is stationary around its mean of 0.50077 ($= \log[1.65]$). But while there is an attractor for the logs of money and income, there is no attractor for their levels. Define f by

$$f(Y_t, M_t) = \log(Y_t) - \log(M_t) - .50077.$$

The candidate attractor set is the kernel of f in Y - M space:

$$\mathbf{A} = \{ (M, Y) : Y - 1.65M = 0 \}.$$

The linear trend in $\log(M_t)$ translates into an exponential trend in the levels of M_t , Y_t , so that the gradient

$$\nabla f(Y_t, M_t) = \left[\frac{1}{Y_t}, -\frac{1}{M_t} \right]$$

is asymptotically driven to $(0,0)$. If log velocity has a constant variance, the variance of the Euclidean distance from (M_t, Y_t) to the line $Y = 1.65M$ grows like e^{2t} . \mathbf{A} is not an attractor for M_t and Y_t ,

despite the fact that they are nonlinearly cointegrated.

The point of this example is that the existence of an attractor between two or more series is not robust even to invertible transformations of the individual series. This is not true for nonlinear cointegration. If x_t and y_t are nonlinearly cointegrated, then so too are $g(x_t)$ and $h(y_t)$, if g and h can be inverted.

III. Estimation

Cointegrating transformations are not generally unique. Granger and Hallman (1990) show that if x_t , y_t are cointegrated, $g(x_t)$ and $g(y_t)$ are also cointegrated if either (i) g is homogenous or (ii) the series are scaled so that the cointegrating vector is $(1,-1)$. Absent further structure, estimating a pair of cointegrating transformations for x_t , y_t is not a well-defined optimization problem.

An optimization problem that can be solved nonparametrically is finding the transformations $\phi(\cdot)$, $\theta(\cdot)$ that maximize the sample correlation of $\phi(x_t)$ and $\theta(y_t)$. Since the asymptotic correlation between cointegrated series is one, one can hope that the correlation-maximizing transformations will also cointegrate. If the "equilibrium error" $\theta(y_t) - \phi(x_t)$ is thought to be stationary as well as SMM, the maximization can be carried out subject to the restriction that the variance of the estimated residual is constant. There is no guarantee that either of these approaches

will always discover a pair of cointegrating transformations if they exist, but applying the methods at least provides a start.

Alternating Conditional Expectation (ACE) is an algorithm proposed by Breiman and Friedman (1985) to find transformations $(\theta, \phi_1, \phi_2, \dots, \phi_n)$ for a set of variables $(y, x_1, x_2, \dots, x_n)$ that

maximize the correlation between $\theta(y)$ and $\sum_{i=1}^n \phi_i(x_i)$. This is

equivalent to maximizing the R^2 from a regression of $\theta(y)$ on $\phi_1(x_1), \dots, \phi_n(x_n)$, or minimizing

$$e^2(\theta, \phi_1, \phi_2, \dots, \phi_n) = \frac{E \left[\theta(y) - \sum_{i=1}^n \phi_i(x_i) \right]^2}{E \theta^2(y)} . \quad (3)$$

The steps in the ACE algorithm are as follows:

(i) Set $\theta(y) \leftarrow \frac{y}{\|y\|}$ and $\phi_1(x_1), \dots, \phi_n(x_n) \leftarrow 0$;

(ii) Iterate until $e^2(\theta, \phi_1, \phi_2, \dots, \phi_n)$ fails to decrease:

(a) Iterate until $e^2(\theta, \phi_1, \phi_2, \dots, \phi_n)$ fails to decrease:

$$\text{For } k = 1 \text{ to } n : \text{ Set } \phi_k \leftarrow E \left[(\theta(y) - \sum_{i \neq k} \phi_i(x_i)) \mid x_k \right];$$

End inner iteration loop;

$$(b) \text{ Set } \theta(y) \leftarrow \frac{E \left(\sum_{i=1}^n \phi_i(x_i) \mid y \right)}{\left\| \left(\sum_{i=1}^n \phi_i(x_i) \mid y \right) \right\|};$$

End outer iteration loop.

Upon completion of the algorithm, the transformations $\theta, \phi_1, \phi_2, \dots, \phi_n$ minimize equation (3).

Tibshirani's (1988) additivity and variance stabilization (AVAS) algorithm is a modification of ACE that chooses $\theta(y)$ so as to achieve a stable variance for the residual $e_t = \theta(y_t) - \sum_{i=1}^n \phi_i(x_{it})$.

At each iteration the variance function

$$v(u) = \text{var} \left(\theta(y) \mid \sum_{i=1}^n \phi_i(x_i) = u \right)$$

is used to compute the variance-stabilizing transformation

$$g(t) = \int_{-\infty}^t \frac{1}{\sqrt{v(u)}} du .$$

$\theta(y)$ for the current iteration is then computed as $\theta(y) + g[\theta(y)]$ from the previous iteration, standardized to mean zero and variance one. For the examples in section V with trending economic times series, AVAS yields more sensible transformations than does ACE.

Having estimated the transformations $\{\phi_1, \phi_2, \dots, \phi_n\}$, it may be

desirable to obtain fitted values for y , rather than its transformation. This can be done either by finding

$$E \left(y \mid \sum_{i=1}^n \phi_i(x_i) \right), \quad (4)$$

or by simply inverting $\theta(y)$ if it is monotone.

Of course, the conditional expectations appearing in equations (3) and (4) are not usually known and have to be estimated. Breiman and Friedman suggest using data smooths in their place. Any one of several scatterplot smoothers can be used, including splines, nearest neighbor, and regression smooths. (See Silverman [1985] and his discussants for a survey on the use of splines for scatterplot smoothing, and Cleveland [1979] for his lowess procedure.) In AVAS, the variance function $v(u)$ is obtained by smoothing the logs of the squared residuals $\{r_t\}$ against the fitted

values $\left(\sum_{i=1}^n \phi_i(x_{it}) \right)$ and exponentiating. See Tibshirani (1988) for details.

In the ACE routines used for this paper, both fixed and variable window regression smooths are employed. A fixed window smooth of size k computes $E(y \mid x)$ as follows:

- (i) Sort the observations by x value.
- (ii) Define the window W_n as the set of all observations $\{x_j, y_j\}$ such that $|j-n| \leq k$, so that the minimum window size is $k + 1$.

(iii) $E(y_n | x)$ is the fitted value of y_n from a linear regression of y on a constant and x , using only the observations in the window W_n .

(iv) Repeat steps (ii) and (iii) for each individual observation y_n in y .

(v) For technical reasons detailed in Breiman and Friedman, the data smooths must always have a zero mean, so the sample mean of the computed $E(y | x)$ is subtracted away before the observations are sorted back into their original order.

If $k + 1 = T$, the sample size, the smooth is just the linear regression $y = \alpha + \beta x$ and the returned values are the demeaned fitted values $\{\hat{y} - \bar{y}\}$. At the other extreme, $k = 0$ will return y minus its mean, a perfect fit. In between, larger values of k trade more smoothness for less ability to trace discontinuities and sharp changes in the slope of $y | x$. The effect of reducing the window size is similar to what happens in a linear regression as more variables are allowed to enter.

The smoother used in Breiman and Friedman's ACE implementation is the variable window "supersmoother" of Friedman and Stuetzle (1982). It differs from the fixed window smoother by making several passes with different window sizes and then choosing one of these for each observation based on a local cross-validation measure. When there is substantial autocorrelation among the prediction errors of the sorted data, the supersmoother tends to

choose window sizes that are too small, so that a plot of the smoothed data may still appear somewhat jagged. Experience so far indicates that this effect is mitigated by a high signal-to-noise ratio, as when the series are highly correlated after very smooth transformations. Nonlinear cointegration is expected to be such a case, and the transformations of economic series found by the supersmoother in section V appear acceptably smooth. Nonetheless, both fixed and variable window smooths are employed in the ACE regressions given in sections IV and V to explore the effects of changing window sizes. Only a variable window smooth was available in the AVAS implementation.

Breiman and Friedman prove that for a stationary, ergodic process, ACE converges to the optimal transformations if the smooths used are (i) uniformly bounded as $T \rightarrow \infty$, (ii) linear, and (iii) mean-squared consistent. Marhoul and Owen (1984) show regression smooths to be mean-squared consistent under conditions that are not satisfied by integrated time series. No one has yet derived conditions under which ACE and AVAS are asymptotically guaranteed to find cointegrating transformations if they exist. The approach taken here is to use the algorithms to find candidate transformations and then test for cointegration as outlined in the next section.

IV. Testing

If x_t and y_t are LMM series with cointegrating transformations $f(x_t)$ and $g(y_t)$, then $z_t = g(y_t) - f(x_t)$ is SMM. Evidence that $f(x_t)$

and $g(y_t)$ are LMM while z_t is not is one way to test for nonlinear cointegration. Granger and Hallman (1988) propose using both the Augmented Dickey-Fuller (ADF) test and a rank version of the ADF called RADF to test the LMM property in a univariate series.

The ADF statistic for testing the unit root hypothesis is the t-statistic for α in the regression

$$\Delta z_t = \alpha z_{t-1} + \sum_{i=1}^p \Delta z_{t-i} + \epsilon_t . \quad (5)$$

If $p = 0$, no lags of Δz_t appear in equation (5). The resulting statistic is then referred to as the Dickey-Fuller (DF) statistic. A one-sided test of the hypothesis of a unit root in z_t is rejected if the statistic falls below a critical value. If z_t has a nonzero mean, either it is subtracted off before performing the test, or a constant is included in the regression. If z_t is a residual from ACE or from a regression including a constant term, it has mean zero by construction.

To construct the RADF statistic, let $r_t = \text{rank}(z_t)$; that is, r_s is one if z_s is the largest of the $\{z_t\}$, or two if z_s is the second largest of the $\{z_t\}$, and so on. Replace the $\{z_t\}$ in equation (5) by their ranks and then compute the RADF as the t-statistic for α just as before. By construction, the RADF and RDF (rank counterpart of the DF) statistics are invariant to monotone transformations of z_t . Granger and Hallman (1988) show that this is a considerable advantage in that the usual DF and ADF tests perform badly when z_t is a nonlinear transformation of an

integrated series with a linear generating process.

Use of the ADF as a test for linear cointegration was first suggested by Engle and Granger (1987), and its distribution has been studied by Phillips (1987), Engle and Yoo (1986), and Yoo (1987). Engle and Yoo provide tables of critical values for the test. These depend on both the number of observations in the sample and the number of parameters estimated in the cointegrating regression.² This presents a problem because ACE and AVAS do not estimate parameters. However, shrinking window sizes in ACE is much like allowing for more parameters in a regression. What is needed is an indication of the effects of changing window sizes on the distribution of ADF and RADF statistics constructed from ACE and AVAS residuals.

