ON THE APPLICABILITY OF ECONOMETRIC METHODS TO SYSTEM DYNAMICS MODELS

M. Sommer, Department of Education, University of Bielefeld, Federal Republic of Germany

1. BACKGROUND

It is well known that System Dynamics (SD), from its very beginning, has been reluctant to employ statistical and econometric tools for estimation and evaluation of model parameters and equations. What has changed up to now is mainly a shift in the justification for this refusal. For more than the first decade of SD, Forrester's philosophy of selecting reasonable parameter values and of judging a model's validity has expressed the dominant attitude[4]. Because of the insensitivity of model behavior to most parameter values[1] and the direct observability of parameters from the real system[2] statistical methods were held to be superfluous. GRAHAM describes an arsenal of SD methods for parameter formulation and estimation. We have to sustain here from repeating our detailed discussion of this first period of SD rejection of econometric methods[28] pp. 244-291, 298-299 and[29].

With respect to parameter observability there is a fundamental dissent between SD and Econometrics (EC): on the scale, which ranks data and a-priori-information used in the process of parameter determination, we find as extreme cases SD models at one end, which use only a-priori-information, and EC models at the other end relying mainly on data information[24]. But there seems to be a contradiction in the SD argumentation against formal parameter estimation: while claiming that parameters do not have to be estimated because they are observable, variables are held to be frequently unobservable which in turn prohibits statistical parameter estimation. At least for socio-economic models on a macro level, the official statistics provide a vast amount of time-series and cross-sectional data on variables but almost none on parameters. This is implicitly acknowledged by LEHMANN, whose SD-model of the Federal Republic of Germany partially draws on time-series data for variables but not for parameters.

Probably as a by-product of the extensive work on a SD National Model of the USA, the econometric challenge has encouraged the reoccupation with the statistical estimation and testing of parameters and specifications. A remarkable shift within FORRESTER's justification to abolish econometric techniques from SD modeling has resulted. While in "Industrial Dynamics" statistical tests were assumed to be acceptable though mostly useless and only exceptionally necessary they are now judged as even dangerous because their internal validity criteria are supposed to provide wrong inferences[6,7].

It is the central purpose of this article to reinvestigate these experiments and the negative conclusions drawn by FORRESTER and his colleagues.

