CREST

École Nationale de la Statistique
et de l'Administration Économique

DÉPARTEMENT DE LA RECHERCHE

Direction des Études
et Synthèses Économiques

N°9222

RANDOM WALKS WITH DRIFT, SIMULTANEOUS EQUATION ERRORS, AND SMALL SAMPLES : SIMULATING THE BIRD'S-EYE VIEW

H. ENTORF[1]

Document de travail

Institut National de la Statistique et des Etudes Economiques

INSTITUT NATIONAL DE LA STATISTIQUE ET DES ETUDES ECONOMIQUES
Série des documents de travail
du CREST (Centre de Recherche en Economie et Statistique)
et du Département de la Recherche

n°9222

Random walks with drift, simultaneous equation errors, and small samples : simulating the bird's-eye view

H. ENTORF[1]

November 1992

Correspondence to:
Horst Entorf
CREST
15 Boulevard Gabriel Péri
F- 92244 Malakoff, France

1 Previous versions of the paper were presented at the conference "Bayesian and Classical Econometric Modelling of Time Series", Marseille, June 1992, and at the ESEM meeting in Brussels. My thanks go to Jan Kiviet for his constructive discussion of my paper and to Svend Hylleberg and Peter Phillips for helpful comments and suggestions. Moreover, I would like to thank Luc Bauwens for providing computer facilities at CORE so that the CPU-intensive work could be finished within a reasonable time. I gratefully acknowledge financial support by the Commission of the European Communities under contract B/SPES-900042.

ABSTRACT

The paper illustrates finite sample problems of regression models with I(1)-variables. Particular care is practised with respect to the role of drifts. First, their impact on the regression of two independent random walks is analysed both analytically and by simulation methods. Second, Haavelmo's famous simultaneous equation bias is considered in the presence of cointegrated variables. Simulation results reveal a decisive role of drifts, possibly being even more important than the length of the time series. Furthermore, some experiments are devoted to the comparison between the Haavelmo bias in the stationary case and the error in the case of cointegrated variables. A main purpose of the paper is to provide a graphical exposition of finite sample problems: Most of the simulation results are summarized as three-dimensional density functions.

RESUME

Dans cet article, on analyse les problèmes des petits échantillons rencontrés dans le modèle de la régression lineaire avec des séries integrées d'ordre 1. On examine particulièrement le rôle des tendances déterministes: (i) leur influence dans la régression entre deux marches aléatoires independantes est considerée, (ii) on évalue le biais des estimateurs du système d'équations simultanées (Haavelmo, 1943) en presence des variables cointegrées. Cette étude fait apparaître une sensibilité importante des résultats à la spécification des composantes déterministes. (iii) On compare le biais de simultaneité obtenu sur séries stationaires à celui observé en presence des variables cointegrées. Une des motivations de cet article est de faire une exposition graphique des problèmes présents dans des études sur petits échantillons. L'essentiel des résultats simulés est presenté sous forme de fonctions de densité tridimensionelle.

1. Introduction

Irrespective of serial correlation of error terms or existing correlation between regressors and error terms, least squares lead to consistent estimates if the regression is run using cointegrated variables. Stock (1987) shows even more: The estimate converges to its true value at a rate T^{-1} rather than the usual $T^{-1/2}$ ("superconsistency"). As a very important and often quoted consequence[1], the simultaneous equation bias arising from the endogeneity of the regressor (Haavelmo, 1943) vanishes asymptotically. However, even "superconsistency" is an asymptotic property and this paper investigates the degree of persistence of the simultaneous equation bias in small samples.[2] Applying, among other things, time-dependent histograms, this paper demonstrates the importance of drift on the speed of convergence to the true value: Considerable drifts of stochastic processes lead to a much faster adjustment. Additional experiments reveal a satisfactory performance of simple IV estimation vis-a-vis the simultaneous equation error in the presence of integrated variables.

The same importance of drifts applies when we estimate least squares regressions where both regressor as well as regressand are random walks, but where both random walks contain a drift. This is an important case in empirical macroeconometrics since many macroeconomic time series are considered to be random walks with drift (Nelson and Plosser, 1982), so that regressions between such time series are very likely. The analysis of the problem is not made explicit in Phillips' (1986) well known paper on "understanding spurious regressions", but applying his results as well as results by Sims, Stock and Watson (1990), the appendix of this paper shows the straightforward convergence of the OLS estimator to the ratio of imposed drifts. Again, simulation results show a very slow adjustment of the OLS estimator towards the asymptotical result when individual drifts are small.

This paper extends simulation studies by Banerjee et al. (1986) and Hansen and Phillips (1990) who investigate the small-sample bias in static regressions applying the illustrative data-generation process presented by Engle and Granger (1987). Banerjee et al. (1986) conclude that care should be experienced in trying to parameterize long-run relationships using static regressions. Hansen and Phillips concentrate attention on IV-techniques developed in Phillips and Hansen (1990). This paper qualifies their findings with respect to the role of drifts, the explicit consideration of Haavelmo's simultaneous equation problem, simple TSLS techniques, signal-noise ratios and the investigation of three-dimensional histograms. The latter summarize simulation results from a bird's-eye view so that the graphics can display finite-sample problems in a compact way. Moreover, comparisons between situations with and without drift as well as contrasting cointegrated simultaneous equations with non-cointegrated simultaneous relationships highlight the sensitivity of static regression with respect to "cointegration versus non-cointegration", serial correlation of data-generation processes, sample size, drift, etc.

The paper is organized as follows. Section 2 analyses the case of two random walks, investigates the role of drifts, and provides some simulated evidence on the role of sample size and drift. Moreover, the difficult identification of random walks in short time series by inspecting DW-statistics is highlighted. Section 3 investigates the persistence of the simultaneous equation bias in small samples, provides some evidence on the impact of drifts, compares I(1) results with those obtained for stationary regressions and presents simulation results when the cointegrating regression is estimated using TSLS. Some final experiments are devoted to the role of the signal-noise ratio for the presence or absence of the simultaneous equation bias. Section 4 concludes.

2. Regressing two random walks with drifts.

In the case of random walks without drift, Phillips (1986) shows that the least squares regression

(1) $$y_t = \hat{\alpha} + \hat{\beta}x_t + \hat{u}_t, t = 1,\ldots,T$$

leads to diverging OLS estimates of $\hat{\alpha}$ and to the convergence of $\hat{\beta}$ to a random variable, when y_t and x_t are generated as independent random walks *without* drift (i.e. equation (1) is a "nonsense regression" or "spurious regression" in the sense of Granger and Newbold, 1974). In the following, we will likewise consider random walks with non-zero drifts, i.e. we define

(2) $$y_t = \gamma_y + y_{t-1} + \varepsilon_t, x_t = \gamma_x + x_{t-1} + \nu_t, t = 1,2,\ldots,$$

in which ν_t is iid(0, σ_x^2) and ε_t is iid(0, σ_y^2). Under regularity assumptions made explicit in Phillips (1986) and Sims, Stock and Watson (1990), the following result can be obtained:

Theorem 1: Suppose (1) is estimated by least squares regression. Then, as T goes to infinity,

$$\hat{\beta} \Rightarrow \frac{\gamma_y}{\gamma_x},$$

$$\hat{\alpha}T^{-1/2} \Rightarrow \sigma_y \int_0^1 W(t)dt - \frac{\gamma_y}{\gamma_x}\sigma_x \int_0^1 V(t)dt \sim N\left(0, \frac{\sigma_y^2 + \sigma_x^2(\gamma_y/\gamma_x)^2}{3}\right).$$

where W(t) and V(t) are independent Wiener processes on C[0,1] and where $\Rightarrow$ denotes "convergence in probability".

Proof: Appendix

Thus, contrary to the "spurious regressions"[3] case, OLS estimation of $\hat{\beta}$ does imply convergence to a constant value (in fact, the result turns out to be the ratio of the drift of the regressand and the drift of the regressor) instead of convergence to a random variable.[4] The result of the constant, $\hat{\alpha}$, is qualitatively the same as for random walks without drifts (see Phillips, 1986): The OLS estimate diverges with a rate $T^{1/2}$. However, Banerjee and Hendry (1992) point out that many functionals to which sample moments of I(1) processes converge can be expressed in terms of simple normal densities. Thus, using $\int W(t)dt \sim N(0,1/3)$, the constant term has a simple interpretation.

The different properties of "spurious regressions" and random walks with drift become evident by comparing time-dependent histograms of the OLS slope parameter β. In a simulation experiment we consider equations (1) and (2). The error terms are generated independently and normally distributed with means equal to zero. To avoid particularities arising from different variances, we impose identical variances, i.e. $\sigma_x = \sigma_y = 1$. Starting values are chosen as $y_0 =$ 0 and $x_0 =$ 0.[5]

Phillips (1986) shows the weak convergence of the OLS estimator to a random variable when random walks have no drift. Figure 1 reflects the nature of "nonsense regressions" in terms of time dependent histograms.[6] Irrespective of the length of the time series, we observe a symmetric and widely spread distribution. Figure 2 summarizes some interesting properties of this distribution to which the nonsense OLS slope estimator converges. The average estimate is about zero and the average 95%-fractile (5%-fractile) is slightly higher than unity (smaller than minus unity); average and median t-values diverge with decreasing growth rates (all reported t-values are in absolute values), leading to the wrong conclusion of a significant relationship (see the high percentage share of t-values larger than two).[7] As likewise shown by Phillips (1986), the distribution of the DW-statistic approaches zero for large samples.

Contrary to the spurious regression case, Figure 3 displays the convergence to constants in the presence of drifts. Of course, the magnitude of imposed drifts is debatable. Hylleberg and Mizon (1989) estimate the magnitude of drifts for some macroeconomic time series (log(GNP), log(M1), log(u/(1-u)) and log(1+r), where u = unemployment rate and r = interest rate). The estimated relations between drift and standard errors range between 0.25 and 0.72. To give a general impression of bivariate regressions with drift, we impose two drifts: the drift of the dependent variable is set to 0.5, the drift of the explaining regressor is 0.25, thus leading to a long-run parameter two; standard errors are kept as $\sigma_y = 1$ and $\sigma_x = 1$..

In small samples, however, we observe results very similar to those of random walks without drift: A flat distribution of the OLS estimator pretends significance of results which are far away from their asymptotical values. As T goes to infinity, the empirical distribution reflects the weak convergency to a constant, i.e. the distribution concentrates on the asymptotically correct result. However, as Figure 3 displays, even with sample sizes of about 300 observations, we observe a skewed distribution and critical values are still far from the asymptotic estimate. t-values of the slope parameter are linearly increasing; the DW-statistic has the same behaviour as in the case without drift.

The magnitude of drifts plays an important role. The larger the drift is, the higher is the coefficient of a linear trend in a transformed formulation of equations (1) and (2) (see Appendix). The proof of Theorem 1 shows that the asymptotic behaviour of the OLS slope estimator boils down to the relative magnitudes of these trends, because the remaining terms reveal a slower rate of convergence. Thus, the larger the drifts are, the faster is the convergence of the the trend-dominated terms, and the larger is the difference between the impact of the linear trend and the impact of the remaining terms which converge at a lower rate and which become less and less important when the sample size increases.

Figure 4 highlights the particular role of drifts. The simulation experiment compares the zero drift case with three non-zero situations. When drifts are non-zero, we simulate regression equations with identical large sample OLS slope estimates. To do so, the ratio of drifts is kept constant as $\gamma_y/\gamma_x = 2$. We define the ratio drift/standard error of the explaining random walk to be identical 0.05, 0.25 and 0.75 (thus the corresponding values of γ_y are 0.1, 0.5 and 1.5, respectively), and we compare resulting OLS estimates with regressions applying random walks without drift. Given the estimates by Hylleberg and Mizon (1989), our simulation results might cover the extreme range of possible speeds of convergence.

The estimates based on $\gamma_x = \gamma_y = 0$ fluctuate around zero, but - as already displayed in Figure 1 - no convergence to a constant value occurs. For $\gamma_x = 0.05$ the estimate is substantially different from the the asymptotic estimate (which is indicated by the upper solid line in Figure 4); even with T=300 the estimate remains below 0.5. When $\gamma_x = 0.25$, the asymptotic result is approached for T larger than 170. For the largest drift under consideration, $\gamma_x = 0.75$, even for very small samples of less than 30 observations observed estimates are close to the final result. Thus, given realistic sample sizes of less than 100 observations and ignorance about the true magnitude of drifts, OLS estimation can easily lead to wrong conclusions concerning the true nature of the data generation process - even when the convergence property towards the ratio of two drifts is known.

According to the relative role of x and y in the OLS estimator $\hat{\beta} = (x'x)^{-1}x'y$, the distribution of $\hat{\beta}$ is increasing when the standard error of the dependent variable, σ_y, is growing relative to σ_x. The distribution of the OLS estimator depends on the ratio of σ_y and σ_x. Figure 5 shows this result for nonsense regressions (without drift). However, t-values and DW-statistics seem to be independent of the relative size of standard errors.[8]

The result of Theorem 1 concerning the constant term in regressions applying I(1)-variables is illustrated in Figure 6. The graphic displays the same situation as in Figure 2, i.e. $\gamma_x = 0.25$ and $\gamma_y = 0.5$. According to the divergence rate $T^{1/2}$, we observe a slowly growing variance of the estimated constant term. The two-dimensional summaries in Figure 7 reveal that the distributional shape of the constant term is still very far away from asymptotic normality even after 300 observations. Average constant terms remain above 3 and decrease very slowly.[9]

How to ensure that a nonsense regressions has not been estimated? DW is $O_p(T^{-1})$ and tends in probability to zero as T goes to infinity (see Phillips, 1986, and Banerjee and Hendry, 1992). Thus, DW should reject nonsense regressions with probability one in large samples (see also Sargan and Bhargava, 1983). However, how does DW perform in small samples? We analyse the practical relevance of DW for T=30. Figure 8 summarizes the distribution of the DW-statistic depending on the number of independent nonsense regressors. The graphic shows that for more than 5 regressors the 95%-fractile of the DW distribution is larger than two. Thus, even regressions with DW-values of about two do not necessarily ensure that we do not estimate nonsense regressions.[10]

3. The simultaneous equation bias in the presence of cointegration, small samples, and drift

Ever since Haavelmo (1943), students in the field of econometrics were taught to be aware of a potential simultaneous equation bias which arises when regressions contain endogenous variables on both the left hand side and the right hand side of the equation. However, as follows from Stock's theorem on superconsistency, in the presence of cointegrated variables the regression goes to the true value asymptotically so that the fit is perfect no matter how strong is the simultaneous equation bias. Thus, 50 years after Haavelmo's seminal work, should students of econometrics dispense with studying the simultaneous equation bias?