A simple Monte Carlo experiment was conducted using 300 repetitions of the following:

(i) Generate \mathbf{u} and \mathbf{e} as vectors of 100 i.i.d. $N(0,1)$ random variables,

(ii) Form summations $x_t = \sum_{j=1}^t u_j$, $\epsilon_t = \sum_{j=1}^t e_j$, and

(iii) Form y_t by

$$(a) \quad y_t = \epsilon_t,$$

$$(b) \quad y_t = \frac{1}{3}x_t + \epsilon_t, \text{ and}$$

² See table 3 (panel b) for percentiles of the RADF as a test for linear cointegration.

$$(c) \ y_t = 3x_t + \epsilon_t.$$

If the series were stationary, (a), (b), and (c) would correspond to R^2 values of 0, 0.1, and 0.9, respectively. In fact, y_t and x_t are correlated random walks that are not cointegrated.

- (iv) The ACE algorithm was applied to the series with various fixed window sizes, as were both ACE and AVAS algorithms using the variable window size smoother. All transformations were restricted to be monotone. After forming the residual series z_t for each case, the ADF and RADF statistics were computed with four lags of Δz_t appearing on the right side of equation (5).

Results of the simulations are summarized in tables 1 and 2, which show the percentiles of the ADF and RADF distributions generated by the experiment. As in Engle and Yoo, the minus signs are omitted for simplicity. For comparison, table 3 shows the distributions of the ADF and RADF statistics using residuals from ordinary least squares (OLS) regressions of a pure random walk on a constant and one, two, three, or four independent random walks. Again, four lags of Δz_t were used in the ADF regression. This table is based on 5,000 replications of each test.

Several patterns are evident in the tables. From the fixed window entries, it is apparent that both the ADF and RADF distributions shift to the right with increasing window size and

increasing β . The RADF results for the variable window ACE and AVAS appear stable across the three β values, as do the ADF results for AVAS. The ADF distribution for the variable window ACE shifts right as β increases.

Table 1 ADF Percentiles

(a) $\beta = 0$								
Method	Window	5%	10%	20%	50%	80%	90%	95%
ACE	9	1.97	2.17	2.40	3.17	3.66	4.08	4.35
ACE	14	1.67	1.91	2.26	2.96	3.58	3.98	4.32
ACE	19	1.53	1.74	2.11	2.81	3.49	3.79	4.20
ACE	24	1.33	1.63	1.99	2.66	3.33	3.77	4.15
ACE	29	1.18	1.49	1.90	2.55	3.37	3.72	4.11
ACE	34	0.91	1.37	1.77	2.53	3.30	3.67	3.88
ACE	39	0.89	1.35	1.74	2.48	3.18	3.53	3.88
ACE	44	0.92	1.27	1.72	2.42	3.17	3.52	3.77
ACE	49	0.60	1.19	1.63	2.36	3.07	3.49	3.72
ACE	Variable	1.01	1.36	1.81	2.55	3.43	3.95	4.45
AVAS	Variable	0.84	1.35	1.76	2.45	3.16	3.47	3.77
(b) $\beta = 0.333$								
ACE	9	1.82	2.11	2.38	3.04	3.73	4.09	4.24
ACE	14	1.59	1.97	2.20	2.85	3.50	3.83	4.16
ACE	19	1.42	1.78	2.11	2.73	3.38	3.75	4.06
ACE	24	1.35	1.67	1.95	2.59	3.25	3.62	3.99
ACE	29	1.24	1.58	1.86	2.54	3.20	3.47	3.85
ACE	34	0.97	1.42	1.85	2.46	3.14	3.45	3.71
ACE	39	1.05	1.37	1.76	2.38	3.08	3.40	3.62
ACE	44	0.88	1.29	1.69	2.36	3.02	3.39	3.60
ACE	49	0.80	1.28	1.66	2.26	3.02	3.31	3.54
ACE	Variable	1.18	1.51	1.90	2.67	3.51	3.92	4.45
AVAS	Variable	0.91	1.31	1.75	2.46	3.11	3.47	3.82
(c) $\beta = 3$								
ACE	9	1.26	1.59	1.95	2.53	3.21	3.54	3.78
ACE	14	1.12	1.53	1.78	2.42	3.11	3.45	3.76
ACE	19	1.00	1.40	1.71	2.35	3.05	3.36	3.59
ACE	24	1.02	1.29	1.62	2.29	2.94	3.31	3.53
ACE	29	0.78	1.19	1.58	2.22	2.89	3.31	3.46
ACE	34	0.81	1.12	1.51	2.20	2.86	3.27	3.44
ACE	39	0.75	1.12	1.48	2.16	2.83	3.25	3.48
ACE	44	0.66	1.07	1.46	2.14	2.82	3.22	3.48
ACE	49	0.66	1.00	1.41	2.12	2.81	3.25	3.42
ACE	Variable	1.37	1.64	1.95	2.59	3.20	3.51	3.74
AVAS	Variable	1.24	1.56	1.89	2.51	3.24	3.50	3.74
OLS		0.51	0.92	1.28	1.98	2.62	2.98	3.23

Source: Author's calculations.

Table 2 RADF Percentiles

(a) $\beta = 0$								
Method	Window	5%	10%	20%	50%	80%	90%	95%
ACE	9	1.89	2.07	2.39	3.05	3.58	3.87	4.11
ACE	14	1.69	1.98	2.19	2.83	3.42	3.77	4.04
ACE	19	1.57	1.84	2.09	2.73	3.31	3.66	3.88
ACE	24	1.53	1.75	2.04	2.61	3.22	3.62	3.85
ACE	29	1.42	1.66	1.89	2.49	3.23	3.65	3.88
ACE	34	1.43	1.56	1.80	2.41	3.21	3.51	3.80
ACE	39	1.35	1.51	1.75	2.38	3.14	3.37	3.72
ACE	44	1.28	1.48	1.73	2.34	3.07	3.31	3.57
ACE	49	1.28	1.50	1.69	2.29	2.95	3.26	3.46
ACE	Variable	1.21	1.47	1.79	2.37	2.95	3.28	3.61
AVAS	Variable	1.18	1.42	1.70	2.30	2.91	3.22	3.50
(b) $\beta = 0.333$								
ACE	9	1.88	2.09	2.34	2.89	3.45	3.85	3.98
ACE	14	1.71	1.92	2.19	2.71	3.30	3.60	3.91
ACE	19	1.59	1.80	2.12	2.63	3.22	3.54	3.85
ACE	24	1.48	1.67	2.02	2.53	3.11	3.43	3.69
ACE	29	1.42	1.58	1.92	2.45	3.08	3.33	3.58
ACE	34	1.37	1.54	1.86	2.38	3.00	3.27	3.52
ACE	39	1.31	1.53	1.82	2.28	3.02	3.30	3.55
ACE	44	1.26	1.43	1.77	2.24	2.95	3.27	3.48
ACE	49	1.22	1.42	1.72	2.24	2.93	3.22	3.36
ACE	Variable	1.29	1.52	1.84	2.40	3.01	3.37	3.65
AVAS	Variable	1.15	1.43	1.73	2.29	2.88	3.25	3.53
(c) $\beta = 3$								
ACE	9	1.48	1.65	1.89	2.47	3.13	3.47	3.68
ACE	14	1.38	1.51	1.80	2.34	3.04	3.34	3.57
ACE	19	1.28	1.49	1.74	2.30	2.98	3.30	3.49
ACE	24	1.25	1.43	1.68	2.25	2.86	3.21	3.42
ACE	29	1.22	1.38	1.65	2.21	2.80	3.13	3.40
ACE	34	1.12	1.40	1.61	2.17	2.78	3.13	3.35
ACE	39	1.12	1.35	1.60	2.13	2.80	3.15	3.31
ACE	44	1.09	1.33	1.58	2.10	2.76	3.11	3.26
ACE	49	1.05	1.30	1.56	2.10	2.74	3.09	3.24
ACE	Variable	1.38	1.61	1.89	2.47	3.08	3.40	3.69
AVAS	Variable	1.36	1.56	1.83	2.43	3.08	3.41	3.66
OLS		0.36	0.79	1.21	1.95	2.61	3.02	3.23

Source: Author's calculations.

Table 3 ADF and RADF as a Linear Cointegration Test

(a) ADF Percentiles									
No. of Regressors	1%	5%	10%	20%	50%	80%	90%	95%	99%
1	-0.22	0.53	0.89	1.29	1.95	2.60	2.96	3.29	3.82
2	0.22	0.98	1.33	1.68	2.31	2.95	3.29	3.61	4.18
3	0.79	1.41	1.69	2.02	2.62	3.24	3.60	3.90	4.46
4	1.17	1.72	1.98	2.28	2.87	3.47	3.81	4.12	4.63
(b) RADF Percentiles									
1	0.45	0.93	1.14	1.41	1.96	2.56	2.89	3.21	3.77
2	0.82	1.27	1.50	1.75	2.32	2.92	3.25	3.57	4.12
3	1.12	1.57	1.80	2.06	2.63	3.22	3.57	3.89	4.43
4	1.42	1.83	2.05	2.33	2.88	3.46	3.79	4.09	4.60

Source: Author's calculations.

The most interesting results are those for $\beta = 3$. In this case there is considerable correlation between $\phi(x_t)$ and $\theta(y_t)$, even though they are not cointegrated. This is the most likely null hypothesis in practice. When $\beta = 3$, the higher percentiles (80, 90, and 95) of the ADF are about 0.1 greater than the corresponding RADF percentiles. The upper percentiles of the two statistics for OLS (table 3) and the variable window procedures are even closer. Looking at the fixed window results, it appears that for window sizes of one-fourth to one-half the sample size, the higher percentiles fall between those found in lines 1 and 2 of table 3, panel a. The critical values for the OLS ADF with two regressors thus provide a conservative test for the ADF and RADF when ACE with a fixed window smoother is used. For the variable window procedures, adding 0.2 (for an ADF test) or 0.1 (for an RADF

test) to these same critical values gives a test of about the right size.

V. Applications

The estimation and testing techniques of sections III and IV were applied to two bivariate data sets: (i) monthly observations of prices and dividends of the Standard & Poor's common stock composite index from January 1957 through February 1990 and (ii) quarterly U.S. nominal GNP and M2 money supply from 1959:IQ through 1989:IVQ. For each data set, the first variable was regressed on the second using OLS, ACE, and AVAS.

The present value model maintains that the price of a stock is the discounted sum of expected future dividends; that is,

$$p_t = E_t \sum_{h=0}^{\infty} \beta_t^h d_{t+h}. \quad (6)$$

If dividends, d_t , follow a difference-stationary process and the discount rate, β_t , is less than one and constant, then Campbell and Shiller (1986) argue that equation (6) implies cointegration of dividends and prices. To see why, rewrite it as

$$\begin{aligned} p_t &= \frac{1}{1-\beta} d_t + E_t \sum_{h=1}^{\infty} \beta^h (d_{t+h} - d_t) \\ &= \frac{1}{1-\beta} d_t + \sum_{h=1}^{\infty} \beta^h E_t \Delta_h d_{t+h} \end{aligned}$$

using the notation $\Delta_h d_{t+h} \equiv (d_{t+h} - d_t)$. Since $\Delta_h d_t$ follows a stationary process, goes the argument, so too does its expectation.