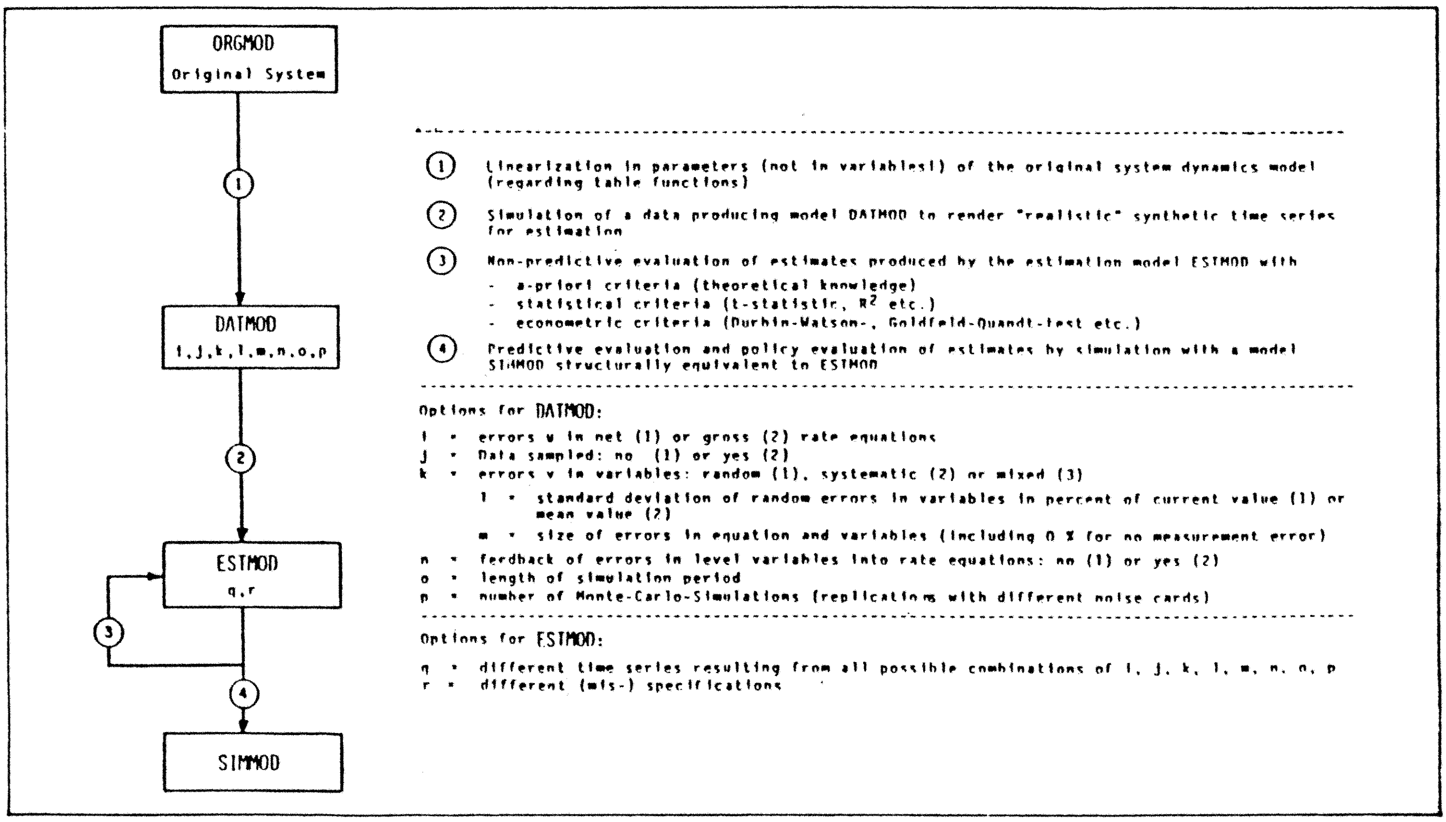


Figure 1: Flow Diagram of the Experimental Design.

2. EXPERIMENTAL REINVESTIGATIONS INTO THE ACCURACY AND RELIABILITY OF ECONOMETRIC METHODS

The contributions to this topic mentioned above[17,19,25,26]. all used a linear-in-the-parameters version of FORRESTER's Market Growth Model to produce synthetic data instead of real world data as input for parameter estimation. These data can be corrupted by errors in variables to take account of sampling and measurement errors. Furthermore, it is possible to study the effects of misspecifications of structural equations and econometric hypotheses about the residuals on the parameter estimates and test statistics. The general methodology of such experiments is well known from Mont Carlo studies which have been performed in EC to asses the small sample properties of various estimators when certain "classical" assumptions of the linear regression model are violated[8,11,13,27].

Fig. 1 shows a sketch of the complex experimental design and gives a rough idea of the numerous factors involved and the almost indefinite combinations to be analysed. The reader should especially be aware of the fact that far more than just one possibility does exist to generate the synthetic data from one SD model. It is, of course, completely impossible to investigate all combinations of factors influencing the data production process. For the non-predictive evaluation of estimation in this section, we have chosen the following conditions because they are similar to those enjoyed by SENGE, MASS and MORECROFT (although there are differences within and between these studies). The DATMOD, taken as the reference model for estimation in section 2, is simulated by pure random errors in all of the six net rate equations (noise) with standard deviations of 1% of the mean rate values and by pure random (measurement) errors in all level variables with zero mean and standard deviations of 10% of their current values without any feedback into the rate equations over one hundred periods.

The reason for corrupting the data by measurement errors in variables is that all experiments by SENGE et. al. without these errors have proven the adequacy of least-squares techniques for SD models. What is at stake is "only" the question, whether this also holds for measurement error conditions. That has rigorously been denied by the MIT authors.

PETERSON has already demonstrated that all the pitfalls of SENGE's experiments might be circumvented by a sophisticated estimation technique like Full Information Maximum Likelihood via Optimal (Kalman) Filtering (FIMLOF). But this proof does not invalidate FORRESTER's argument that the quantitative social and economic sciences heavily rely on regression analysis. It is the aim of this paper to answer the question, whether – contradictory to SENGE's conclusions – econometric techniques based on the least-squares principle being less complicated, less expensive and by far more widespread in common software packages than