Motivated by Haavelmo (1943), we consider the following model with endogenous variables C and Y and an exogenous variable X which might be interpreted as consumption C, income Y, and any exogenous variable X of the national account identity:

(i) $$X_t = \gamma + \rho X_{t-1} + \nu_t$$

(ii) $$C_t = \beta Y_t + \varepsilon_t$$

(iii) $$Y_t = C_t + X_t, t = 1,2,\ldots,T.$$

After substituting (iii) into (ii), the data-generation process (i)-(iii) calculates C_t according to

(ii') $$C_t = \frac{1}{1-\beta}(\beta X_t + \varepsilon_t).$$

The innovations are chosen as independently distributed with $\nu_t \sim N(0,1), \varepsilon_t \sim N(0,1)$. The starting value is chosen randomly according to the distribution of ν_t. For most of the simulation experiments the parameter β is set equal to 0.7. The length of the time series, T, the degree of serial correlation ρ, and the drift γ are varied depending on the particular simulation experiment. Most of the experiments impose the assumption $\rho = 1$ so that X_t, C_t and Y_t are I(1) and equation (ii) is a cointegrating relationship. The following results summarize results of the OLS regression of C_t on Y_t which is estimated for each replication of the simulation model, i.e. we simulate the estimation of a "consumption function" (ii) disregarding any problem arising from the correlation between the regressor and the error term.

First we consider the T-dependent distribution of the cointegrating OLS slope parameter ($\rho = 1$) in the absence of any drift. Figure 9 reveals a heavily skewed distribution, slowly converging to the long-run parameter 0.7. However, in case of small sample sizes estimates smaller than 0.7 (the true value) are rare events, and most outcomes lie beween 0.7 and 0.75, and many estimates are larger than 0.8 or even 0.85. Thus, applied researchers who have to rely on small samples sizes should be aware of this particular shape of the distribution function. As can be seen from the fractiles of $\hat{\beta}$ presented in Figure 10, results of about 0.78 are - despite super-consistency - not significantly different from the true parameter 0.7 when the sample size is smaller than 40 observations, whereas lower critical values can be found already for estimates of about 0.70.

Figure 10 contains some interesting statistical properties of the model (i)-(iii). The average DW-statistic converges to two, t-values are linearly increasing and the R^2-statistic approaches unity. The last point is related to an increasing signal-noise ratio of the cointegrating regression and will be analysed in some more detail below. The performance of the DW-statistic is likewise affected by the growing R^2: Observed error terms contain only white noise in the long run.

Table 1 gives a general impression of the importance of the various parameters under consideration. First, unless $\rho = 1$, OLS estimates of β are persistently higher than the true parameter 0.7 so that in case of I(0) variables the Haavelmo bias will keep in place irrespective of sample size. Even so, the higher the autocorrelation of the exogenous process is, the smaller is the simultaneous equation bias. Nevertheless, even for the high coefficient $\rho = 0.75$, the bias remains substantial (0.79 instead of 0.70 for T=500). Moreover, as for I(1) variables, the bias becomes smaller with increasing T in case of stationary variables with $\rho > 0$, though the reduction is almost invisible (it is more profound in the non-zero drift; see Table 1b).

Table 1: The impact of sample size, drift and degree of serial correlation on the simultaneous equation bias

a) $\gamma = 0$

	T=20	T=80	T=150	T=500
$\rho=0$	0.8490 (26.2)	0.8496 (51.1)	0.8499 (69.6)	0.8505 (126.7)
$\rho=0.75$	0.8042 (26.7)	0.7932 (51.6)	0.7926 (70.6)	0.7919 (128.5)
$\rho=1$	0.7524 (37.1)	0.7176 (124.7)	0.7102 (223.9)	0.7031 (742.8)

b) $\gamma = 0.5$

	T=20	T=80	T=150	T=500
$\rho=0$	0.8333 (25.6)	0.8335 (50.4)	0.8334 (68.9)	0.8335 (125.0)
$\rho=0.75$	0.7575 (31.0)	0.7445 (63.4)	0.7433 (87.4)	0.7417 (160.3)
$\rho=1$	0.7154 (67.6)	0.7008 (487.9)	0.7002 (1252.4)	0.70001 (7566.7)

Note: Each entry is based on 2000 replications of the simulation model (i)-(iii); T, γ and ρ denote sample size, drift and degree of serial correlation, respectively; the figures represent the average OLS estimate $\hat{\beta}$, as well as corresponding (conventional) average t-values (in parentheses).

The bias is further reduced in the presence of drifts, as can be seen by comparing Table 1,a) and Table 1,b). The simulation experiment imposes γ=0.5, thus assuming a drift within the range being estimated by Hylleberg and Mizon (1989). Again, the reduction of the bias becomes more profound when ρ is high. In the extreme case of unit roots, we observe a very quick adjustment in the presence of drifts, but a slow adjustment to the asymptotic result in the absence of drifts.

As expected from the econometrics of stationary variables, refering to conventional t-values without checking for potential endogeneity problems leads to wrong conclusions also for I(0) variables: All t-values pretend reliable estimates. In case of I(1) simulations we have to be aware of the non-standard distribution of OLS estimates (Stock, 1987), here presented in Figure 9, and in case of regressors having a drift, of necessary transformations of the computer output t statistic in order to derive a "t-statistic" on $H_0{:}a = 0$ (see West, 1988).

Table 2: The role of drifts

	$\beta = 0.7$			$\beta = 0.3$		
	γ=0	γ=0.25	γ=0.75	γ=0	γ=0.25	γ=0.75
T= 30	0.739	0.723	0.702	0.389	0.352	0.305
T= 90	0.716	0.704	0.7002	0.337	0.308	0.3004
T=250	0.706	0.7003	0.70003	0.316	0.3008	0.30004

Note: Each entry is based on 1000 replications of the simulation model (i)-(iii), imposing ρ = 1. Two alternative true values are considered: $\beta = 0.7$ and $\beta = 0.3$; T and γ denote sample size and drift; the figures represent the average OLS estimate $\hat{\beta}$

Table 2 highlights the substantial role of drifts on the performance of superconsistent estimates. Whereas random walks X_t without drift result in a considerable bias, models including significant drifts reveal almost no bias even for small samples (see γ=0.75 which leads to a rounded estimate 0.70 even for T=30). Again, a simple intuition of this result can be given by the representation of X_t as $X_t = \gamma t + \Sigma v_j$ and the dominant role of deterministic trends (confer the arguments in the Appendix, or West's (1988) analysis of variables of order one with drift, for instance). Deterministic trends which are of a higher order than the stochastic counterparts lead to a negligible impact of the stochastic trend in the long-run. Hence, when the deterministic part is particularly strong because of a high coefficient of time γ, the estimate approaches the asymptotic result very quickly so that the "long-run" can be very short.[11]

Table 2 suggests that the ratio "bias / true value" is higher if the true value is small. This result is based on a positive relationship between the signal-noise ratio and β, and on the fact that a high signal-noise ratio coincides with a small simultaneous equation bias (see Figure 10 and the discussion of signal-noise ratios below). Figure 11 displays the bias $(\hat{\beta} - \beta)$ as a function of β.

The relationship between the average bias and β is strictly negative with widely spread and positive errors for small β; for $\beta \rightarrow 1$ the distribution converges to the true value (so that R^2 approaches unity).

Of course, irrespective of the presence of drifts, the simultaneous equation bias would be even higher in the *absence* of cointegration, as Figure 12 displays. It illustrates the slow adjustment of the OLS distribution from no serial correlation at all (i.e. X_t is white noise) to the limiting case of unit roots (such that C_t and Y_t are cointegrated). We consider a small sample with T=30; the degree of persistence, ρ, is increased from 0 to 1 with increments 0.05. The histogram reveals a marked step towards "perfect" cointegration, i.e. from $\rho = 0.95$ to $\rho = 1$, where the distribution changes more than for smaller degrees of serial correlation.

The higher the distance from unit roots is, the higher would be the probability of a (false) conclusion that the estimate is significantly different from the imposed true value: For $\rho = 0$ we observe a rather symmetric distribution with mean 0.85 and a probability for estimates smaller than 0.7 being almost zero.

It is interesting to study the results of model (i)-(iii) in the neighbourhood of $\rho = 1$ (Figure 13). Particularities can be observed for t-values, which remain almost constant up to $\rho = 0.7$, and which rise very quickly thereafter. DW-statistics and OLS estimates adjust slowly; R^2 increases significantly for ρ larger than 0.90.

Having identified potential drawbacks arising due to the presence of the Haavelmo bias, the next step according to elementary text book reasoning is to eliminate the bias running Theil's (1953) two-stage-least-squares (TSLS). In the following, we try the same applying cointegrated variables.

Figure 14 contrasts the unit root case with the performance of TSLS in stationary regressions. We simulate small samples (T=30). The empirical distribution reveals that TSLS bears some problems for stationary time series, as is to be expected because of the small sample and the consistency property of the TSLS estimator. Rare but extreme outliers lead to a very skew distribution function, where minimal estimates are about -1.62 (average estimate = 0.688 for $\rho = 0$).[12] However, the distribution becomes sharply peaked and concentrated on the true value when ρ increases (range=[0.53,0.78] and average estimate = 0.700 in case of $\rho = 1$).[13] Thus, the application of a conventional TSLS estimator instead of the "superconsistent" OLS estimator still improves the performance in the presence of cointegrated relationships, at least in small samples and for low signal-noise ratios, as is discussed below.[14]

From Figure 10 we know about increasing R^2 for growing T. The contribution of noise in the "consumption function" $C_t = \beta Y_t + \varepsilon_t$ - held constant because of $\varepsilon_t \sim N(0,1)$ in the data generation process - becomes less and less important relative to the signal of the equation, i.e. βY_t, which is determined by the infinitely growing variance of the exogenous random walk X_t. Actual time series, however, are perhaps more realistically represented by a likewise growing variance of the error term in the simultaneous cointegrating equation. Thus our final experiments deviate from the assumption $\varepsilon_t \sim N(0,1)$ and adjust the variance σ_ε^2 of the error term ε_t so that the noise-signal ratio remains constant.[15]

Inspecting the reduced form (ii') $C_t = \beta/(1-\beta)X_t + 1/(1-\beta)\varepsilon_t$ of the cointegrating equation (ii), the ratio between the exogenously determined signal and the noise term can be expressed as

$$\frac{\text{signal}}{\text{noise}} \rightarrow r \equiv \frac{\sigma_x^2\beta^2/(1-\beta)^2}{\sigma_c^2/(1-\beta)^2}.$$

The variance of the exogenous process (no drift) evolves according to $\sigma_x^2(t) = t\sigma_v^2 = t$. On the other hand, $\sigma_c = 1$ is held constant so that r goes to infinity and R^2 approaches unity. Observed time series regressions, however, have R^2s well below unity - irrespective of the number of observations. Hence it is important to study the sample-size dependent behaviour of the Haavelmo bias when signal-noise ratios or the R^2 are held constant.

The practical implementation uses a somewhat different approach, but is still following the general idea.[16] We first determine the exogenous time series $X_t = X_{t-1} + v_t$,t=1,2,...T. Second, the empirical variance of $X_1, X_2, \ldots, X_T$, $S_x^2(T)$,is estimated.[17] In the third step, we calculate $\sigma_c^2(T) = \beta^2 S_x^2(T)/r$ and we generate equations (ii') and (iii) assuming $\varepsilon_t \sim N(0, \sigma_c^2(T))$. The variance is then adjusted for each new simulation, depending on sample size T.

Figure 15 summarizes the results of the cointegrating OLS regression. Imposing r=2 in the up-dating formula $\sigma_c^2(T) = \beta^2 S_x^2(T)/r$ leads to R^2s about 0.973. Moreover, Figure 15 contains some information about *estimated* noise-signal ratios. We calculate "(signal - estimated noise)/(signal)", i.e.

$$(1 - r_{est.}) \equiv 1 - \frac{(C-\hat{\beta}Y)'(C-\hat{\beta}Y)}{(\hat{\beta}Y)'(\hat{\beta}Y)}.$$

As intended by construction, the measure remains constant at values about 0.983.

The most important consequence of imposed constant signal-noise ratios is the estimate of β, remaining almost constant at 0.73, implying thus a persistent simultaneous equation bias - even in the "long run". The power of superconsistency is offset by growing noise of the error term in the cointegrating relationship.[18] The speed of convergence - if any - of the distibution function of $\hat{\beta}$ is very low (see the presentation of fractiles in Figure 15).[19]

Again, TSLS instead of OLS for cointegrating equations leads to estimates approaching the true value even for small T. Figure 16 is based on the same construction $\sigma_c^2(T)$ as before. As for TSLS simulation in case of small ρ (see Figure 14), we observe extreme outliers leading to a very skewed distribution for small samples. For T larger than 50, however, outliers die out almost completely, the distribution becomes markedly peaked, and the average estimate coincides almost perfectly with the true value.[20]

The results confirm a related simulation study by Hansen and Phillips (1990). They also find evidence that the signal-noise ratio is the critical factor among possible problems concerning the estimation and inference in models of cointegration, not the degree of long-run endogeneity or serial correlation. Similar to Phillips and Hansen (1990), we conclude that OLS works well for high signal-noise ratios so that the bias is neglible in this case. If, however, the relative signal is low, TSLS techniques easily outperform ordinary least squares.

4. Conclusions.

This paper illustrates finite sample problems of regression models with I(1)-variables. For the purpose of graphical exposition, simulation results are summarized in three-dimensional distribution functions. They give a compact impression of well known problems arising from the simultaneous equation bias which should vanish asymptotically when cointregrated relationships occur. The second main point discussed in the paper is the distribution of the regression estimate in the presence of non-cointegrated I(1) variables. Here as well as in the case of the simultaneous equation bias, we focus on the particular role of drifts.

Applied researchers should be cautioned against relying on the "mystery" of cointegration when dealing with Haavelmo's (1943) classical simultaneous equation bias: The distribution of the estimated cointegrating vector reveals a flat curvature and a strong asymmetry in small samples, still leading to a potential bias when we would solely rely on the power of cointegration. We have to be aware of the fact that - despite superconsistency - the problem arising from the endogeneity of both regressand and regressor vanishes only "in the long run" - which is indeed a very long run when regressors have no drift, or when the true slope parameter is well below unity (cf. Figure 11). Moreover, superconsistency is based on the fact that the relative importance of the noise term in the cointegrating relation vanishes with respect to the signal when $T \rightarrow \infty$. Assuming instead *constant* noise-signal ratios leads to a persistent and even constant simultaneous equation bias.

The presence of drifts, on the other hand, substantially reduces the bias in small samples. The impact of drifts, their effect on equivalent deterministic trends of the transformed model, and the implication of the faster convergence of these deterministic parts compared to stochastic trends has been investigated in several simulation studies. From these one may conclude that the introduction of drifts can easily have more significant effects than extending the sample size (cf. Table 2, for instance: For T=250 and no drift the average estimate is 0.706, for T=30 but drift included ($\gamma = 0.75$) the corresponding estimate amounts to the almost identical estimate 0.702).

A further experiment is concerned with drifts in stationary models. We observe the reduction of the simultaneous equation bias also in this case, a point being related to the vanishing simultaneous equation bias in the presence of deterministic trends and in this context analysed by Krämer (1984), for instance.

TSLS would be suitable in case of stationary models. In fact, applying the same technique for cointegrated I(1)-variables leads to a strong and quick reduction of the simultaneous equation error when compared to the superconsistent OLS estimator. As before, the presence of a drift would additionally increase the speed of convergence.

Despite presented reservations against being overly optimistic in case of small samples, some comparisons between stationary models and cointegrated models highlight the advantages cointegrated models still would have. We observe a distribution of OLS estimates for stationary regressions which definitely lead to biased estimates when endogeneity problems arise. Though this bias would be less severe after running TSLS, the combination of both TSLS *and* cointegration turn out to be more reliable than TSLS in stationary models.