A discounted sum of stationary variables is also stationary, so

$\sum_{h=1}^{\infty} \beta^h E_t \Delta_h d_{t+h}$ is stationary and p_t , d_t are cointegrated.

Unfortunately, the argument that the stationarity of $\Delta_h d_t$ guarantees the stationarity of $E_t \Delta_h d_{t+h}$ is incorrect. The expectation can change each period due to influences on agents' expectations that are not stationary. The argument does hold if the optimal forecast $E_t \Delta_h d_{t+h}$ is a linear function of past values of p_t and d_t with constant coefficients. But as seen in table 4, a unit root cannot be rejected in the residual from a regression of prices on dividends. The low Durbin-Watson statistic indicates that this is a spurious regression of the kind discussed in Granger and Newbold (1974), and the values of the ADF and RADF statistics are not nearly large enough to reject the hypothesis of a unit root in the OLS residuals. Figure 1(a) is a scatterplot of stock prices and dividends with the regression line superimposed. The LMM behavior of the residual is evident.

Table 4 Stock Prices and Dividends

Procedure Series	ADF(4)	RADF(4)	DW	R ²
OLS				.850
p_t	2.46	-1.23		
d_t	5.50	.56		
\hat{e}_t	-2.18	-1.98	.038	
ACE				.984
$\theta(p_t)$	1.43	-1.23		
$\phi(d_t)$	2.97	.37		
\hat{e}_t	-5.13	-4.98	.240	
AVAS				.963
$\theta(p_t)$	-.09	-1.23		
$\phi(d_t)$	1.90	.56		
\hat{e}_t	-5.44	-5.26	.147	

Source: Author's calculations.

Figures 1(b) and 1(c) show the transformations of stock prices and dividends estimated by the variable window ACE, while figures 1(d) and 1(e) show the AVAS transformations. The dividend transformation looks similar for both procedures, but the AVAS price transformation shows some evidence of nonlinearity not present in the corresponding ACE transformation. The reason for the difference is evident in plots of the ACE and AVAS residuals, figures 1(f) and 1(g). The ACE residual variance shows a clear trend that the AVAS price transformation has eliminated.

The DW is low for both the ACE and AVAS residuals, but the ADF

and RADF statistics are well above the 95th percentiles noted in tables 1 and 2. This suggests that prices and dividends are cointegrated after an appropriate transformation has been applied to dividends. Upon closer examination of the original series, however, it becomes apparent that the nonlinearity in the dividend transformation, particularly the flat spot between $d = 3$ and $d = 7$, is almost entirely due to the behavior of the two series over the 1970s. In January 1967, prices and dividends were \$84.45 and \$2.96, respectively. Fifteen years and seven months later, dividends had risen by 155 percent (to \$7.56), while stock prices had climbed only 40 percent (to \$117.86). Since that period, both series have trended mostly upward. Because there appears to be little likelihood that dividends will ever again be in the \$3.00 to \$7.50 range, there is no way to tell the difference between nonlinear cointegration and linear cointegration with time-varying parameters for these series. Given the well-known problems resulting from inappropriate detrending of $I(1)$ time series, ACE and AVAS transformations of trending series should be interpreted with caution.

Economically, the dividend transformation is not very satisfying. The present value theory implies cointegration between prices and dividends, not transformations of prices and dividends. The cause of economic understanding would be better served through an exploration of why cointegration is not found in the data. An obvious starting point would be to allow for time variation in the discount rate. Another explanation may be that investors in the

1970s thought dividend payouts were unsustainably high, perhaps due to the inadequate adjustment of depreciation allowances for inflation or obsolescence. Some support for the latter view is found by Campbell and Shiller, who report that the dividend-price ratio Granger causes dividends.

A second example clearly shows the differences between ACE and AVAS. Engle and Granger (1987) report that GNP and M2 are cointegrated in logarithms. Running either ACE or AVAS on the logarithms of the series results in transformations (figures 2[a] through 2[d]) that appear linear. However, if ACE and AVAS are used with the levels of M2 and GNP (figures 2[e] through 2[h]), only the AVAS algorithm finds the log transformation. The ACE algorithm finds that the very strong linear relationship it starts out with improves only slightly on subsequent iterations, so it stops. There is, however, an exponential trend in the residual variance. AVAS tries to eliminate it, and the resulting variance-stabilizing transformations look very much like scaled logarithms. Table 5 shows some statistics from OLS, ACE, and AVAS. The OLS results are for the logs of M2 and GNP, but the others are not.

Table 5 GNP and M2

Procedure Series	ADF(4)	RADF(4)	DW	R ²
OLS				.99834
$\log(y_t)$	3.90	.70		
$\log(m_t)$	3.98	.00		
\hat{e}_t	-3.37	-3.44	.15	
ACE				.99965
$\theta(y_t)$.11	.70		
$\phi(m_t)$.88	.00		
\hat{e}_t	-5.00	-5.42	.97	
AVAS				.99977
$\theta(y_t)$	-.25	.70		
$\phi(m_t)$	-.12	.00		
\hat{e}_t	-5.19	-5.29	.86	

Source: Author's calculations.

VI. Conclusion

Attractor sets are the special case of nonlinear cointegration in which the cointegrating function is the Euclidean distance function. However, series can be nonlinearly cointegrated in an economically interesting way without having an attractor. It may be better to aim future research at methods of discovering interesting cointegrating functions rather than at looking for attractors.

If several series are cointegrated only after they are individually nonlinearly transformed, this can be thought of as an additively separable cointegrating function. Granger and Hallman

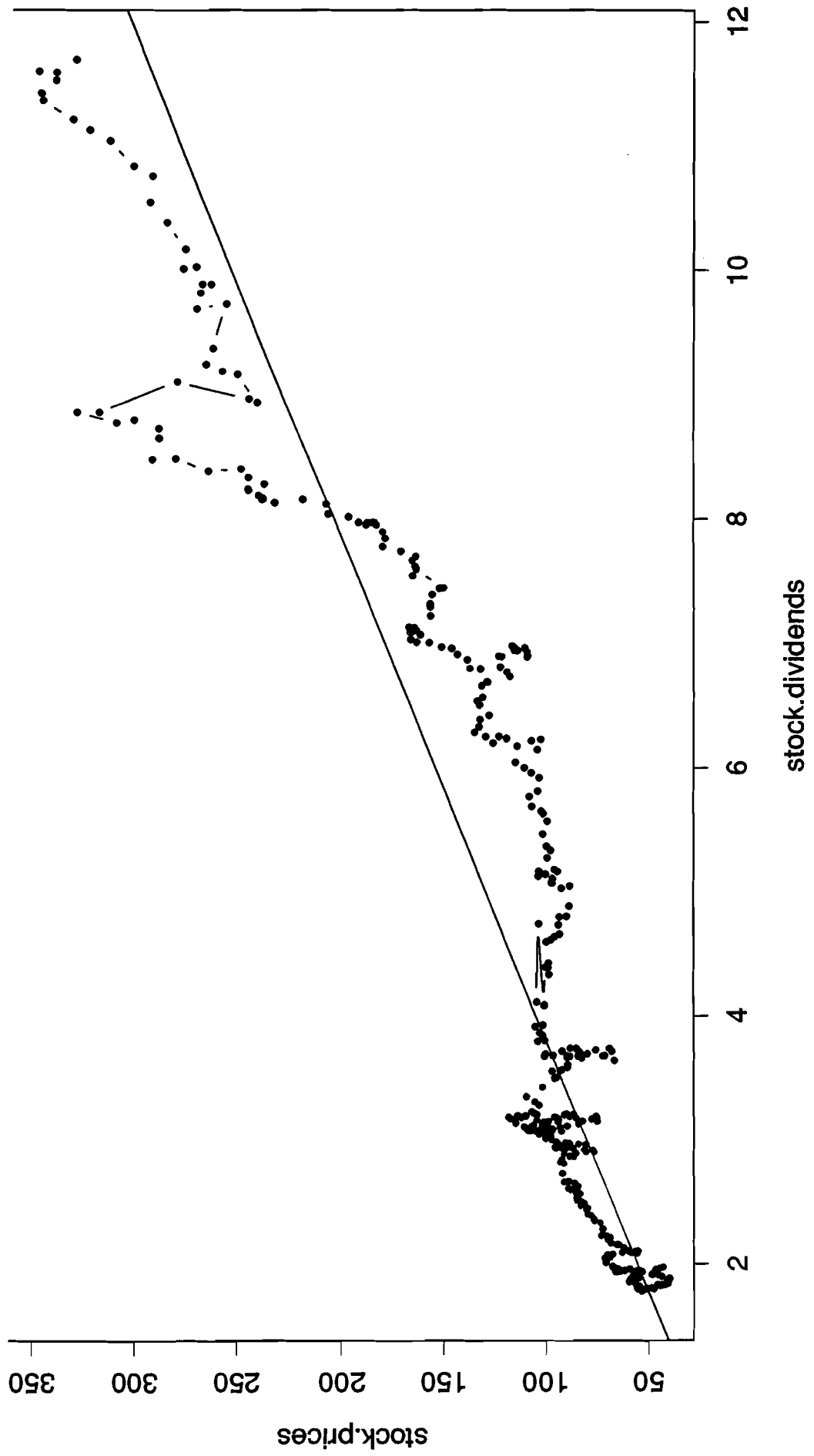
(1990) propose using ACE to estimate the transformations and the ADF to test for nonlinear cointegration. In this paper, it appears that a version of ACE modified to stabilize the residual variance may be more useful. Once the possibility of nonlinear transformations of the data is acknowledged, it would be sensible to employ a unit root test that is robust to such changes. The RADF is expressly designed for this purpose, so both it and the conventional ADF are employed.

References

1. L. Breiman and J. H. Friedman, "Estimating Optimal Transformations for Multiple Regression and Correlation," Journal of the American Statistical Association, vol. 80, pp. 580-97, 1985.
2. J. Y. Campbell and R. J. Shiller, "Cointegration and Tests of Present Value Models," Working Paper No. 1885, National Bureau of Economic Research, 1986.
3. W. S. Cleveland, "Robust Locally Weighted Regression and Smoothing Scatterplots," Journal of the American Statistical Association, vol. 74, pp. 829-36, 1979.
4. R. F. Engle and C. W. J. Granger, "Cointegration and Error Correction: Representation, Estimation and Testing," Econometrica, vol. 55, pp. 251-76, 1987.
5. R. F. Engle and B. S. Yoo, "Forecasting and Testing in Cointegrated Systems," Journal of Econometrics, vol. 35, pp. 143-59, 1987.
6. J. H. Friedman and W. Stuetzle, "Smoothing of Scatterplots," Technical Report ORION006, Stanford University, Department of Statistics, 1982.
7. C. W. J. Granger and J. J. Hallman, "The Algebra of $I(1)$," Finance and Economics Discussion Series, Board of Governors of the Federal Reserve System, 1988.
8. C. W. J. Granger and J. J. Hallman, "Long Memory Series with Attractors," Oxford Bulletin of Economics and Statistics, forthcoming, 1990.
9. C. W. J. Granger and P. Newbold, "Spurious Regressions in Econometrics," Journal of Econometrics, vol. 2, pp. 111-20, 1974.
10. J. C. Marhoul and A. B. Owen, "Consistency of Smoothing with Running Linear Fits," Technical Report LCS 008, Stanford University, Department of Statistics, 1984.
11. P. C. B. Phillips, "Time Series Regression with a Unit Root," Econometrica, vol. 55, pp. 277-301, 1987.
12. B. W. Silverman, "Some Aspects of the Spline Smoothing Approach to Non-parametric Regression Curve Fitting (with discussion)," Journal of the Royal Statistical Society, vol. 47, pp. 1-52, 1985.