Parameters test statistics	SENGE (1974)	SENGE (1977)2)	MASS/ SENGE (1978)	MORECROFT (1977)1)		TRUE VALUE
				a	b	
KØ1 |T-STAT|	-109,8 (0,4)	-	-	-	-	400
KØ2 |T-STAT|	425,6 (1,5)	-	-	-	-	-11,24
KØ3 |T-STAT|	-137,4 (1,6)	-	-	-	-	-11,80
KØ4 |T-STAT|	12,5 (1,5)	-	-	-	-	0,912
K1 |T-STAT|	-	252,8	408,2 (4,5)	120 (5,8)	2.427 (0,1)	475
K2 |T-STAT|	-	-31,6	-26,85 (1,2)	-38,4 (17,5)	-136,2 (0,2)	-61,5
K3 |T-STAT|	-0,524 (3,4)	-0,2942	-0,5563 (3,6)	0,038 (0,8)	-14,32 (0,1)	-0,6178
K4 |T-STAT|	0,107 (3,1)	0,0660	0,07719 (1,4)	-0,005 (0,3)	0,71 (0,1)	0,1324
K5 |T-STAT|	-0,0081 (2,7)	-0,00586	-0,00411 (0,68)	0,001 (0,6)	-0,04 (0,1)	-0,00975
Estimation method	GLS (2)	GLS (2)	OLS	OLS	IV	
R^2	0,38 (0,99)	? (0,99)	0,849	0,81	>-1	
DW	2,04	2,12		1,19	2,04	
test on homosced.	n.r.	n.r.	n.r.	n.r.	n.r.	

1) 5 % measurement noise

2) All parameters reported to be significant at 90 p.c. confidence level (two-sided test) → t >1,28

Table 1: Estimation Results for the Net Rate Equation RBL.

FIMLOF, do not do the job almost as well. The following discussion will be organized along the specific equations of the Market Growth Model which represent typical SD modeling elements.

2.1 Gross Rate Equations of The Material Network

The backlog equation of the Market Growth Model has been the one most extensively studied by the authors.(*)

(1) $BL(t) = BL(t\text{-}1) + DT * RBL(T\text{-}1)$

(2) $RBL = OB - DR$

(3) $OB = S * SE$

(4) $SE = g_1\ (DDRM)$

(5) $DR = PC * PCF$

(6) $PCF = g_2\ (DDM)$

(7) $DDM = BL / PC$

The equations show that the net rate equation RBL (periodical change in order backlog) consists of two gross rate equations, orders booked OB and delivery rate DR, which incorporate different policies. OB depends on the numbers of salesmen S and – in a linearized way – on the delivery delay recognized by market DDRM, while DR is governed by the production capacity PC and in nonlinear (i.e. cubic) way by the delivery delay minimum DDM. All papers have chosen to estimate the net rate equation (2) after reduction of their independent variables to levels only.

(2a) $RBL = K1*S + K2*S*DDRM + K3*BL + K4*BL^2/PC + K5*BL^3 / PC^2$

Only SENGE[25] has specified g as a cubic instead of a linear function with parameters KØ1, ..., KØ4. Furthermore SENGE[25, 26], MASS/SENGE[17] and MORECROFT[19] have lagged the right-hand-side variables by one period.

It is very interesting to see what interpretion and conclusions have been suggested by the authors. SENGE[25] (p. 53) holds that an econometrician would accept the estimated RBL equation with its inaccurate parameters, because with the exception of the t-statistic for KØ1 (Tab. 1, col. 1) he has no warning of the inaccuracies. We would raise at least two objections: first, an equation with four out of seven parameters exhibiting a t-statistic smaller than the "rule-of-thumb"-value of 2(t=1,96 for α= 0,05) deserves some reservations and, second, there is no reason to trust the R^2 of the level equation (0,99) when the R^2 for the actually estimated rate equation is very poor (0,38). One reason could be the cubic functional form. The linear functional form used by SENGE in his later paper[26] (p. 182; Tab. 1, col. 2) indeed gives more accurate parameters for the OB component of RBL. But a reestimation with second order GLS shows that R^2 is still very poor (0,49) and thus prevents – contradictory to SENGE's view[26] (p. 183) – the acceptance of inaccurate estimates.