The role of the drift is also decisive in case of regressions applying integrated (but not cointegrated) variables. Time-dependent histograms point out the difference between the convergence to random variables in the spurious regression case without drifts (made popular by Granger and Newbold, 1974, and analysed by Phillips, 1986) and the convergence to constants in case of regressing random walks with drifts (cf. the Appendix). Again, we see that drifts lead to dominating deterministic trends, and the higher the drift is, the faster is the dominance over the stochastic elements of the OLS estimator, and the faster is the speed of convergence.

References:

Banerjee, A., J.J. Dolado, D.F. Hendry and G.W. Smith (1986), Exploring equilibrium relationships in econometrics through static models: Some Monte Carlo evidence, *Oxford Bulletin of Economics and Statistics* 48, 253-278

Banerjee, A. and D.F. Hendry (1992), Testing integration and cointegration: An overview, mimeographed, forthcoming *Oxford Bulletin of Economics and Statistics*

Drèze, J.H. and C.R. Bean, Eds. (1990), *Europes's Unemploymnet Problem,*, MIT-Press

Engle, R.F. and C.W.J. Granger (1987), Co-integration and error correction: Representation, estimation, and testing, *Econometrica 55*, 251-276

Engle, R.F. and C.W.J. Granger (1991), Introduction to *Long-Run Economic Relationships: Readings in Cointegration,* edited by R.F. Engle and C.W.J. Granger, Oxford University Press

Entorf, H. (1992), Supplement to "Random walks with drift, simultaneous equation errors and small samples", mimeographed, Catholic University of Louvain-la-Neuve

Granger, C.W.J. and P. Newbold (1974), Spurious regression in econometrics, *Journal of Econometrics 2*, 111-120

Haavelmo, T. (1943), The statistical implications of a system of simultaneous equations, *Econometrica* 11, 1-12

Hansen, B.E. and P.C.B. Phillips (1990), Estimation and inference in models of cointegration: A simulation study, in T.B. Fomby and G.F. Rhodes (eds.), *Advances in Econometrics, Vol. 8, Cointegration, Unit Roots and Spurious Regression,* JAI Press

Hylleberg, S. and G.E. Mizon (1989), A note on the distribution of the least squares estimator of a random walk with drift, *Economics Letters* 29, 225-230

Krämer, W. (1984), On the consequences of trend for simultaneous equation estimation, *Economics Letters* 14, 23-30

Nelson, C.R. and C. Plosser (1982), Trends and random walks in macroeconomic time series, *Journal of Monetary Economics* 10, 139-162

Phillips, P.C.B (1986), Understanding spurious regression in econometrics, *Journal of Econometrics 33*, 311-340

Phillips, P.C.B. and B.E. Hansen (1990), Statistical inference in instrumental variables regression with I(1) processes, *Review of Economic Studies* 57, 99-125

Sargan, J.D. and A. Bhargava (1983), Testing residuals from least squares regression for being generated by the Gaussian random walk, *Econometrica* 51, 153-174

Sims, C.A., J.H. Stock and M.W. Watson (1990), Inference in linear time series models with some unit roots, *Econometrica 58*, 113-144

Stock, J.H (1987), Asymptotic properties of least squares estimators of cointegrating vectors, *Econometrica 55*, 1035-1056

Theil, H. (1953), Repeated least-squares applied to complete equation systems, mimeograhed, The Hague: Central Planning Bureau

West, K.D. (1988), Asymptotic normality, when regressors have a unit root, *Econometrica 56* , 1397-1417

Yule, G.U. (1926), Why do we sometimes get nonsense-correlations between time series?, *Journal of the Royal Statistical Society* 89, 1-69

Appendix.

Proof of Theorem 1

We write x_t and y_t as

$$x_t = t\gamma_x + S_t^x,\ y_t = t\gamma_y + S_t^y,$$

where

$$S_t^x = \sum_{j=1}^{t} \nu_j,\ S_t^y = \sum_{i=1}^{t} \varepsilon_i,$$

and $\nu_0 = \varepsilon_0 = 0$. The OLS estimator of β is

$$\beta = \frac{\sum_{t=1}^{T} (x_t - \bar{x})(y_t - \bar{y})}{\sum_{t=1}^{T} (x_t - \bar{x})^2} = \frac{\frac{1}{T^3}\Sigma x_t y_t - \frac{1}{T^2}\overline{xy}}{\frac{1}{T^3}\Sigma x_t^2 - \frac{1}{T^2}\overline{x}^2}$$

Next, we consider the convergence properties of individual terms:

$$\lim_{T\to\infty} \frac{1}{T^2}\overline{xy} = \lim_{T\to\infty} \frac{\Sigma(t\gamma_x + S_t^x)}{T^2} \lim_{T\to\infty} \frac{\Sigma(t\gamma_y + S_t^y)}{T^2}.$$

We know from Phillips (1986), Lemma 1, that

$$\frac{1}{T^{3/2}} \sum_{t=1}^{T} S_t^x \Rightarrow \sigma_\nu \int_0^1 V(t)dt \qquad (A1)$$

so that this term as well as ΣS_t^y vanish when we divide by T^2 instead of $T^{3/2}$.

It remains

$$\lim_{T\to\infty} \frac{\gamma_x}{T^2} \sum_{t=1}^{T} t \to \frac{\gamma_x}{2},\ \lim_{T\to\infty} \frac{\gamma_y}{T^2} \sum_{t=1}^{T} t \to \frac{\gamma_y}{2}$$

so that

$$\frac{1}{T^2}\overline{xy} \Rightarrow \frac{\gamma_x\gamma_y}{4},\ \frac{1}{T^2}\overline{x}^2 \Rightarrow \frac{\gamma_x^2}{4}.$$

Next we consider

$$\lim_{T\to\infty} \frac{1}{T^3} \sum_{t=1}^{T} x_t y_t = \lim \frac{\Sigma(t\gamma_x + S_t^x)(t\gamma_y + S_t^y)}{T^3}$$

$$= \lim \gamma_x\gamma_y \frac{\Sigma t^2}{T^3} + \lim \gamma_x \frac{\Sigma t S_t^x}{T^3} + \lim \gamma_y \frac{\Sigma t S_t^y}{T^3} + \lim \frac{\Sigma S_t^x S_t^y}{T^3}.$$

From Phillips (1986, Lemma 1) and Sims, Stock and Watson (1990, Lemma 1), we know the weak convergence respectively convergence of individual terms:

$$\frac{1}{T^{5/2}}\Sigma tS_t^x \Rightarrow \int_0^1 t\sigma_v V(t)dt,\ \frac{1}{T^{5/2}}\Sigma tS_t^y \Rightarrow \int_0^1 t\sigma_\varepsilon W(t)dt$$

$$\frac{1}{T^2}\Sigma S_t^x S_t^y \Rightarrow \sigma_v\sigma_\varepsilon \int_0^1 V(t)W(t)dt,\ \frac{1}{T^3}\Sigma t^2 \rightarrow \frac{1}{3}.$$

Given the different rates of convergence $T^{-5/2}$, T^{-2} and T^{-3}, the only remaining term is the quadratic trend term so that

$$\frac{1}{T^3}\sum_{t=1}^{T} x_t y_t \Rightarrow \frac{\gamma_x\gamma_y}{3},\ \frac{1}{T^3}\sum_{t=1}^{T} x_t^2 \Rightarrow \frac{\gamma_x^2}{3}.$$

Collecting terms leads to

$$\beta \Rightarrow \frac{\gamma_x\gamma_y\frac{1}{3}-\gamma_x\gamma_y\frac{1}{4}}{(\gamma_x)^2\frac{1}{3}-(\gamma_x)^2\frac{1}{4}} = \frac{\gamma_y}{\gamma_x}.$$

q.e.d.

Since $\hat{\alpha} = \bar{y} - \beta\bar{x}$, we can write

$$\lim \hat{\alpha}T^{-1/2} = \lim \frac{1}{T^{3/2}}\Sigma(t\gamma_y + S_t^y) - (\lim \beta)\left(\lim \frac{1}{T^{3/2}}\Sigma(t\gamma_x + S_t^x)\right).$$

Applying the convergence of β, the trend terms drop out and again using Phillips' Lemma (see A1), we have the result

$$\hat{\alpha}T^{-1/2} \Rightarrow \sigma_\varepsilon \int_0^1 W(t)dt - \frac{\gamma_y}{\gamma_x}\sigma_v \int_0^1 V(t)dt.$$

q.e.d.

Footnotes

1. See, for instance, Engle and Granger (1991, p.9): "... the regression can be run with either variable as the dependent variable, and there is no correction for simultaneous equation bias or serial correlation".

2. Given annual time series, a sample period 1961-1990 would lead to the small sample of T=30 observations. This is a realistic case in many macroeconomic applications (see, for instance, the country models of the "European Unemployment Programme", published in Drèze and Bean, 1990).

3. Despite a more general use of the expressions "nonsense regression" and "spurious regression" since Yule's (1926) seminal work, here the term is limited to regressions applying random walks *without* drift (thus in the narrow sense of simulation experiments by Granger and Newbold, 1974, and the mathematical analysis by Phillips, 1986).

4. Including a linear trend does change the results as follows: Estimating $y_t = \hat{\alpha} + \hat{\beta} x_t + \hat{\delta} t + \hat{u}_t$ leads to the convergence of $\hat{\delta}$ to γ_y and $\hat{\beta}$ does converge to a random variable (Entorf, 1992).

5. The random numbers are generated applying the RNDN command of GAUSS.

6. The three-dimensional histograms in the paper are constructed using GAUSS Puplication Quality Graphics. At each step (for instance time units), most of the indivividual simulations are replicated 5000 times and the results are captured using 30 different categories. All GAUSS programs can be received from the author by request.

7. In their famous simulation experiment, Granger and Newbold (1974) report a rejection rate of 76 percent (50 observations, 100 replications).

8. Results reported elsewhere (Entorf, 1992) show that *non-zero* drifts lead to the same pattern of the three-dimensional distribution (depending on σ_y/σ_x), with 5% and 95%-fractiles diverging slower than in the case without drift. However, t-values decrease with increasing σ_y/σ_x and DW-statistics are growing slowly.

9. Some unreported simulations reveal that, given imposed parameters of Figure 6, after T=7000 the average constant term is 0.53, and even after T=12000 it is 0.33. Experiments with much higher drifts ($\gamma_x = 5$), however, lead to an average estimate of about 0.05 for T=100 (500 replications in either case). It should be noted, however, that these averages have almost no practical meaning, since the variance of the estimate goes to infinity.

10. Alternative simulations with imposed drifts lead to the identical empirical distribution of the DW-statistic.

11. It should be noted that the general suggestion to include trends because of presumed drifts (confer, for instance, Hansen and Phillips, 1990) is misleading for the considered cointegrated model under consideration:
$C_t = \hat{\alpha} + \hat{\beta} Y_t + \hat{\delta} t + \hat{u}$ leads to $\hat{\beta} = 0.807, 0.746, 0.720$ in case of $\gamma = 0.75$, $\beta = 0.7$ and T= 30, 90, 250 (without constant term: 0.792, 0.743, 0.717). Thus, the bias is even higher.

12. A similar result is reported by Hansen and Phillips (1990) who observe a fat tail property for their simulation study of IV techniques.

13. For cointegrated variables, West (1988) showed the consistency and normalilty of TSLS estimators when regressors have a drift, which, however, is absent in the simulations underlying Figure 14.

14. Non-reported results show that this result holds a fortiori in case of drifts, which lead to almost perfect estimates even for very small sample sizes of less than 30 observations.

15. I owe the suggestion of analysing constant noise-signal ratios to Jan Kiviet.

16. The *direct* approach would be to calculate $\sigma_c^2(t) = \beta\sigma_x^2(t)/r = \beta t/r$ and to generate $\varepsilon_t \sim N(0,\sigma_c^2(t))$ at each t. This, however, implies nonstationarity of ε_t and thus non-cointegration.

17. The expected value is

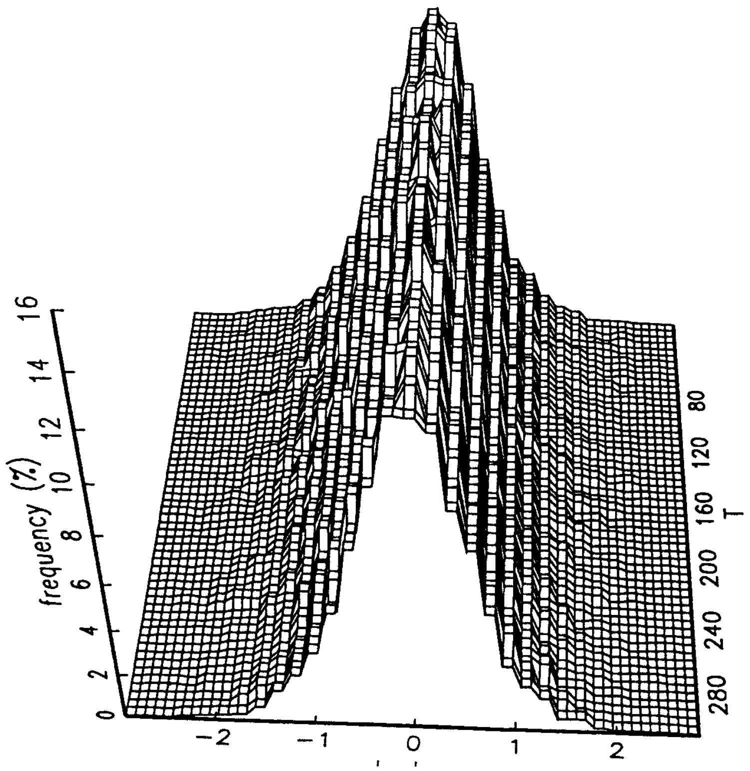
$$E\left(\left(\frac{1}{T}\right)\Sigma x_t^2 - \bar{x}^2\right) = \frac{(T+1)}{2} - \frac{T(T+1)(2T+1)}{6T^2} = (T+1)\frac{(1-1/T)}{6} \rightarrow \frac{T}{6}.$$

18. A comparison with Figure 10 reveals that we "freeze" the situation of T=40 with $\sigma_c^2 = 1$: There we find almost the same R^2 and β.

19. Alternative calculations (Entorf, 1992) imposing r=3 lead to a slightly incresing β (from 0.721 at T=30 to 0.722 at T=310) with a slowly converging distribution. $(1 - r_{est})$ turns out to be constant at 0.988; R^2 increases slowly (from 0.978 to 0.979).