13. R. Tibshirani, "Estimating Transformations for Regression via Additivity and Variance Stabilization," Journal of the American Statistical Association, vol. 83, pp. 394-405, 1988.
14. B. S. Yoo, "Co-Integrated Time Series: Structure, Forecasting and Testing," unpublished Ph.D. dissertation, University of California, San Diego, 1987.

Figure 1(a) Stock Prices and Dividends



Source: Author's calculations.

Figure 1(b): ACE Transformation of Stock Prices

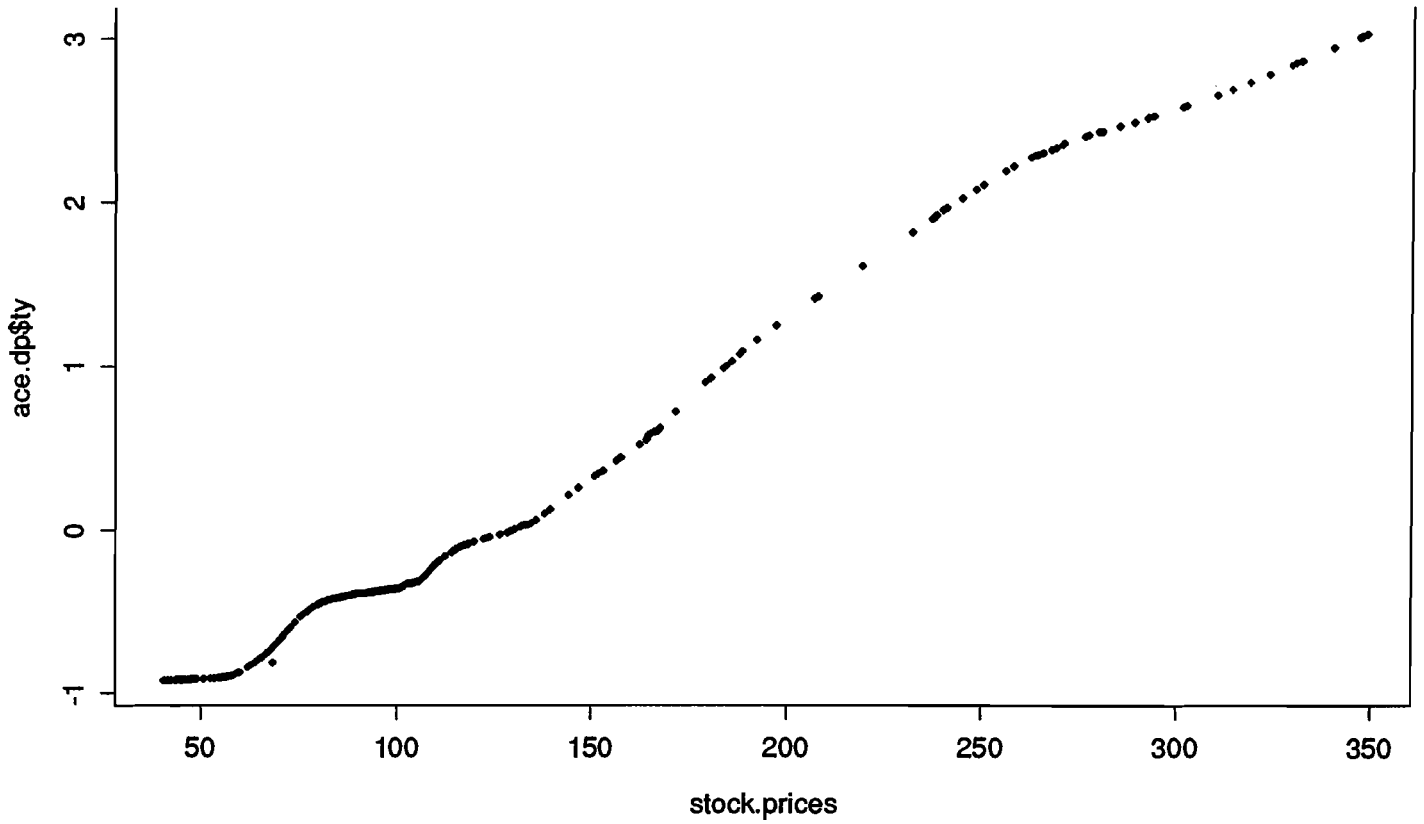
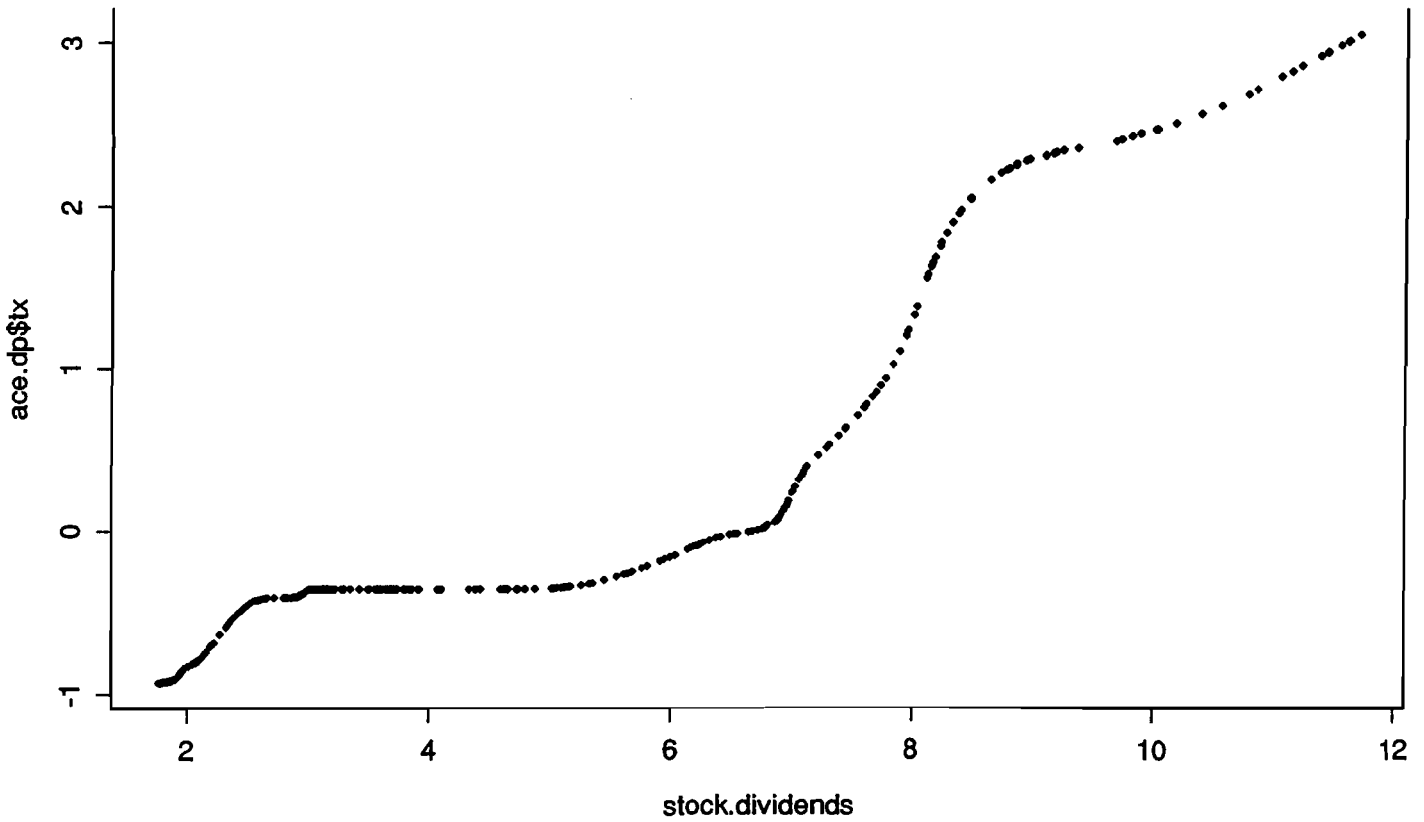


Figure 1(c): ACE Transformation of Stock Dividends



Source: Author's calculations.