(*) *The abbreviations for the variables names are the same as in[5]. A preceeding "R" denotes a rate, "U" an error in equations, "V" an error in variables. Variables without time indices belong to the present period t.*

MASS and SENGE later estimated the same equation by OLS instead of GLS. The resulting t-values (Tab. 1, col. 3) are interpreted in a remarkably different way now. Because of t = −1,248 for K2 "an econometrician viewing at the statistical results might conclude that delivery delay (a variable associated with K2, M.S.) is a relatively unimportant influence on sales"[17] (p. 455). But why should K2 be questioned while K5 with a much worse t-statistic remains unrefuted? Again, we would argue as before: three out of five t-values are undesirably low and call for further thinking. MORECROFT's results demonstrates that a change to another estimator (Instrumental Variables) is not necessarily a good choice.

The order backlog part of the model contains two different policies, one for the inflow, the other for the outflow. Inflow and outflow rate equations – if structurally different – are the SD counterparts of behavioral or stochastic equations in EC and thus have to be estimated seperately (T–STAT/ beneath parameters).

(3a) $OB = \underset{(33,7)}{429,8} * S - \underset{(14,6)}{49,2} * S * DDRM$

DW = 1,59 R^2 = 0,94 Goldfeld-Quandt-test: F = 0,13

(5a) $-DR = \underset{(20,2)}{-0,598} * BL + \underset{(7,1)}{0,1161} * BL^2/PC - \underset{(3,4)}{0,0073} * BL^3/PC^2$

DW = 1,99 R^2 = 0,95 Goldfeld-Quandt-test: F = 0,09

Keeping in mind that we are dealing with strong measurement errors, these estimates can well be judged as very statisfactory. The apparent heteroscedasticity we obtain even better estimates for the OB equation.

(3b) $OB = \underset{(39,6)}{463,4} * S - \underset{(18,9)}{54,6} * S * DDRM$

DW = 1,68 R^2 = 0,98 Goldfeld-Quandt-test; F = 1,57

2.2 Nonlinear Functional Relationships

FORRESTER's original Market Growth Model contains three table functions. In DATMOD one table (SE) has been completely linearized while the other two (PCF and CEF) are linearized in parameters only and are third order polynomials. We have shown in the previous section that it is no problem to estimate SE and PCF – implicitly embodied in OB and DR – in the same functional form that was used in DATMOD. This may not always work out, as can be illustrated by the rate equation for production capacity ordering PCO, which incorporate the third nonlinear function CEF.

(8) $PCO = PC * CEF$

(9) $CEF = g_3\ (DDC)$

The function used in DATMOD is of the form

(9a) $PCO = K12*PC + K13*PC*DDC + K14*PC*DDC^2 + K15*PC*DDC^3$

SENGE[25] (pp. 53) interprets his estimated parameters (Tab. 3, col 1) as moderately accurate but statistically insignificant, supporting his conviction that the application of econometric methods tends to lead into the two error categories B and C cf Tab. 2 (accepting wrong or rejecting right hypotheses) while econometricians hope to avoid these errors as often as possible.

t-statistic indicates	parameter is accurate	parameter is inaccurate
significance	A	B
insignificance	C	D

Table 2: The two pitfalls of econometrics when applied to a SD model (according to SENGE[25]).

We rather feel that with all t-values smaller than 2 and R^2 = 0,57 one is not really inclined to accept these estimates, especially when the true parameters are unknown. Furthermore, the experimental situation shows that the estimates are very sensitive to the random number generator (compare cols. 1 and 2 as well as cols. 3 and 4 in Tabl. 3), but can never be save-guarded against zero. In such a situation, it is wise to try other functional specifications in estimation than have been underlying the data production. We would further suggest not only to draw on inferential but also to apply descriptive statistics, i.e. to plot the relevant variables against each other (Fig. 2).

In spite of the good test statistics obtained by a strictly linear specification (Tab. 3, col. 5) it seems preferable to drop only the quadratic term out of the equation (Tab. 3, col. 6). The comparison shows that we should not strive