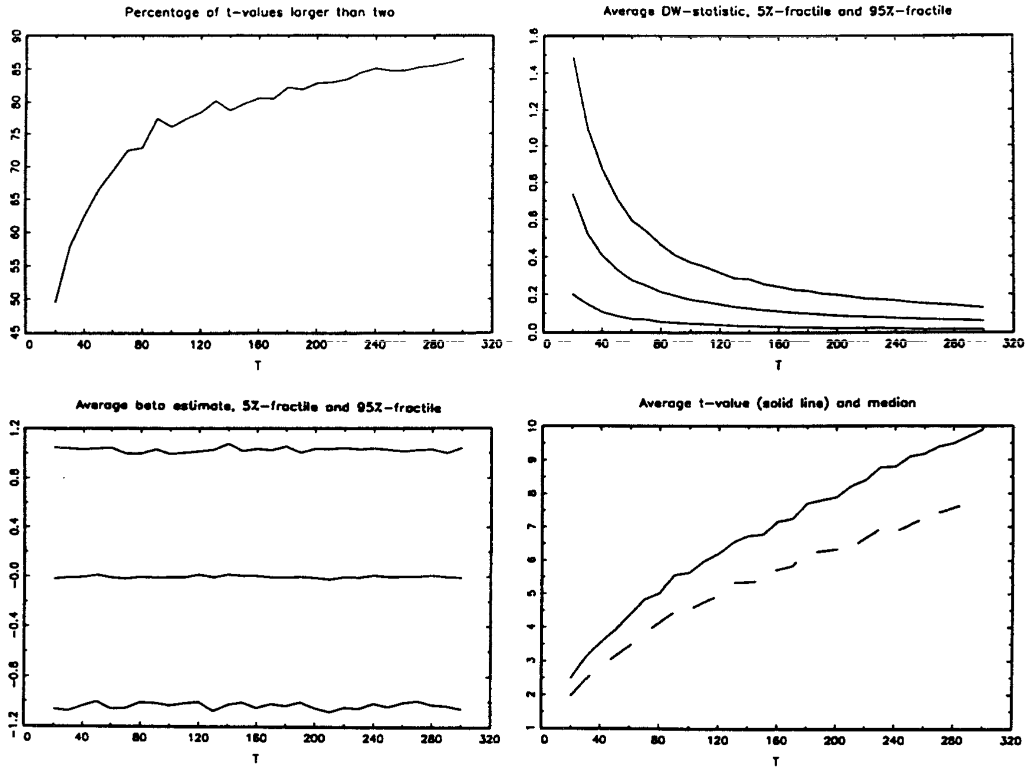
20. Average estimates are 0.698 for T=20, and 0.699 for T=90. Results reported elsewhere (Entorf, 1992) confirm reported constant signal-noise ratios and the shape of the TSLS estimator using two-dimensional summary statistics of presented simulated regressions.

Figure 1: Time-variant histogram of the OLS estimate: Random walks have no drift



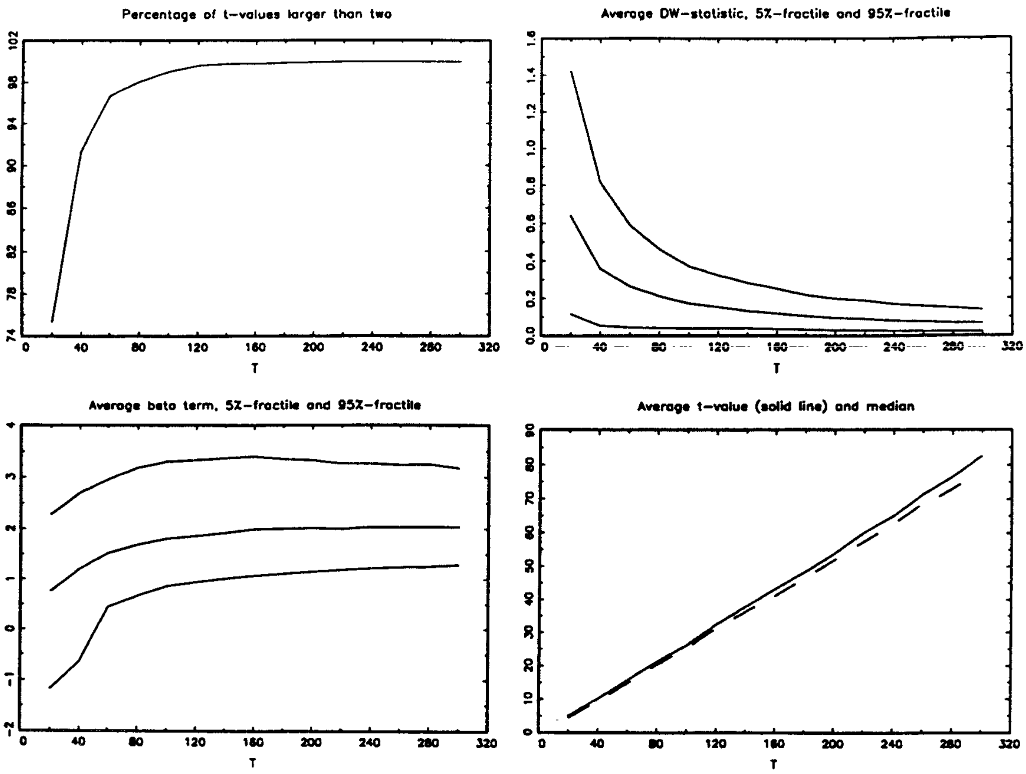
Note: Regressing two random walks as described in (1)-(2) with zero drifts; the length of the simulated time series, T, is extended as follows: T=20,30,...,300; 5000 replications of (1) and (2) are calculated for each T; the graphic displays the time-variant histogram (30 categories) of $\hat{\beta}$. 79 outliers outside the range [-2.8, 2.8] are skipped; extreme values: -5.23, 3.94.

Figure 2: Regressing random walks without drift: Summary graphics



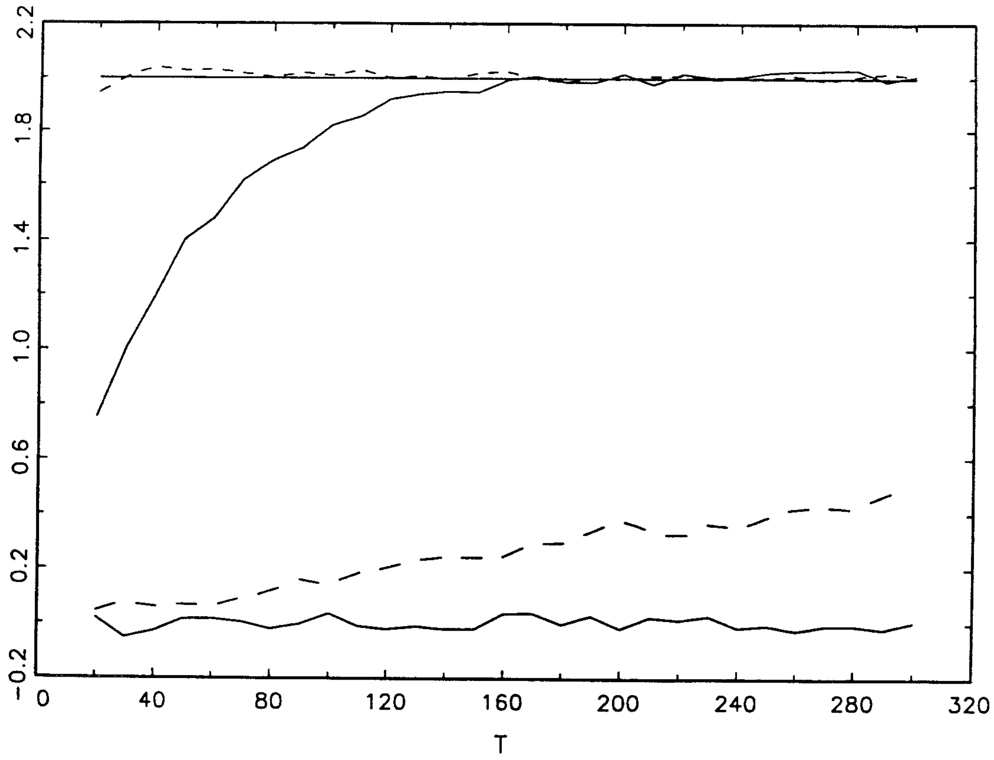
Note: The calculations are based on the simulation experiment described in Figure 1.

Figure 3: Regressing random walks with drift: Summary graphics



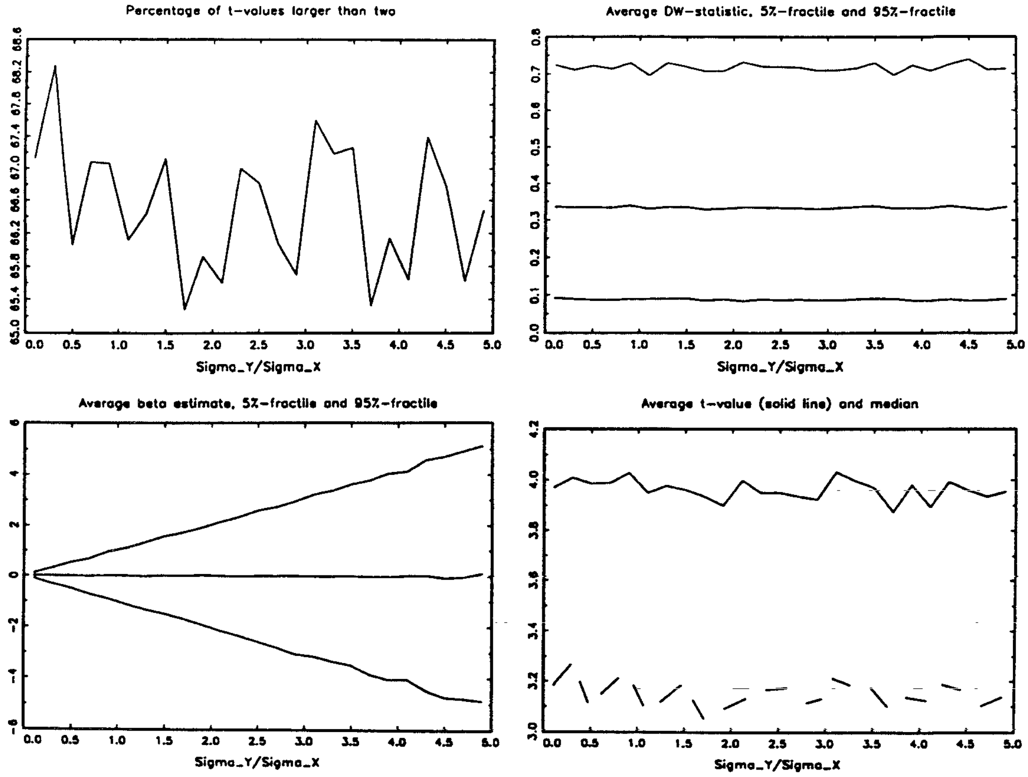
Note: Regression of two random walks as described in (1)-(2); drifts are non-zero: $\gamma_y = 0.5, \gamma_x = 0.25$; the length of the simulated time series, T, is extended as T=20,30,...,300; 5000 replications of (1) and (2) are calculated for each sample size T;

Figure 4: The impact of drift on the speed of convergence



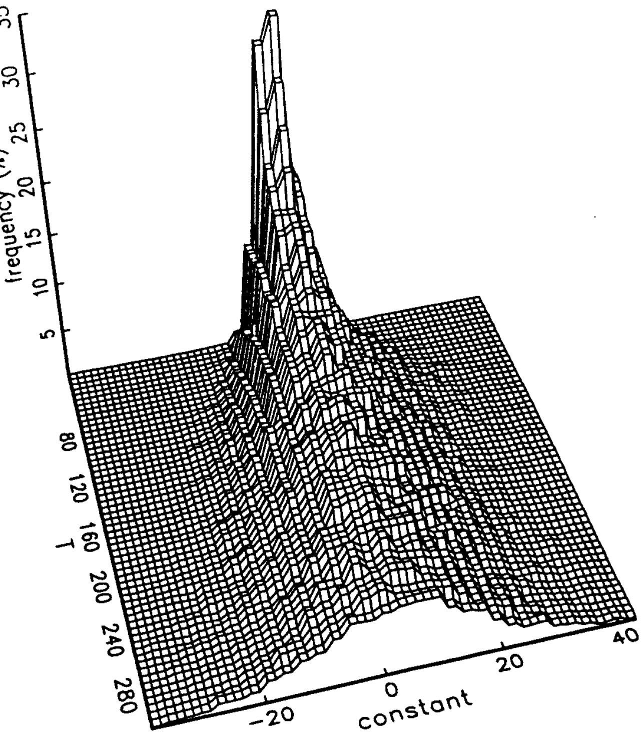
Note: Regression of two random walks as described in (1) and (2); sample sizes T=20,30,...,300; for each T the model is replicated 1000 times; four alternative drifts are considered: $\gamma_x = 0, 0.05, 0.25, 0.75$, $\gamma_y = 2\gamma_x$; the graphic displays the average OLS estimator β depending on drift and T ($\gamma_x = 0$: bottom line, $\gamma_x = 0.75$: top line).

Figure 5: The impact of σ_y/σ_x on nonsense regressions



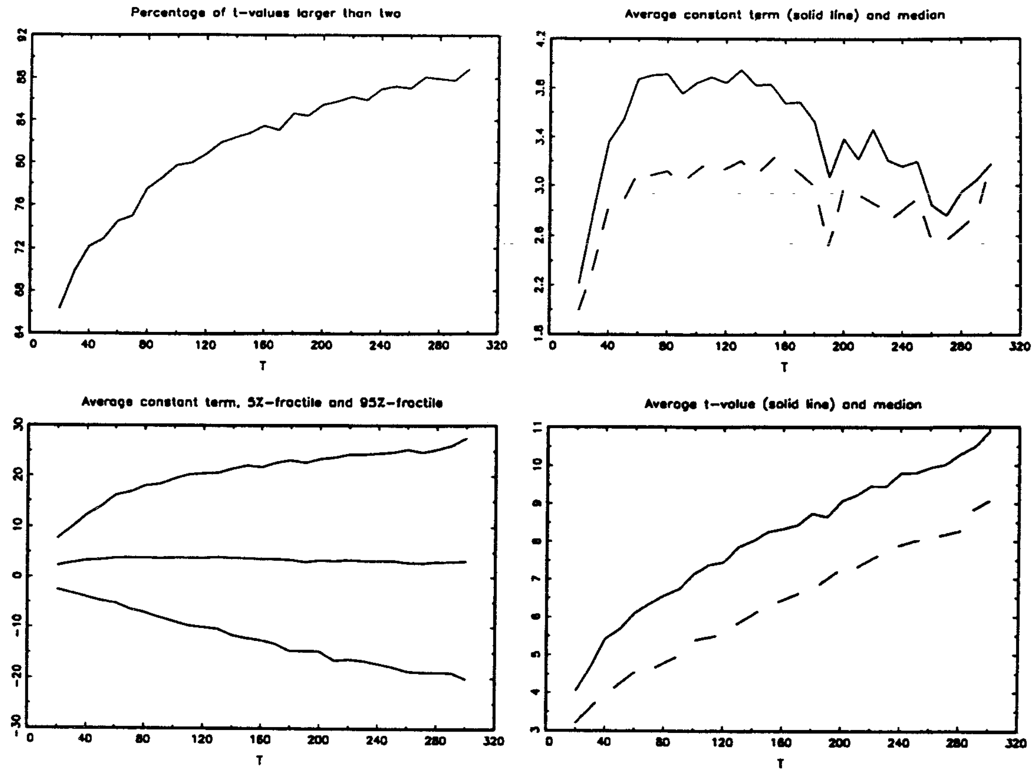
Note: Regression of two random walks as described in (1) and (2) without drift; T=50; the ratio of the standard errors σ_y/σ_x is extended as 0.1, 0.3, 0.5, ..., 4.9; $\sigma_x = 1$; for each ratio the model is replicated 5000 times.