Figure 1(d): AVAS Transformation of Stock Prices

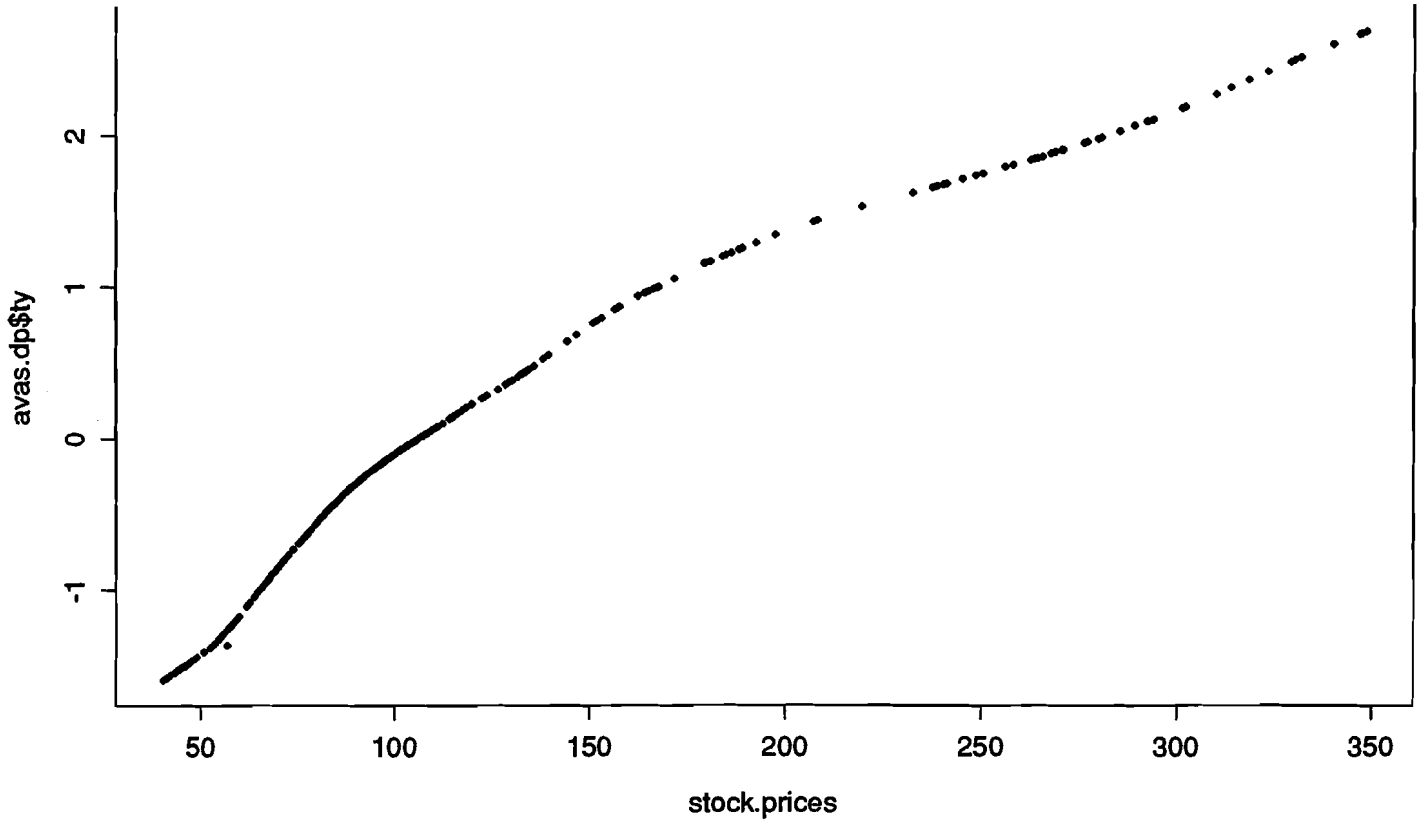
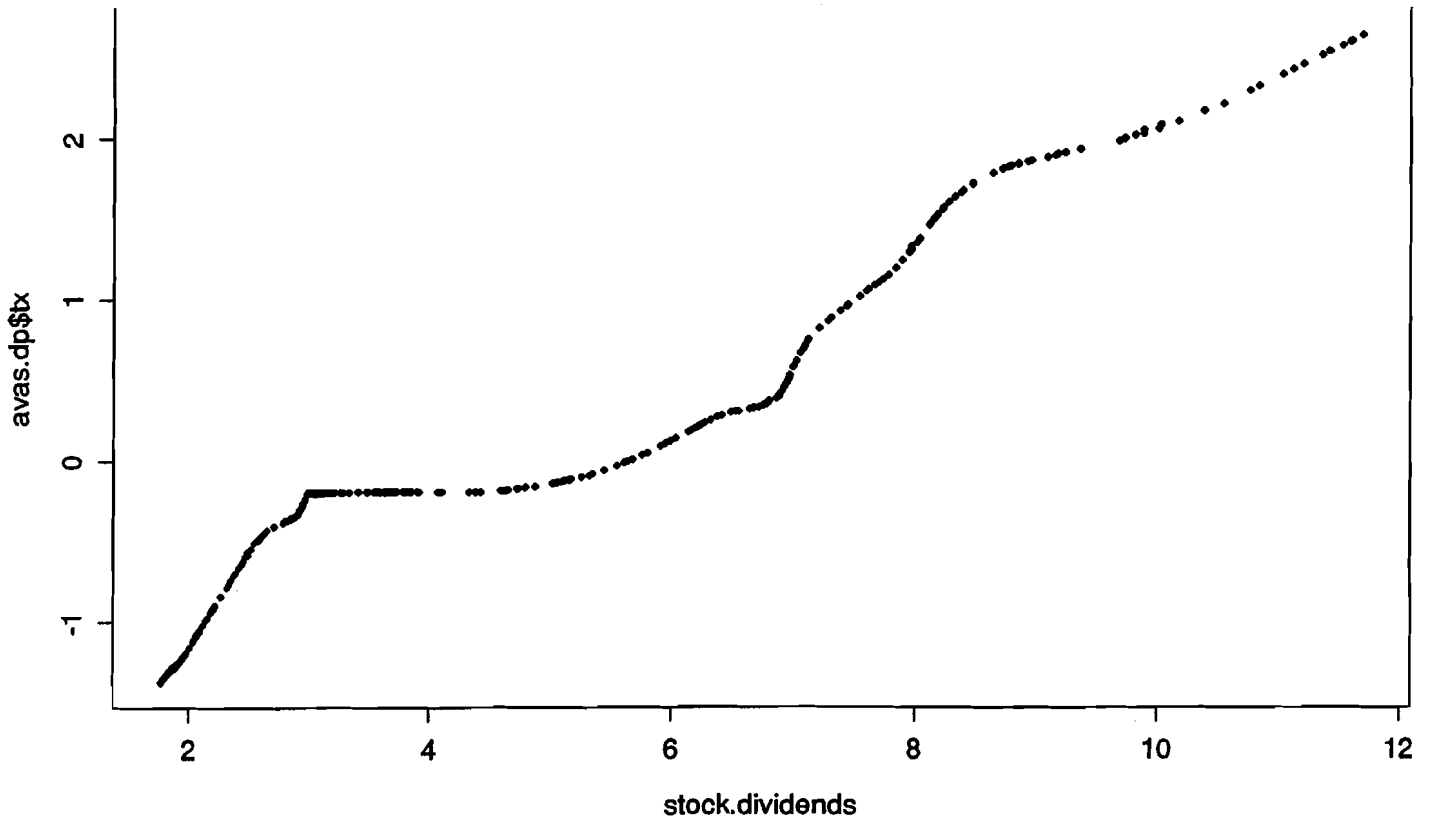


Figure 1(e): AVAS Transformation of Stock Dividends



Source: Author's calculations.

Figure 1(f): ACE Residual for Prices and Dividends

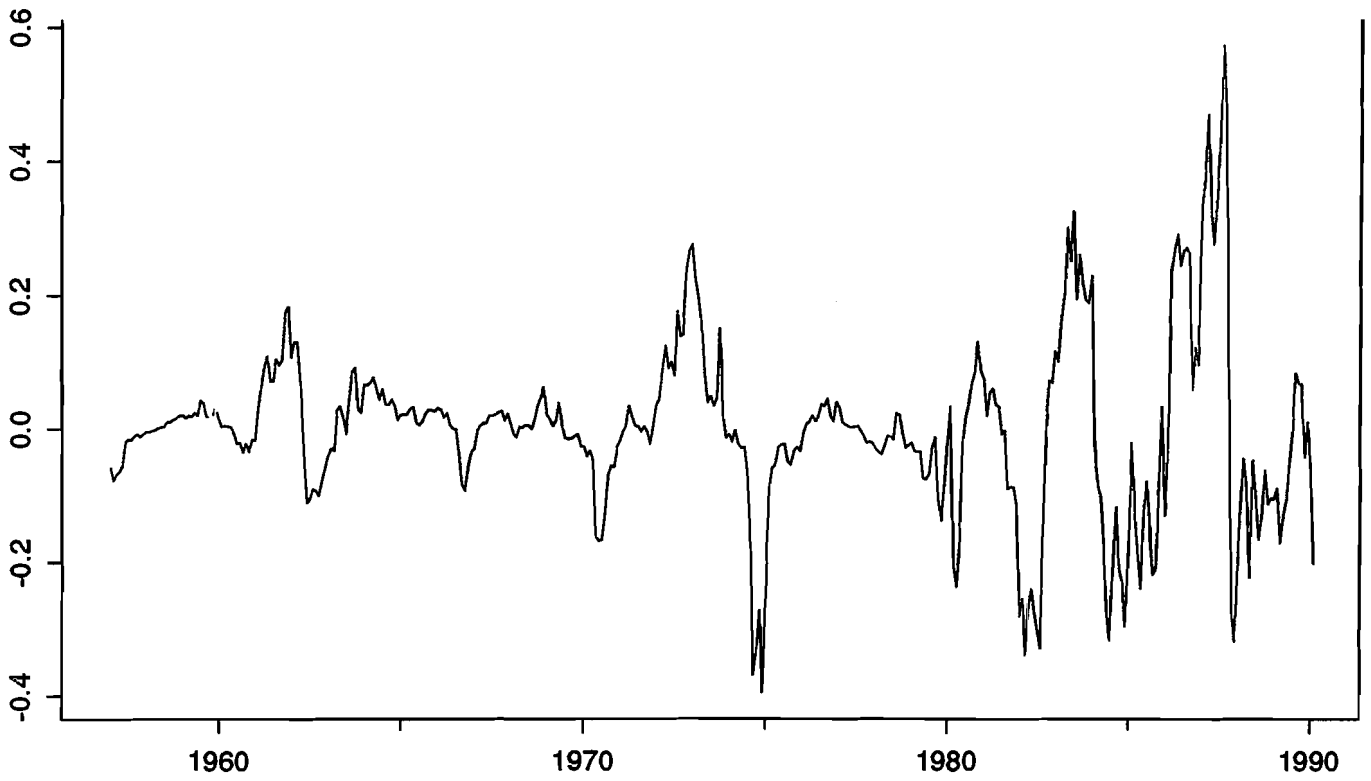
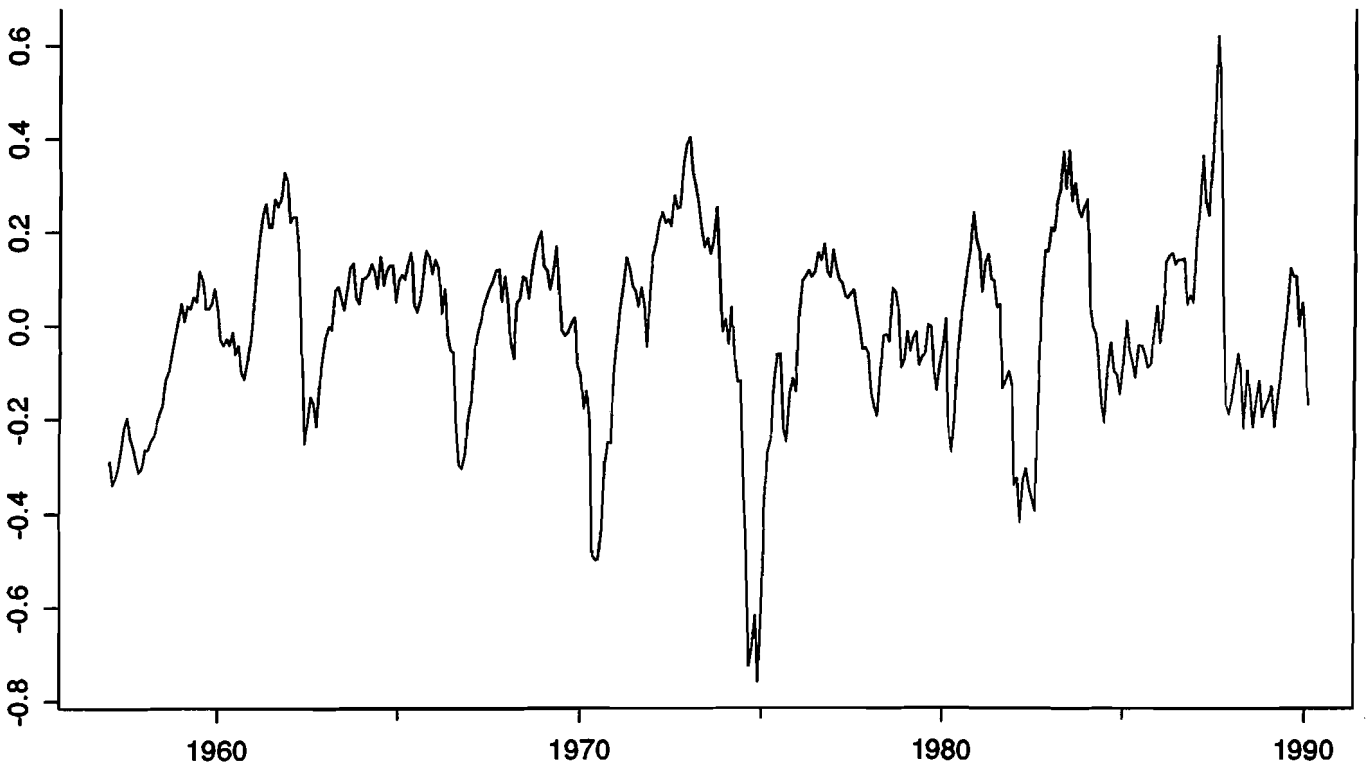