Figure 6: Diverging constant terms in I(1)-regressions



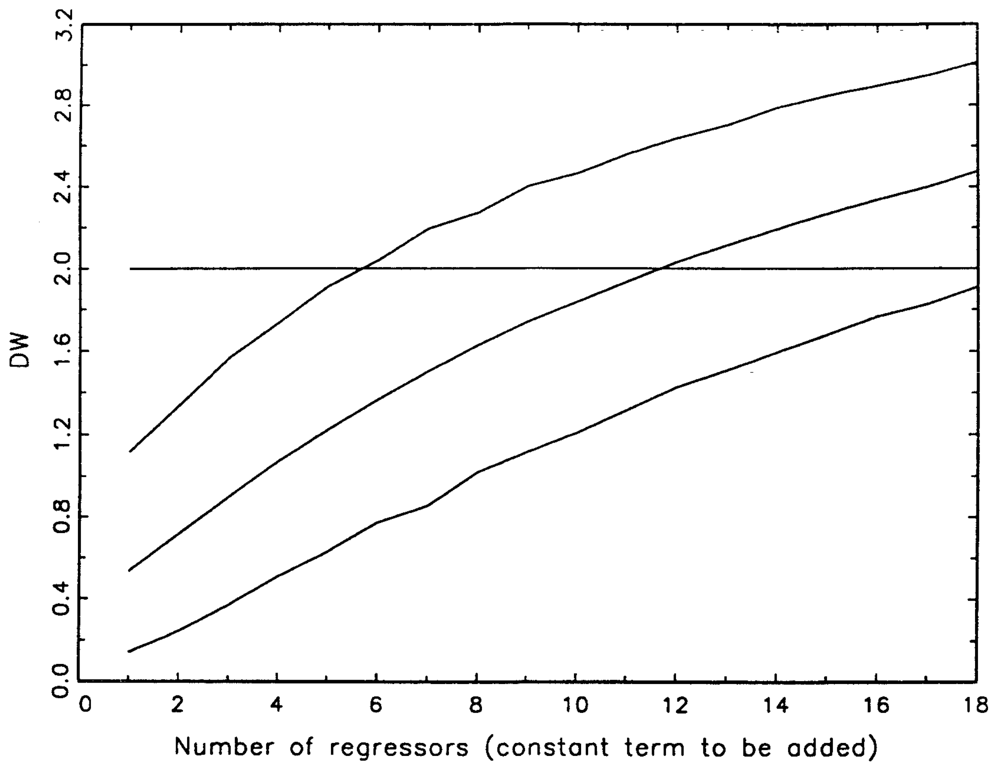
Note: Regression of two random walks as described in (1)-(2); drifts are non-zero: $\gamma_y = 0.5, \gamma_x = 0.25$; the length of the simulated time series, T, is extended as T=20,30,...,300; 5000 replications of (1) and (2) are calculated at each sample size T; the graphic displays the time-variant histogram (30 categories) of $\hat{\alpha}$. 531 outliers outside the range [-37,43] are skipped; extreme values: -60.6, 82.7.

Figure 7: Diverging constant terms in I(1)-regressions: Summary graphics



Note: See Figure 6.

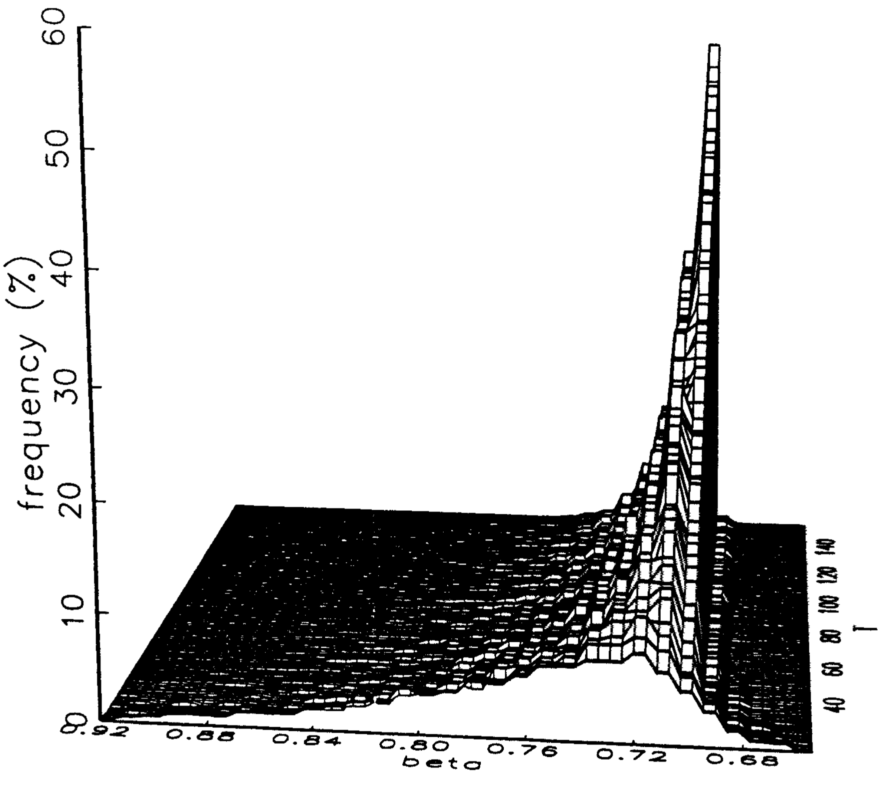
Figure 8: DW-statistics and nonsense regressions



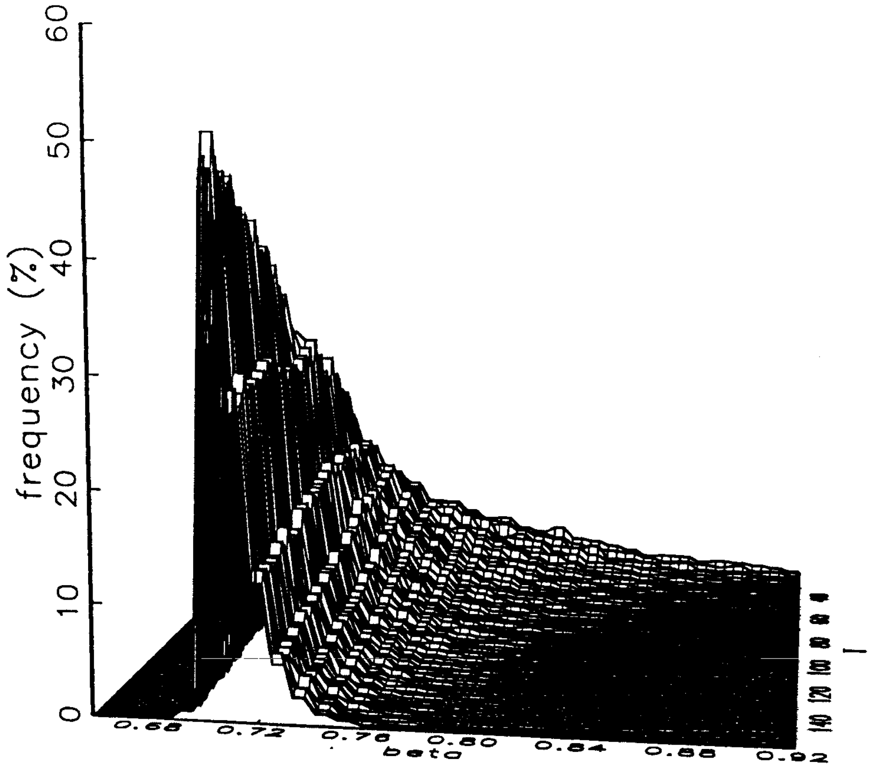
Note: Regression of $y_t = \alpha + \beta x_t, x_t$ being an i.i.d. random walk (without drift) with dimension K=1, 2, 3, ... 18; T=30. For each K the model is replicated 5000 times; the graphic displays average DW-statistics as well as 5% and 95% fractiles of the DW distribution.

Figure 9: The adjustment of the cointegrating OLS estimator towards its asymptotic distribution

Front view:

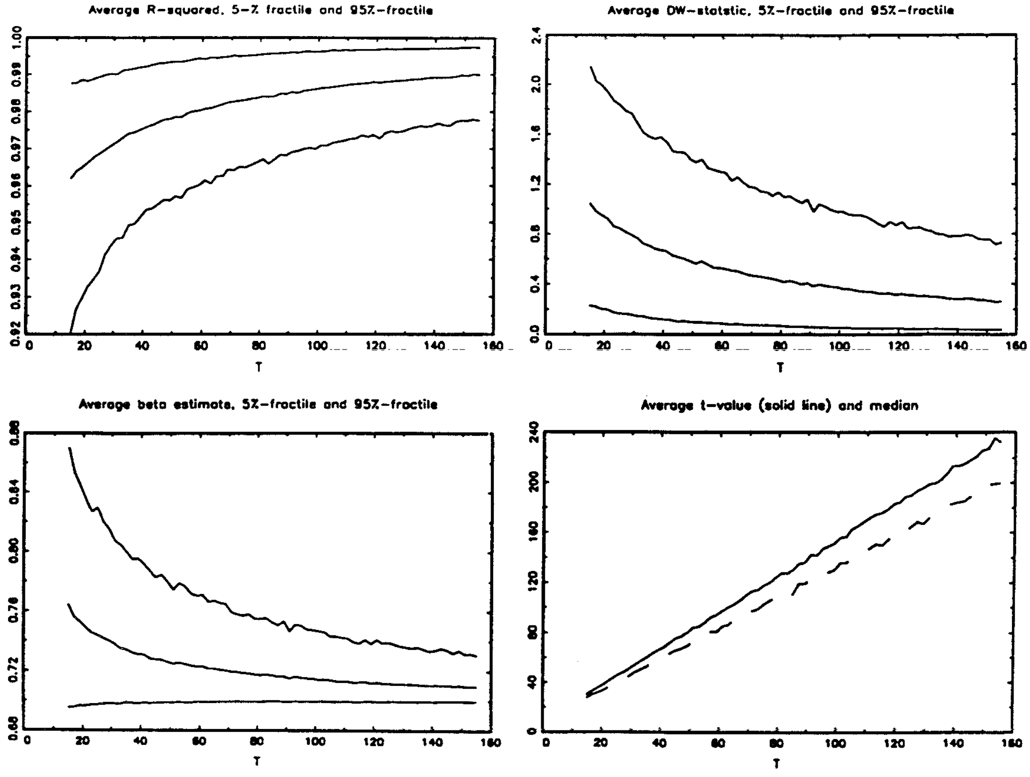


Back view:



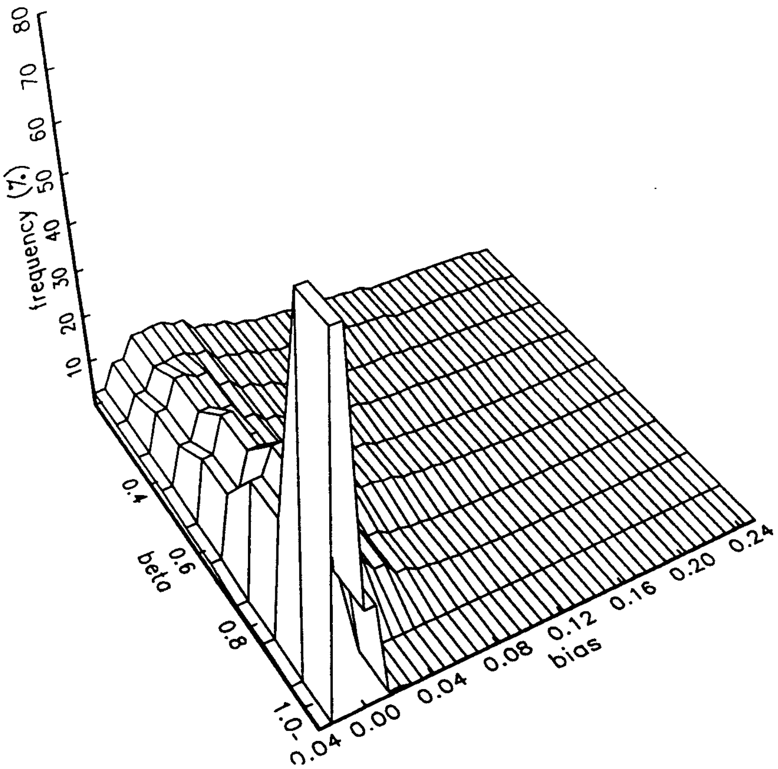
Note: Figure 9 captures 3000 replications of model (i)-(iii) for each sample size T, T=15,17,19,...,155; the exogenous process is a random walk without drift; the graphic displays the time-dependent histogram (30 categories) of the estimator $\hat{\beta}$. 191 outliers outside the range [0.66,0.92] are skipped; extreme values: 0.63, 1.02.

Figure 10: The adjustment of the cointegrating OLS estimator towards its asymptotic distribution: Summary graphics



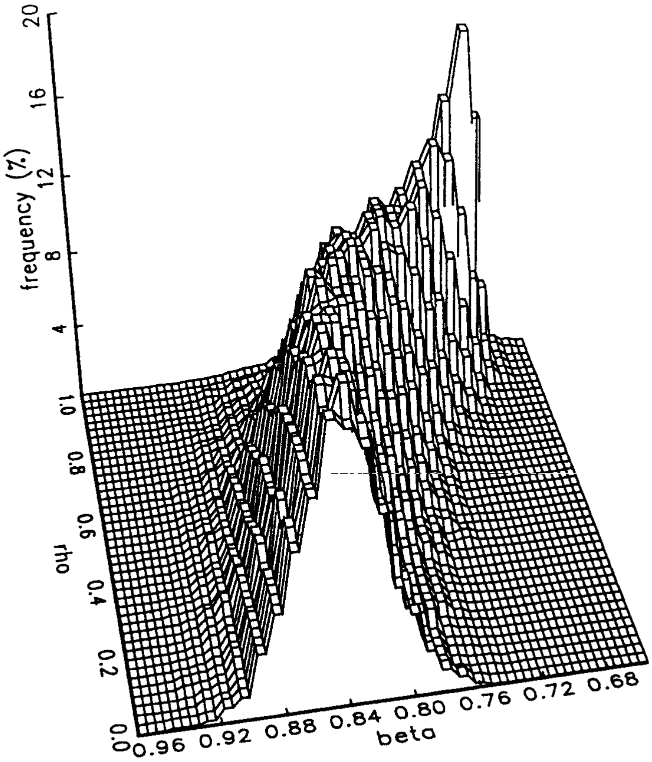
Note: See Figure 9.