Figure 1(g): AVAS Residual for Prices and Dividends



Source: Author's calculations.

Figure 2(a): ACE Transformation of $\log(m2.q)$

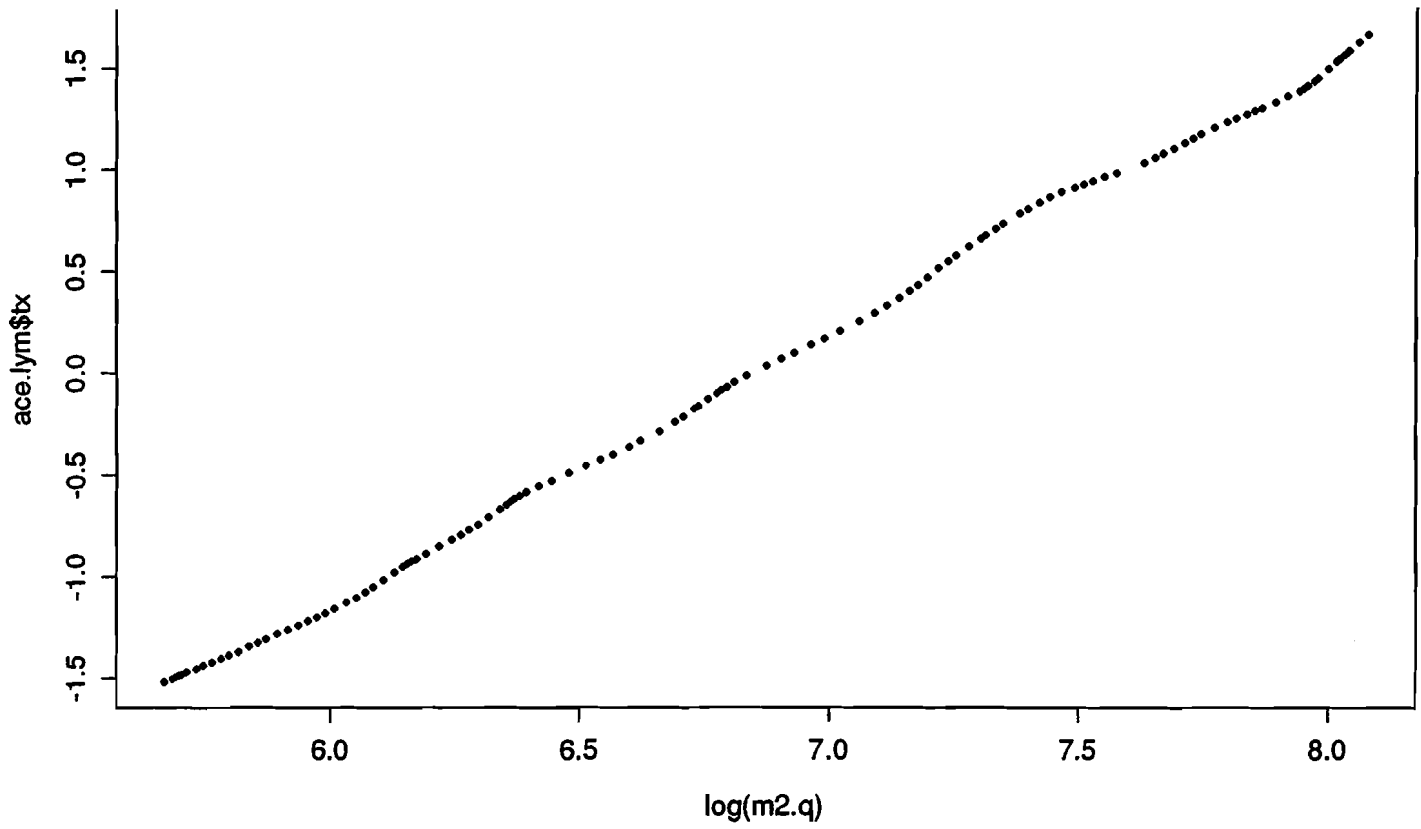
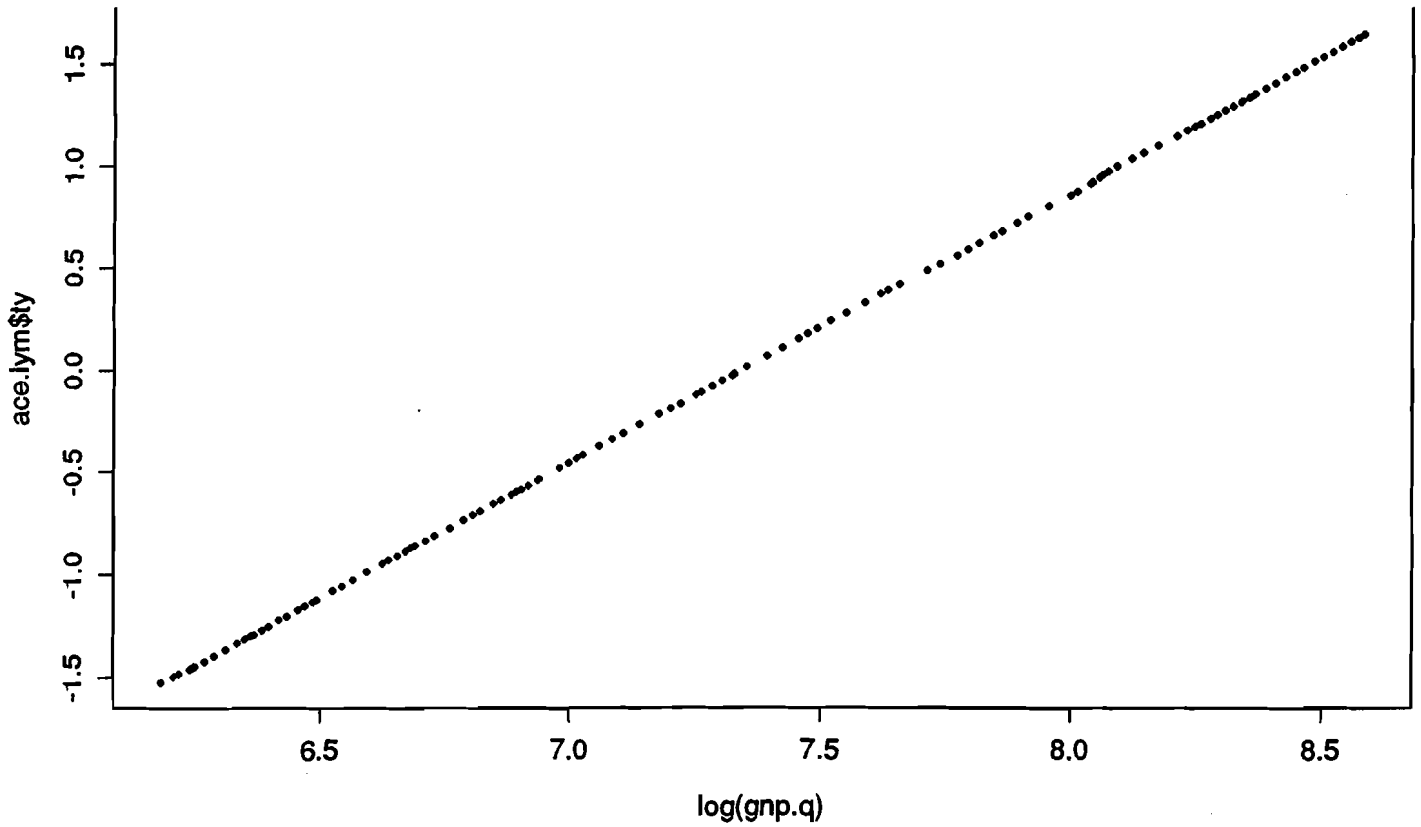


Figure 2(b): ACE Transformation of $\log(gnp.q)$



Source: Author's calculations.

Figure 2(c): AVAS Transformation of $\log(m2.q)$

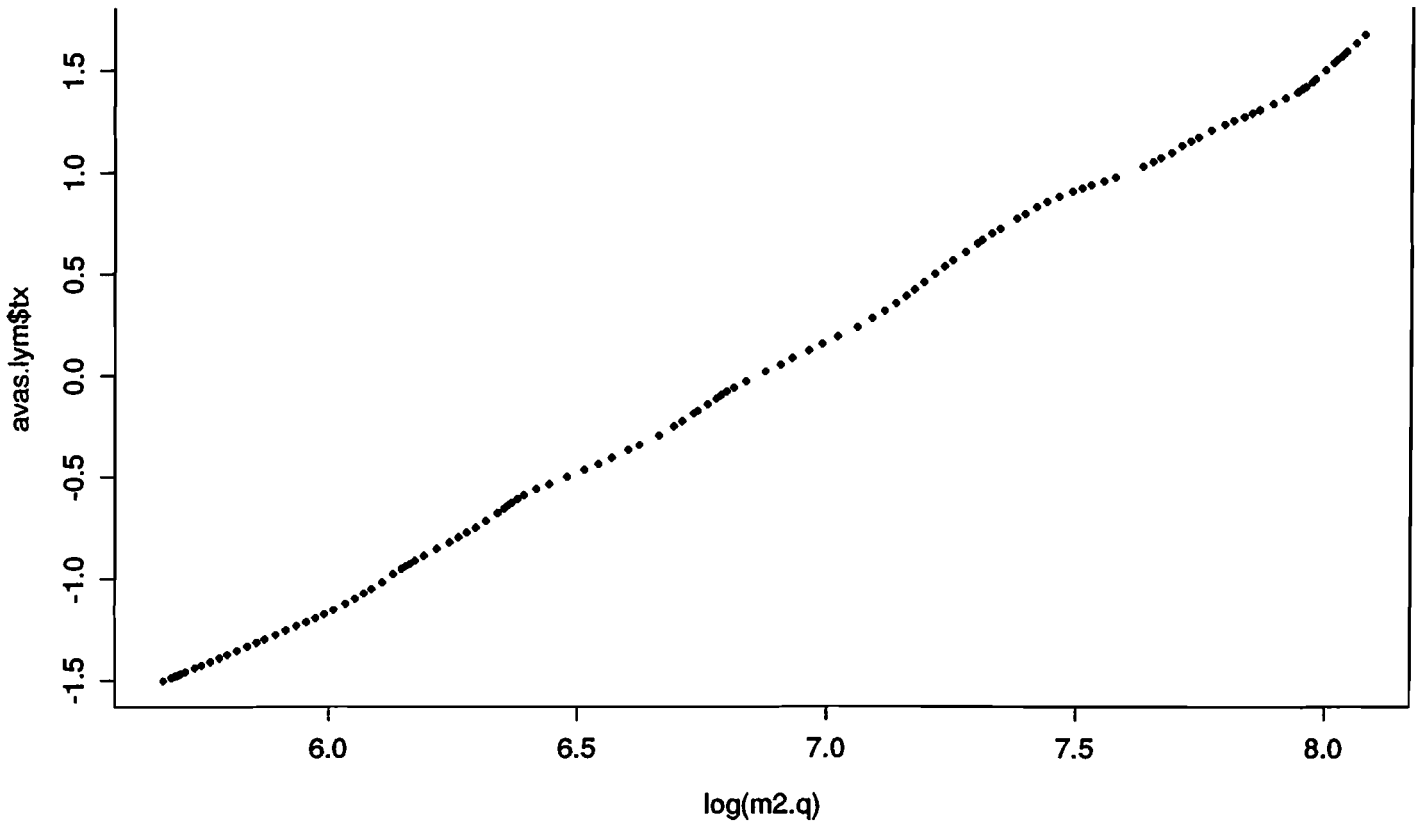
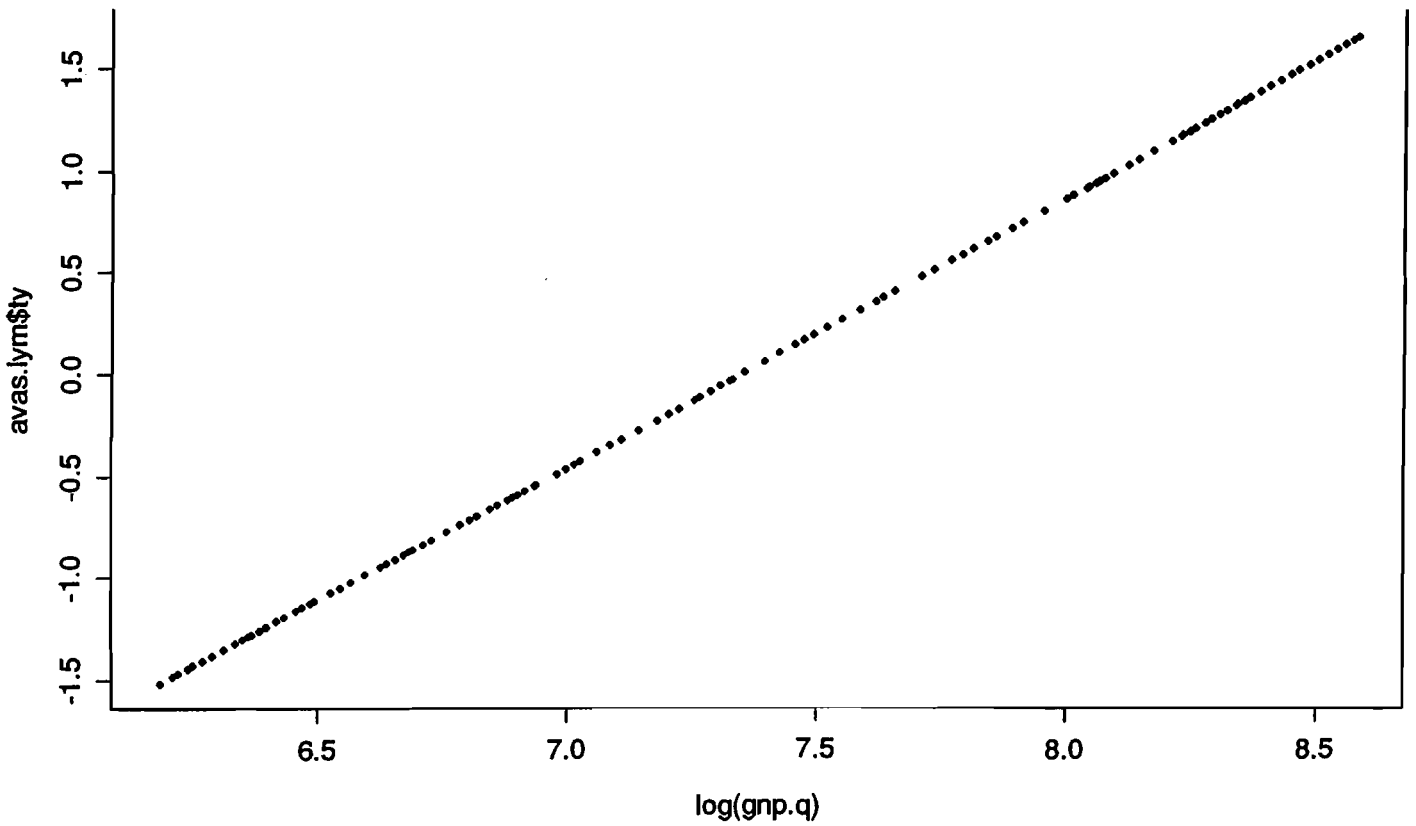


Figure 2(d): AVAS Transformation of $\log(gnp.q)$



Source: Author's calculations.

Figure 2(e): ACE Transformation of m2.q

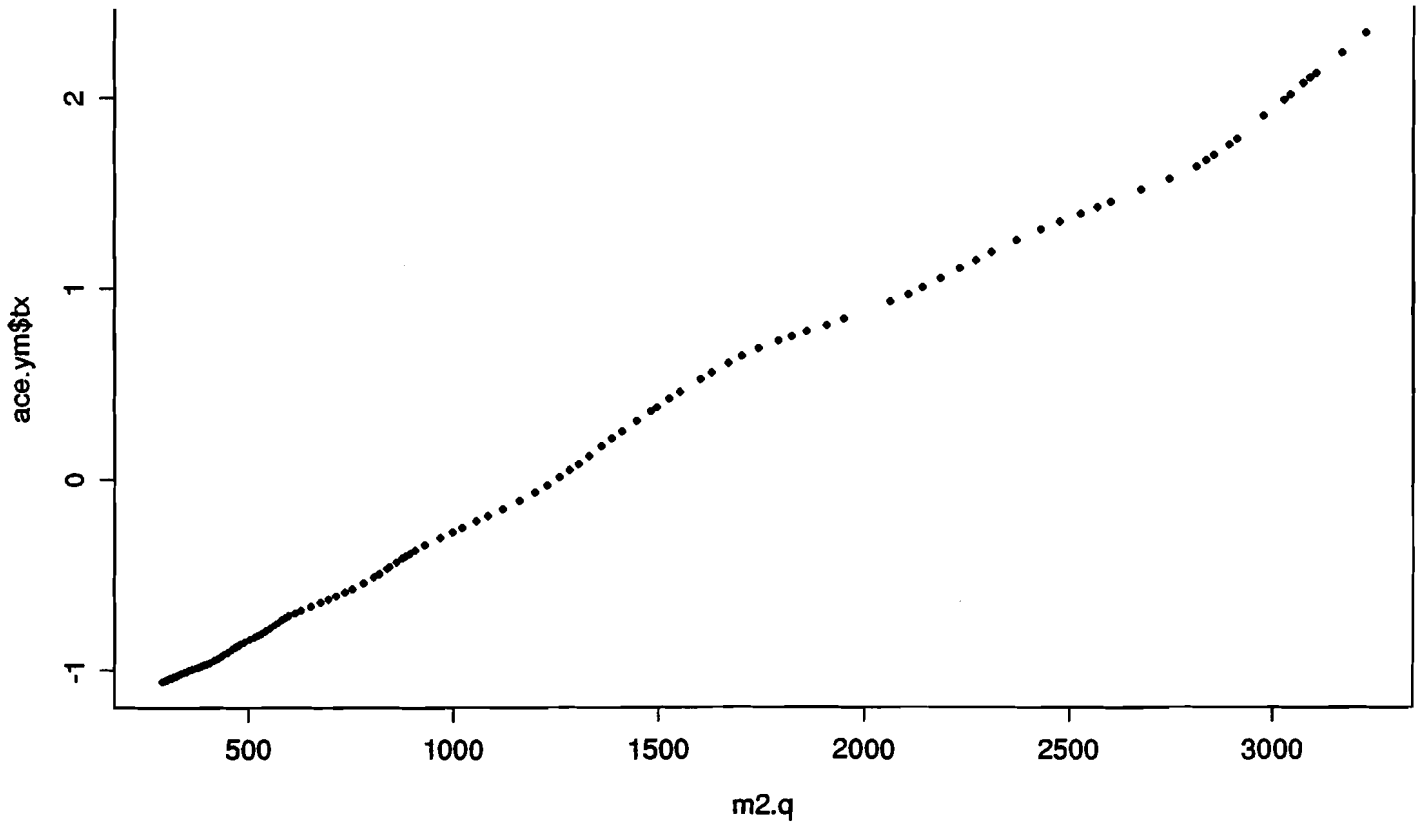
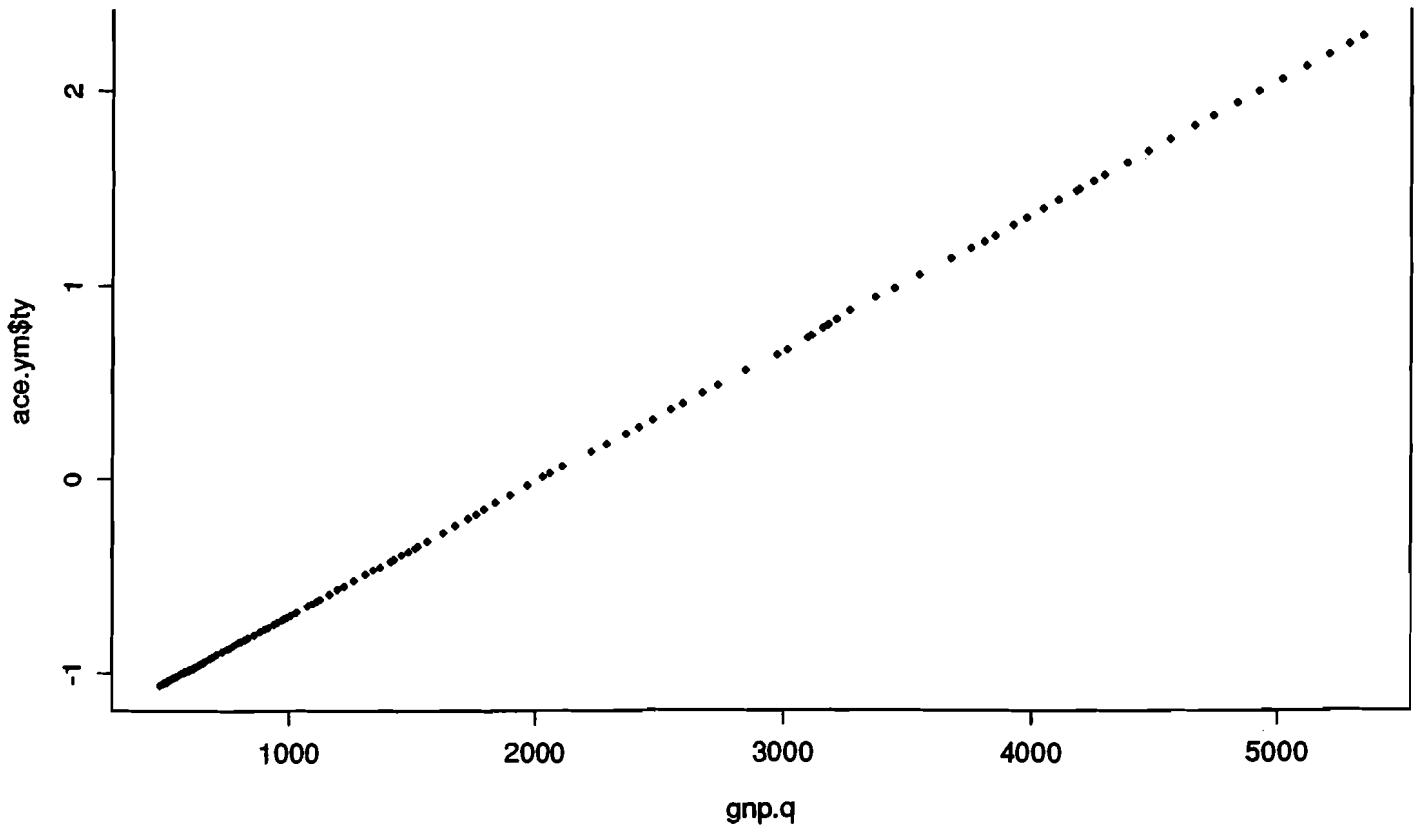