Figure 11: The simultaneous equation bias as a function of the true value



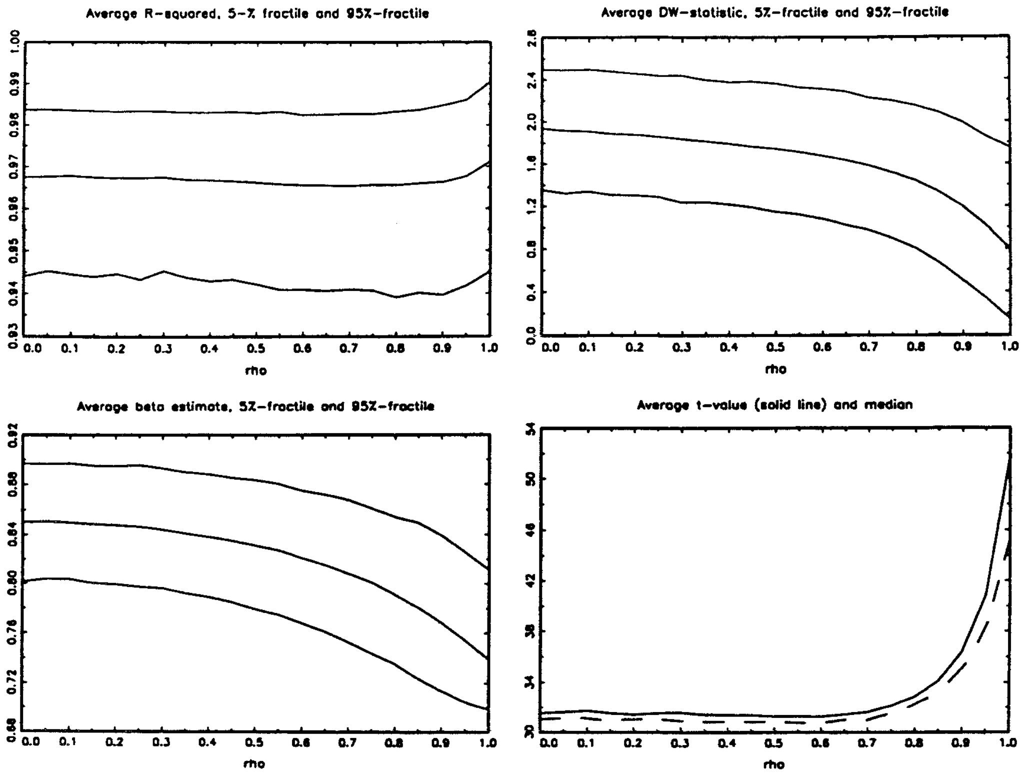
Note: Figure 11 captures 5000 replications of model (i)-(iii), varying the theoretical value β. Considering T=50 and zero drift, the graphic displays the three-dimensional histogram (20 categories) of the bias ($\hat{\beta} - \beta$) depending on $\beta = 0.15, 0.35, 0.55, 0.75, 0.95$. 275 outliers outside the range [-0.02,0.25] are skipped; extreme values: -0.043, 0.485.

Figure 12: Serial correlation, the adjustment towards cointegration, and the (biased) distribution of the OLS estimator in the presence of a simultaneous equation error



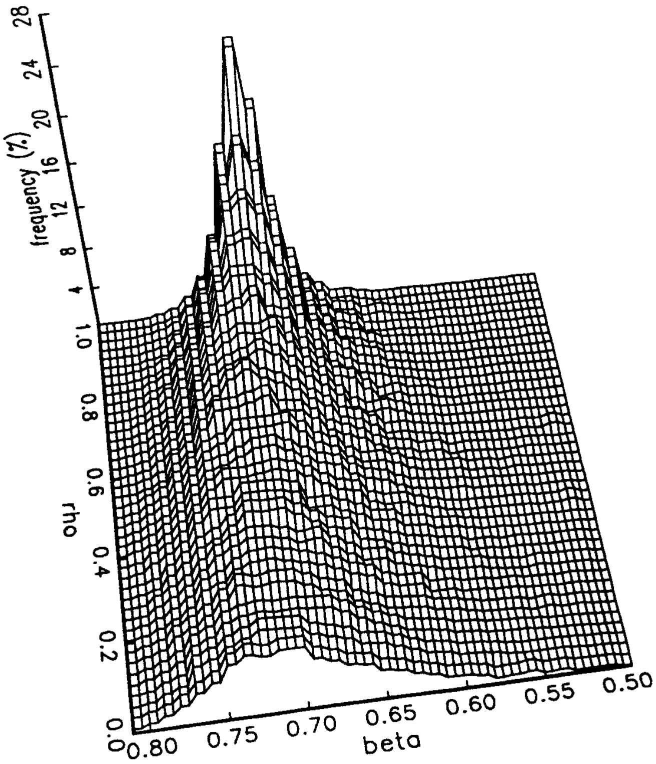
Note: Number of observations: T=30; model (i)-(iii) contains the exogenous process $X_t = \rho X_{t-1} + v_t$, ρ= 0, 0.05, 0.1,..., 0.90, 0.95, 1.00; for each ρ 5000 replications of the model (i)-(iii) are simulated; the graphic displays the distribution (30 categories) of $\hat{\beta}$ depending on ρ. 5 outliers outside the range [0.66,0.97] are skipped; extreme values: 0.65, 0.98.

Figure 13: The adjustment towards cointegration: Summary graphics



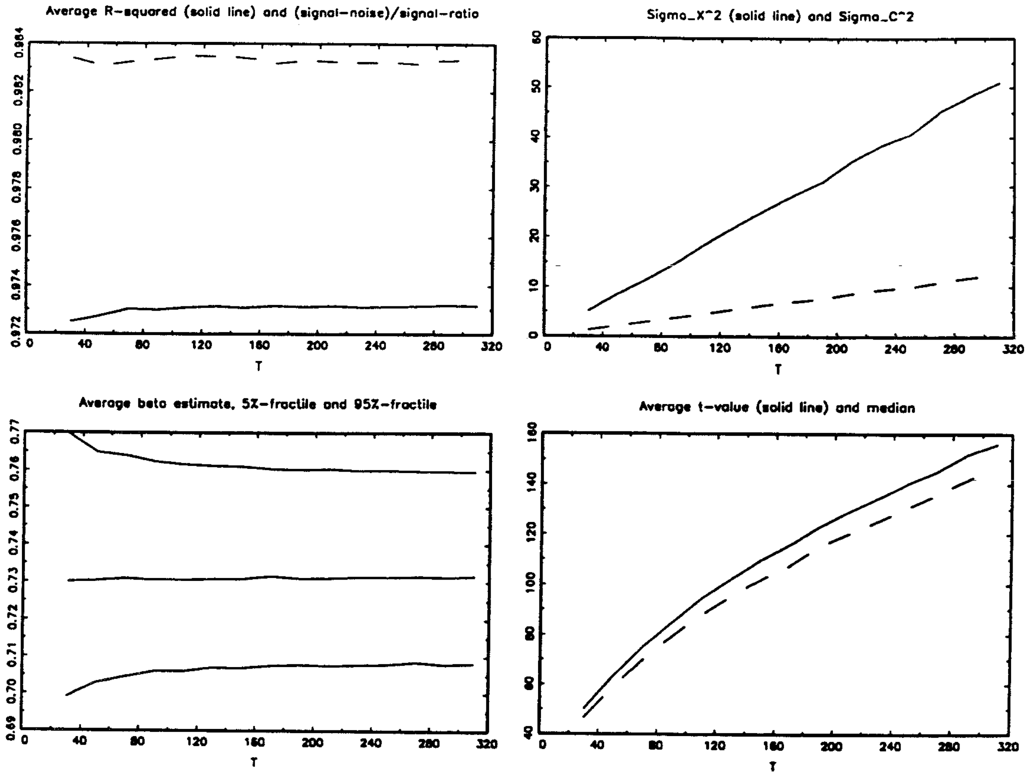
Note: See Figure 12.

Figure 14: The adjustment towards cointegration: Two-Stage-Least-Squares



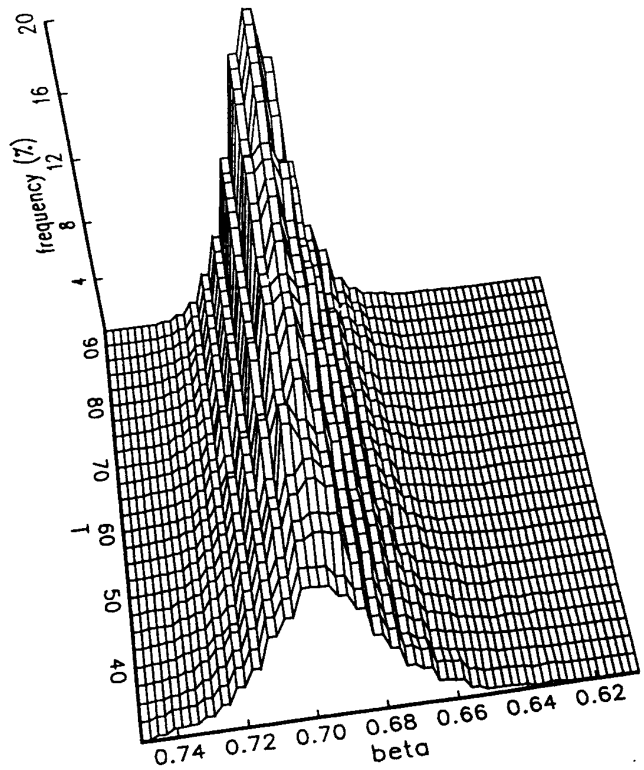
Note: See Figure 12 with TSLS instead of OLS; 1139 outliers outside the range [0.50,0.81] are skipped; extreme values: -1.62,0.84.

Figure 15: Constant signal-noise ratios and OLS estimation



Note: Figure 15 captures 5000 replications of model (i)-(iii) for each sample size T, T=30, 50, 70, ...,310; the exogenous process is a random walk without drift; the graphic displays summary graphics when noise-signal ratios are held constant (see the text for details); Sigma_X^2 and Sigma_C^2 denote the average estimated variance of the exogenous time series $X_1, \ldots, X_T$, and the imposed variance of the error term in the cointegrating equation (ii), respectively.

Figure 16: Constant signal-noise ratios and TSLS estimation



Note: Figure 16 captures 5000 replications of model (i)-(iii) for each sample size T, T=20, 25, 30, ...,90; the exogenous process is a random walk without drift; the graphic displays the time-dependent histogram (30 categories) of the TSLS estimate of β when noise-signal ratios are held constant (see the text for details). 39 outliers outside the range [0.61,0.75] are skipped; extreme values: 0.57, 0.78.

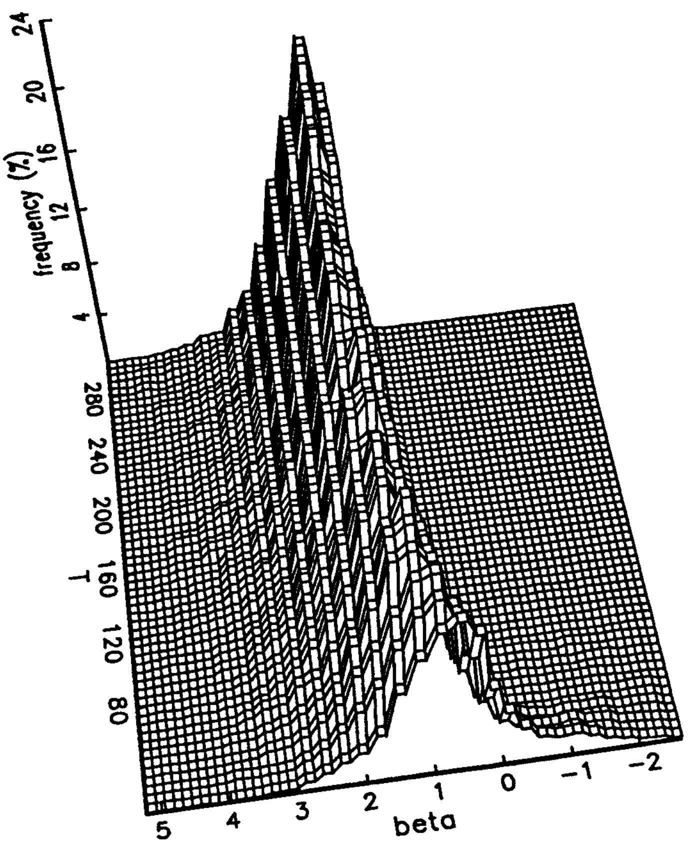
Supplement to "Random walks with drift, simultaneous equation errors, and small samples"

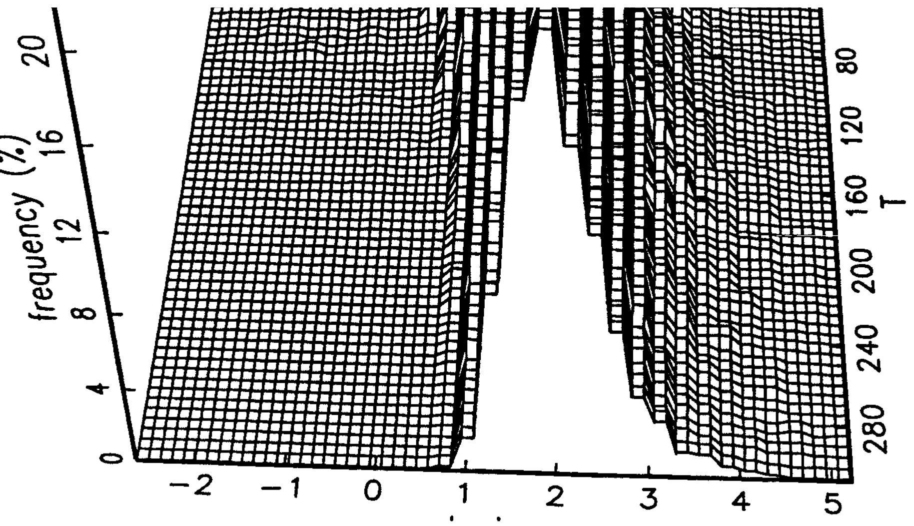
The supplement provides supporting evidence on small sample properties of regressing random walks with drift and simultaneous cointegrated variables. The first part of the supplement corresponds to results presented in the original paper, the second part analyses the regression of two independent random walks when an additional trend is included (see footnote 4 of the paper).

Part I.