Figure 2(f): ACE Transformation of gnp.q



Source: Author's calculations.

Figure 2(g): AVAS Transformation of m2.q

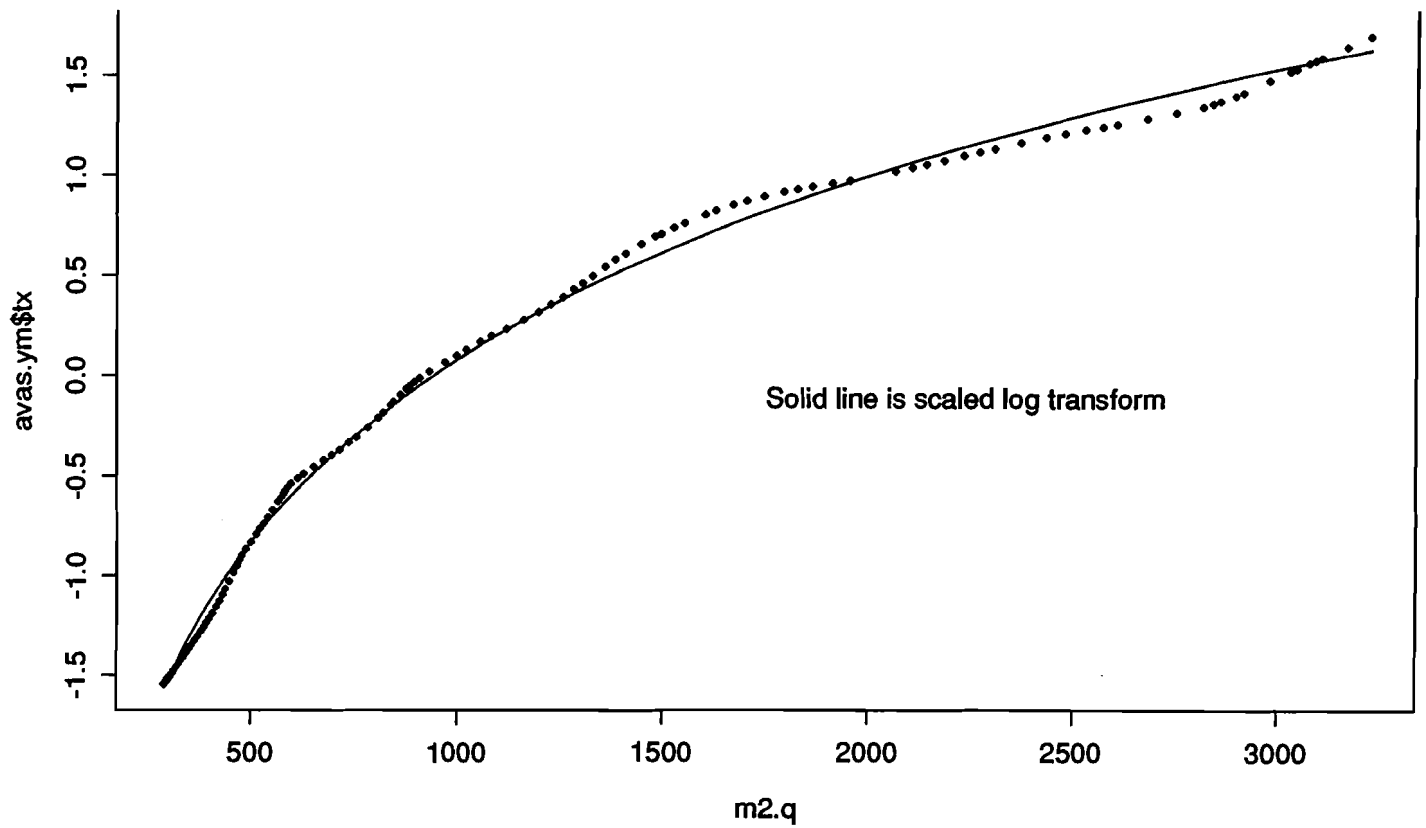
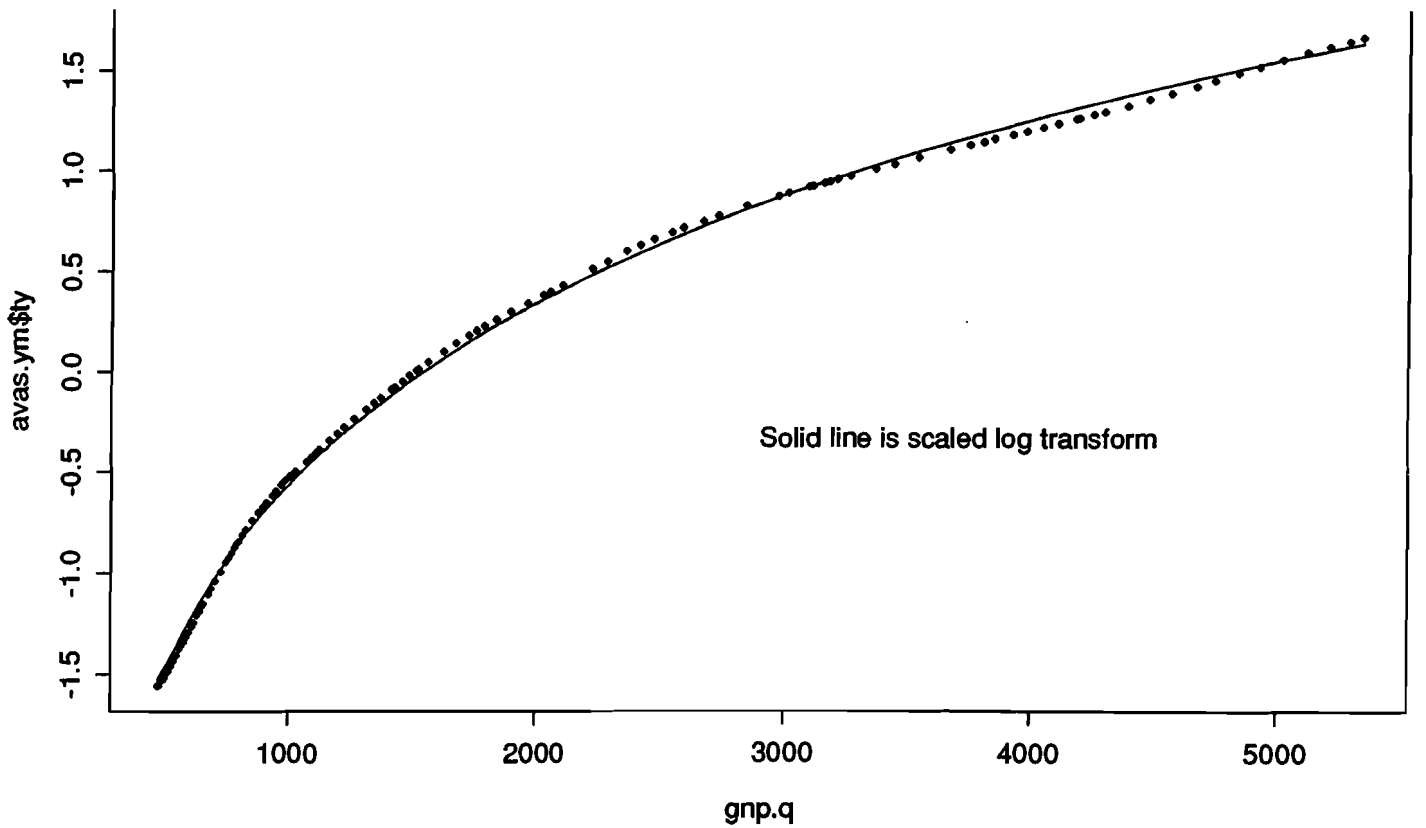


Figure 2(h): AVAS Transformation of gnp.q



Author's calculations.