Cf. "*Figure 3*: Regressing random walks with drift: Summary graphics": Three-dimensional counterparts.

a)

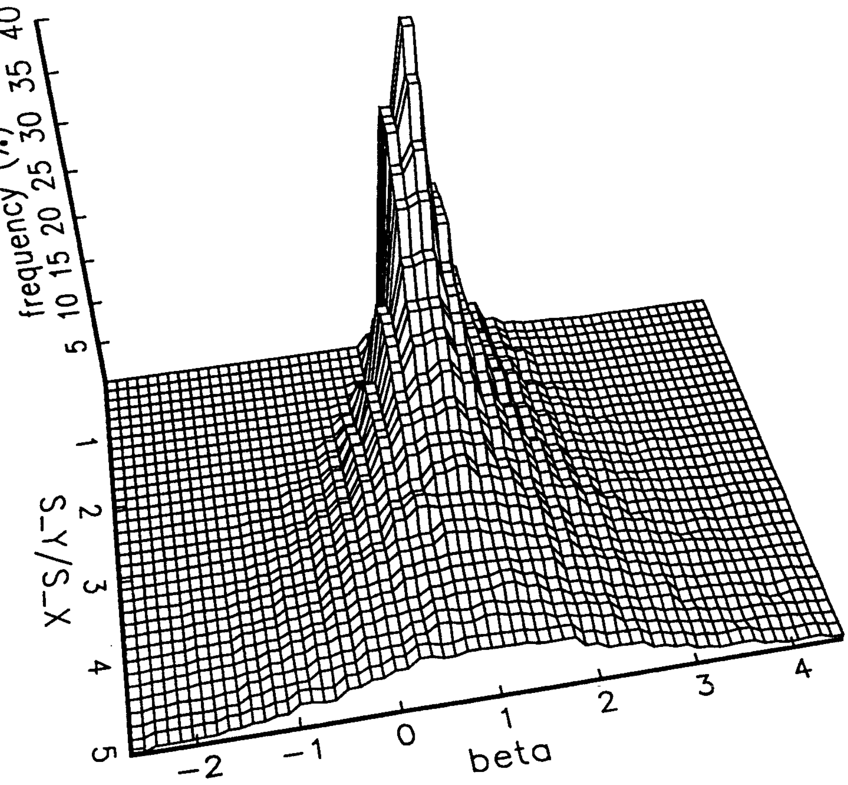




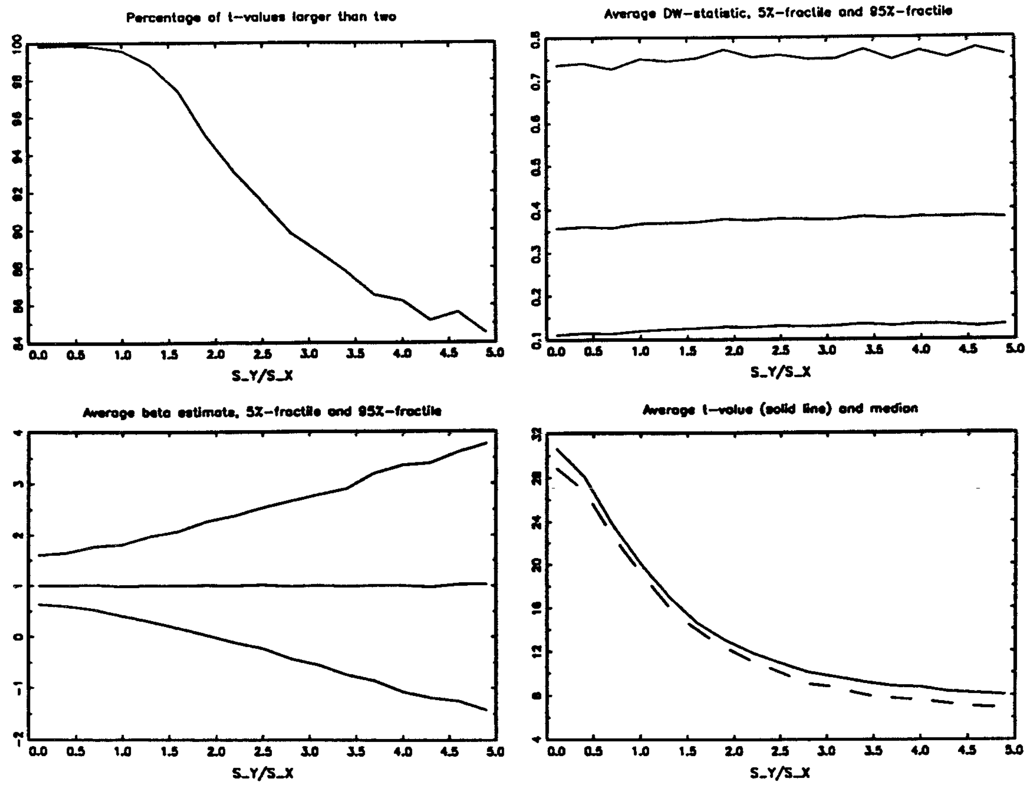
Note: Regression of two random walks as described in equations (1) and (2) of the paper; drifts are non-zero: $\gamma_y = 0.5, \gamma_x = 0.25$; the length of the simulated time series, T, is extended as T=20,30,...,300; 5000 replications of (1) and (2) are calculated for each sample size T; the graphics display the T-variant estimate of the slope parameter β (30 categories); 830 outlier outside the range [-2.5, 5.2] are skipped; extreme values: -6.39, 9.53.

Cf. *"Figure 5:* The impact of σ_y/σ_x on nonsense regressions": The case of random walks with drift.

a) 3d-representation



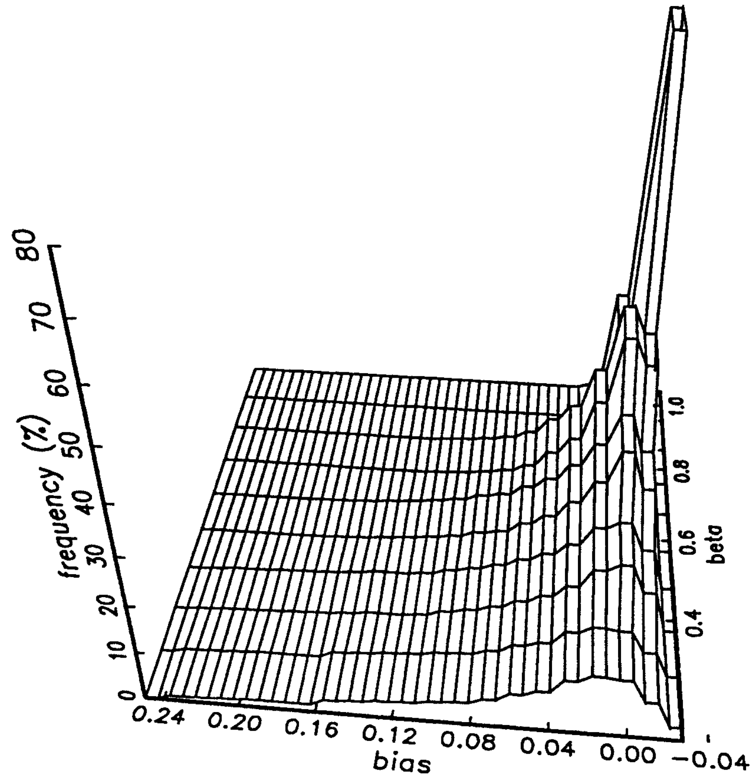
b) Summary graphics



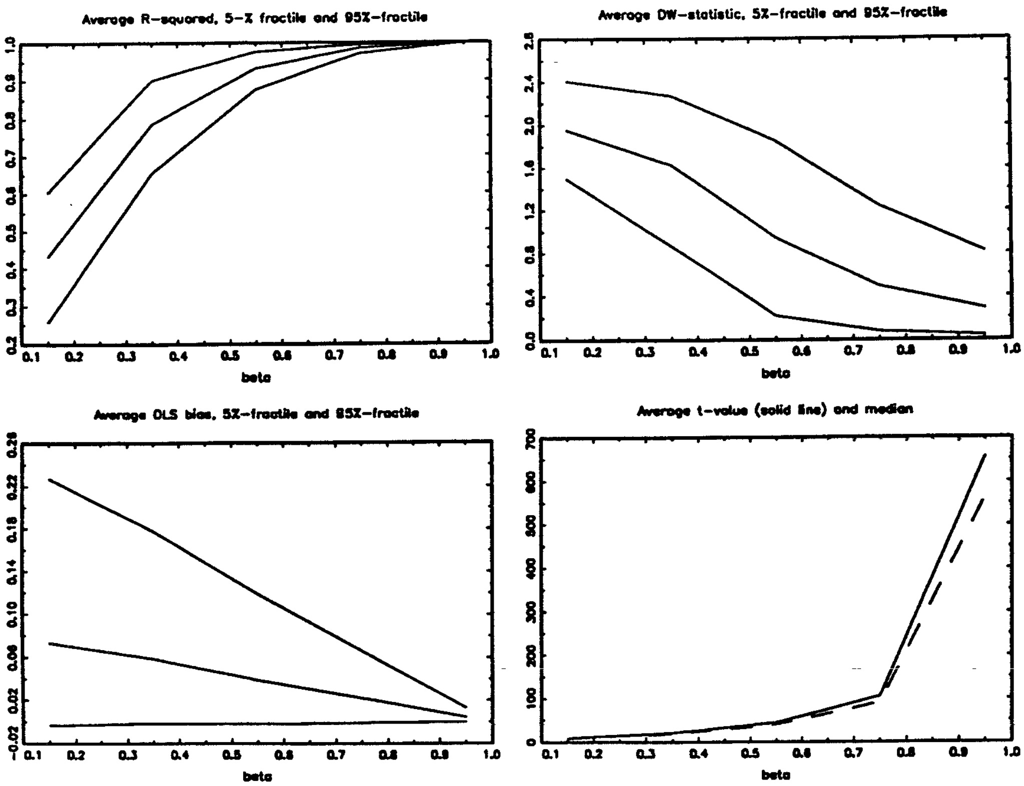
Note: Regression of two random walks as described in (1) and (2) with drifts $\gamma_y = \gamma_x = 1$; T=50; the ratio of the standard errors σ_y/σ_x is extended as 0.1, 0.4, 0.7, ..., 4.9; $\sigma_x = 1$; for each ratio the model is replicated 5000 times. The graphics display the ratio-dependent distribution of $\hat{\beta}$ (30 categories) and corresponding summary graphics; for the 3d-picture 748 outliers outside the range [-2.5, 4.5] are skipped; extreme values: -8.95, 14.36.

Cf. "*Figure 11:* The simultaneous equation bias as a function of the true value": A different view and summary graphics.

a) 3d-representation

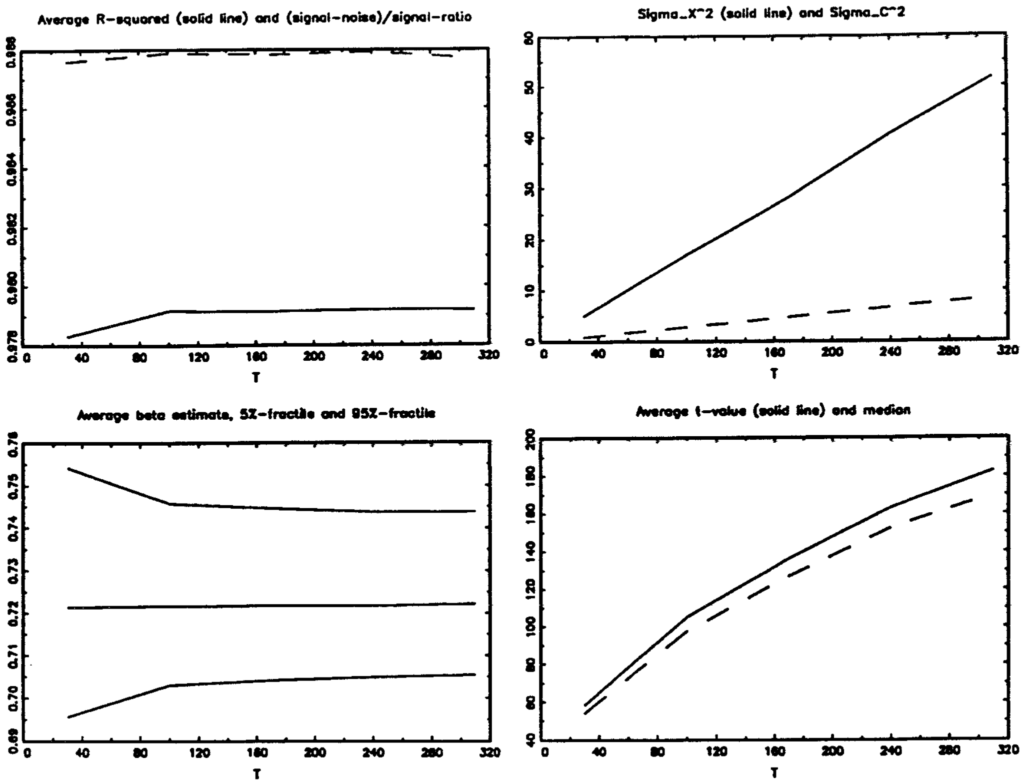


b) Summary graphics



Note: For each theoretical value β, the Figures capture 5000 replications of model (i)-(iii) presented in the paper. Considering T=50 and zero drift, the 3d-graphic displays the three-dimensional histogram (20 categories) of the bias ($\hat{\beta} - \beta$) depending on $\beta = 0.15$, 0.35, 0.55, 0.75, 0.95. 275 outliers outside the range [-0.02,0.25] are skipped; extreme values: -0.043, 0.485.

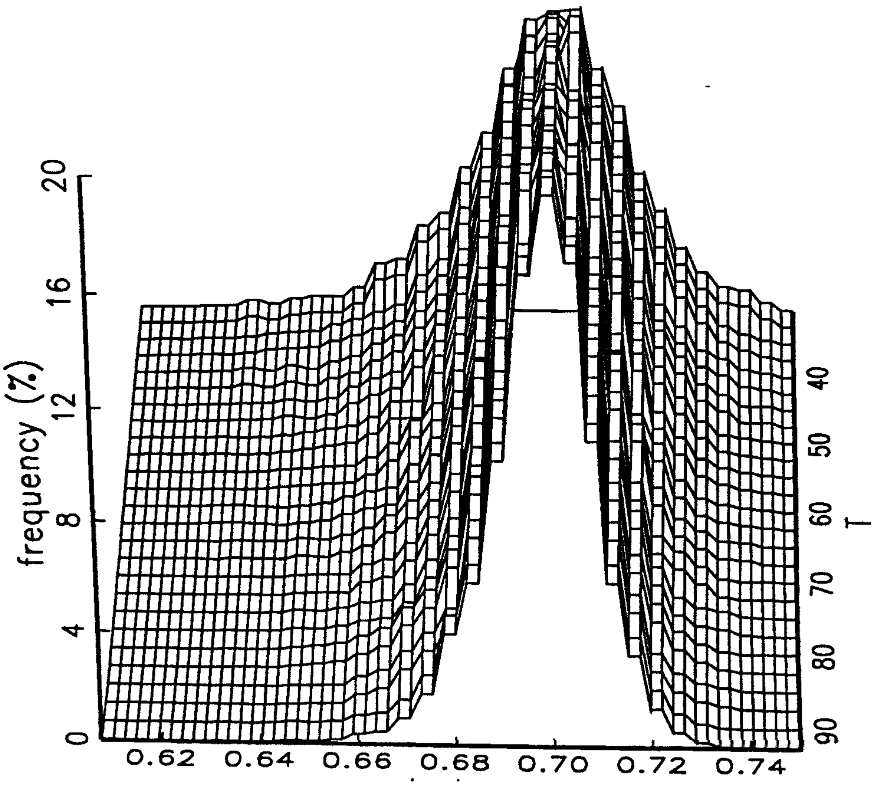
Cf. "*Figure 15:* Constant signal-noise ratios and OLS estimation": Imposing a different signal-noise ratio.



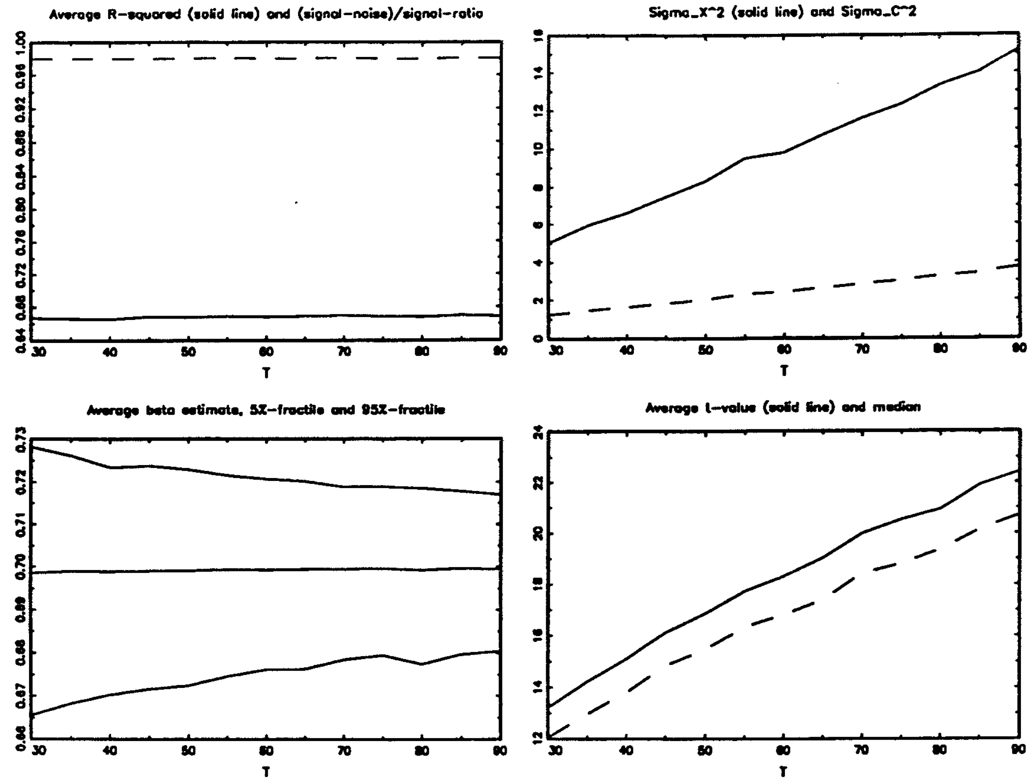
Note: The Figure captures 5000 replications of model (i)-(iii) for each sample size T, T= 30, 100, 170, 240, 310; the exogenous process is a random walk without drift; the graphic displays the summary graphics when signal-noise ratios are held constant using r=3 (see the paper for details); Sigma_X^2 and Sigma_C^2 denote the average estimated variance of the exogenous time series X_t and the imposed variance of the error term in the cointegrating equation (ii), respectively.

Cf. "*Figure 16:* Constant signal-noise ratios and TSLS estimation": A different view and summary graphics.

a) 3d-representaion



b) Summary graphics



Note: The Figures capture 5000 replications of model (i)-(iii) for each sample size T, T=20, 25, 30, ...,90; the exogenous process is a random walk without drift; the graphic displays the time-dependent histogram (30 categories) of the TSLS estimate of β when noise-signal ratios are held constant (see the text for details). 39 outliers outside the range [0.61,0.75] are skipped in the 3d-representation; extreme values: 0.57, 0.78.

Part II.

We consider the least squares regression

$$y_t = \hat{\alpha} + \hat{\beta} x_t + \hat{\delta} t + \hat{u}_t, t = 1, \ldots, T.$$

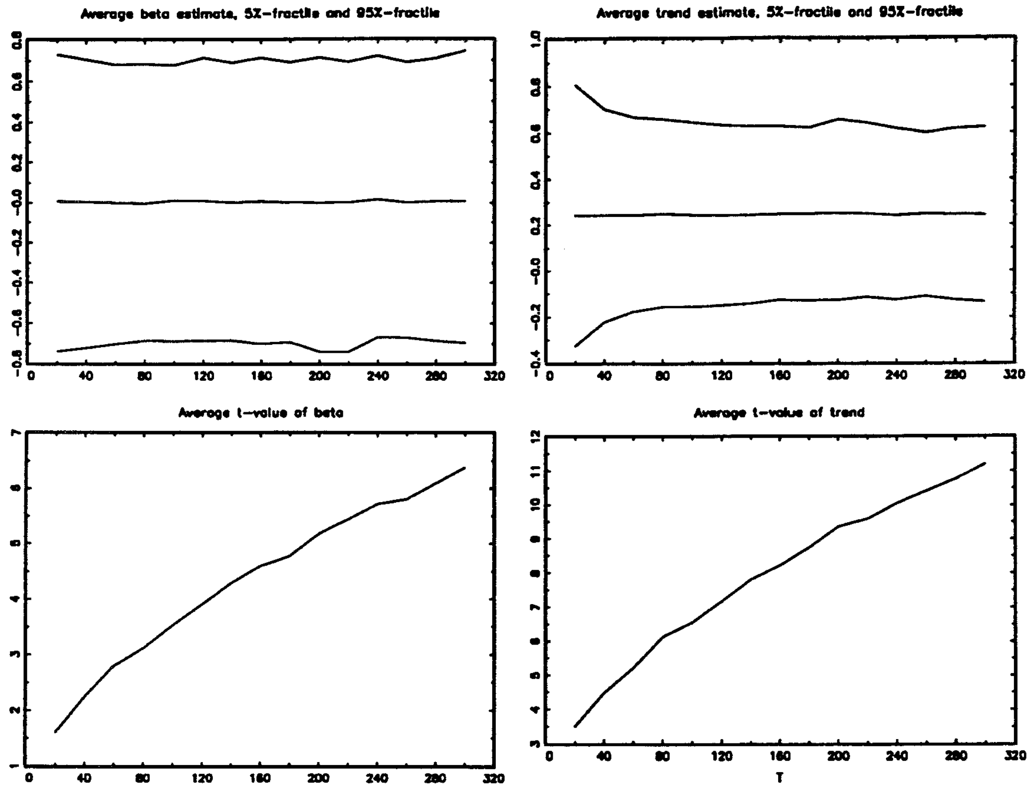
x_t and y_t are independent random walks, i.e. we define

$$y_t = \gamma_y + y_{t-1} + \varepsilon_t, x_t = \gamma_x + x_{t-1} + \nu_t, t = 1, 2, \ldots, T.$$

The error terms have zero mean and variances $\sigma_y^2$and σ_x^2.

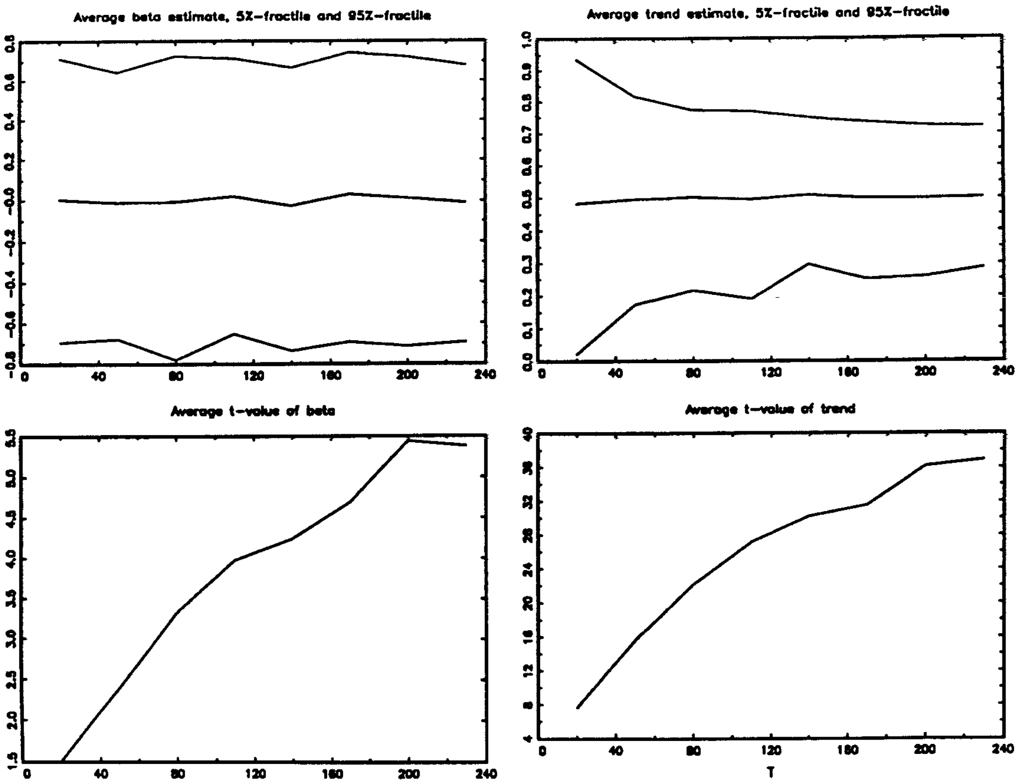
In the following simulation study we impose normal distribution of error terms. Starting values are $y_0 = 0$ and $x_0 = 0$. The simulation experiments (see Figures S1 - S4) vary the length of time series T, the magnitude of drifts, the variance of error terms, and the number of replications, N.

Figure S1:



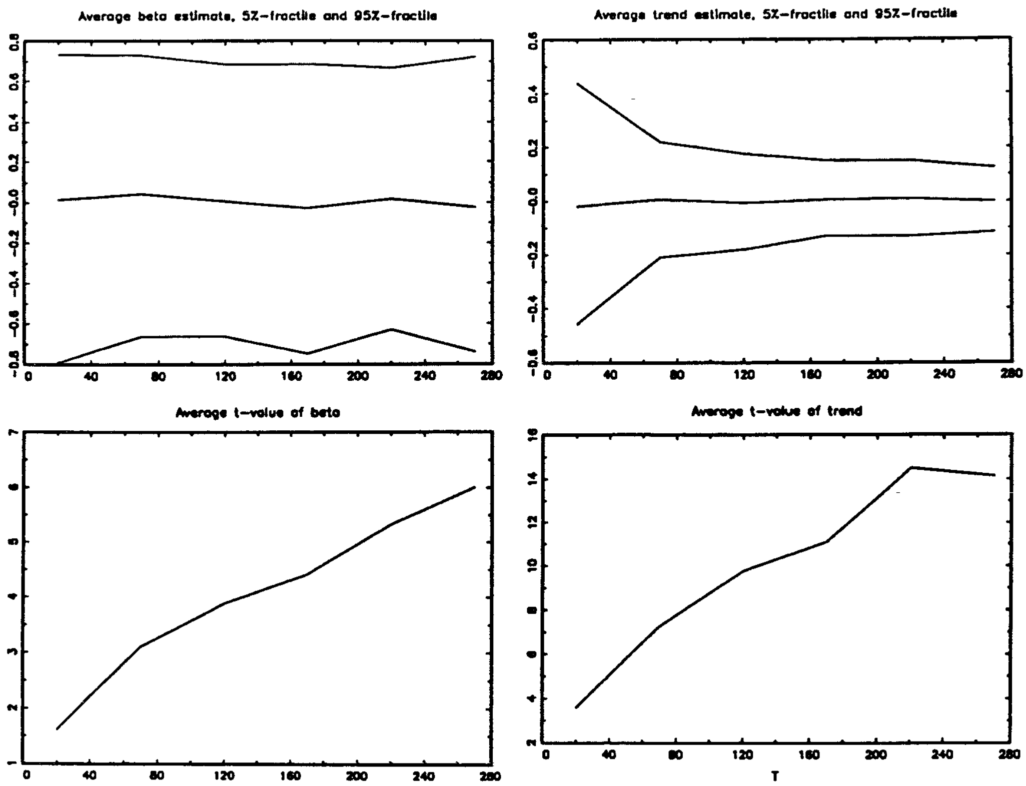
Note: $\sigma_y = \sigma_x = 1, \gamma_y = 0.25, \gamma_x = 0.5$, T= 20, 40, 60, ..., 300; 3000 replications for each T.

Figure S2:



Note: $\sigma_y = \sigma_x = 1$, $\gamma_y = 0.5$, $\gamma_x = 0.25$, T= 20, 50, 80, ..., 230; 800 replications for each T.

Figure S3:

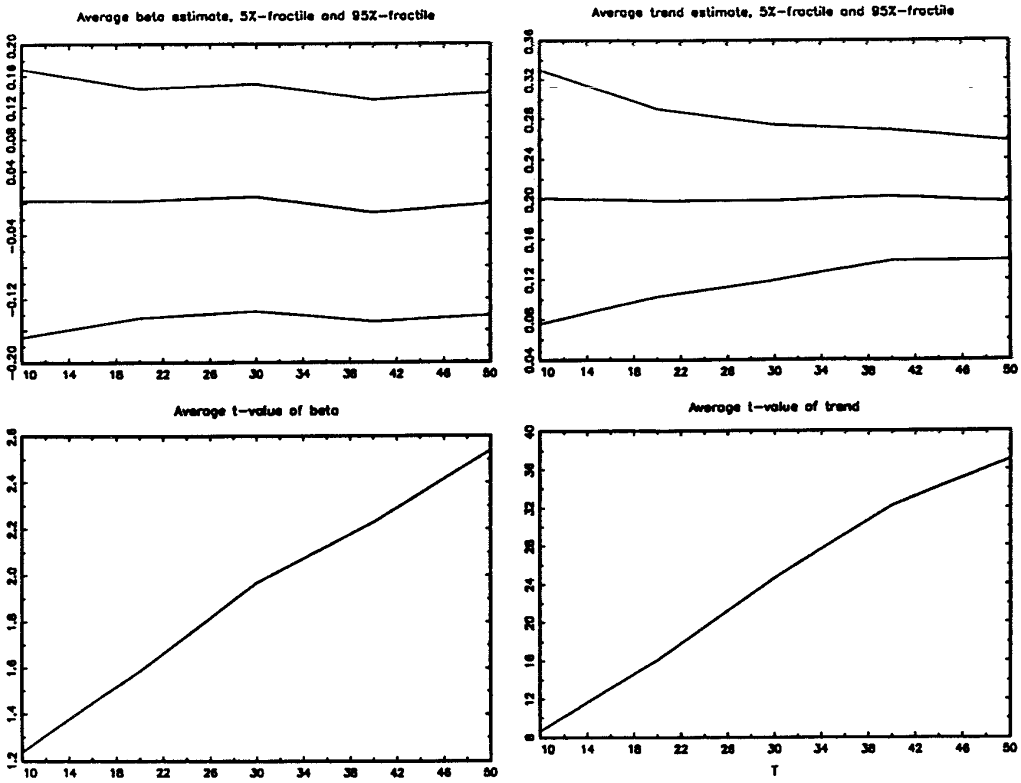


Note: $\sigma_y = \sigma_x = 1, \gamma_y = 0, \gamma_x = 0$, T= 20, 70, 120, ..., 270; 500 replications for each T.

The simulations S1 to S3 reveal convergence of $\hat{\beta}$ to a random variable. The average mean of the random variable fluctuates around zero; the empirical distribution turns out to be symmetric with 95%-fractiles (5%-fractiles) of about 0.7 (-0.7). The distribution of $\hat{\beta}$ is not affected by imposed drifts, which, however, determine the OLS estimation of the trend variable: The parameter of the linear trend converges to γ_y.

As for spurious regressions, we observe diverging (absolute) t-values - irrespective of the nature of zero or non-zero drifts. The estimated trend too has significant t-values, leading thus to "nonsense trends" in case of zero drifts.

Figure S4:



Note: $\sigma_y = 0.2$, $\sigma_x = 1$, $\gamma_y = 0.2$, $\gamma_x = 0.2$, T= 10, 15, 20, ..., 50; N=1000 for all T.

Figure S4 shows the influence of the ratio of imposed standard errors on the distribution of $\hat{\beta}$. The average of the trend is close to γ_y even for small T.