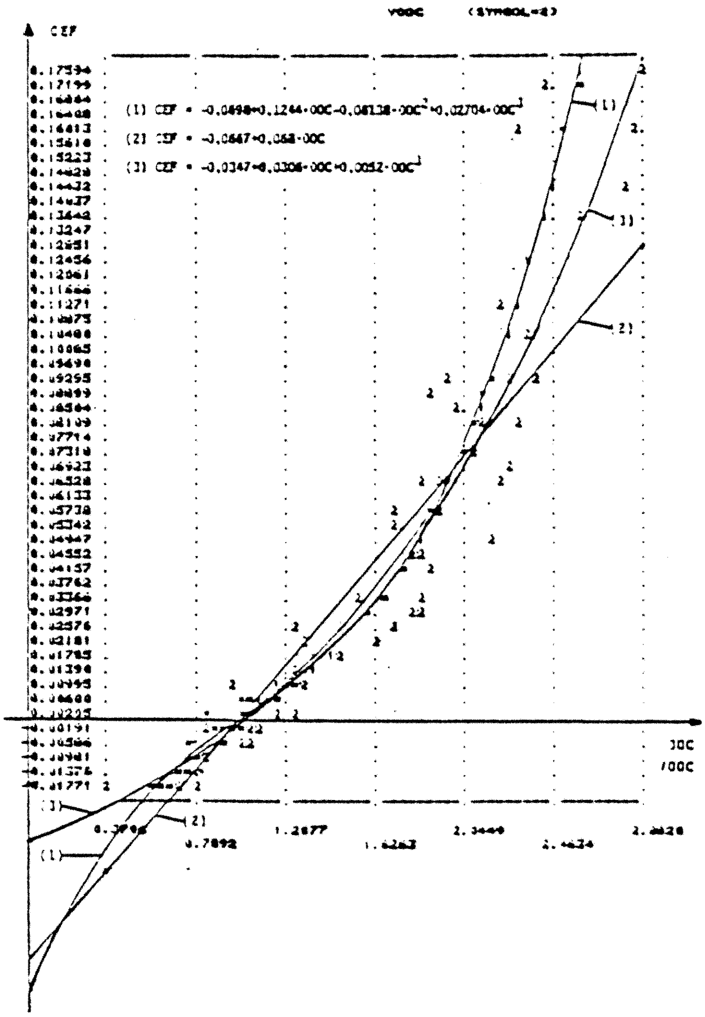


Figure 2: Different functional forms for CEF.

Parameters test statistics	SENGE (1974)	SENGE (1977)	SOMMER a	SOMMER b	SOMMER c	SOMMER d	TRUE VALUE (except for c, d)
K12 |T-STAT|	-0,0433 (1,6)	-0,0345	0,0064 (0,2)	-0,0170 (0,6)	-0,0667 (16,7)	-0,0347 (4,7)	-0,0698
K13 |T-STAT|	0,0693 (1,1)	0,0364	-0,0541 (0,7)	-0,0084 (0,1)	0,068 (26,1)	0,0306 (3,8)	0,1244
K14 |T-STAT|	-0,0423 (0,9)	-0,0133	0,0533 (1,0)	0,0258 (0,7)	-	-	-0,08138
K15 |T-STAT|	0,0174 (1,651)	0,01049	-0,0053 (0,5)	-0,0001 (0,02)	-	0,0052 (4,9)	0,02704
Estimation method	GLS (2)	GLS (2)	OLS	OLS	OLS	OLS	
R^2	0,57 (0,99)	? (0,99)	0,89	0,91	0,86	0,91	
DW	1,91	1,85	2,02	1,99	1,23	1,99	

Table 3: Estimation results for the weakly nonlinear function PCO.

for an isolated maximization of t-values, because this can push us towards linear specifications even where they are inadequate. The heavy measurement error no longer allows to estimate the nonlinear function in the same specification as it has entered DATMOD, but it is still possible to receive good nonlinear estimates if we carefully respecify ESTMOD.

2.3 A-Priori–Informations on Parameters

Another important rate equation of the Market Growth Model's material network is the salesmen hiring equation SH. Since salesmen hiring (inflow) and firing (outflow) are assumed to be regulated by the same decision rule, SH truely is a net rate equation.

(10) S(t) = S(t-1) + DT * SH(t-1)

(11) SH = (1/SAT) * (IS - S)

(12) IS = B/SS

(13) B = RS * DRA

Substitution of auxiliary equations (12) and (13) into (11) leads to:

(11a) SH(t) = (RS/SS)*(1/SAT)*DRA(t) - (1/SAT)*S(t)

The estimation equation used by SENGE and MORECROFT is specified as:

(11b) H = H = K6*DRA + K7*S

SENGE[25] obtained significant parameter estimates by GLS which might be regarded as too inaccurate (Tab. 4, col. 1). The same might be said of MORECROFT's OLS-estimates. MORECROFT further proves that IV are capable of coming up with more exact parameters in this case. It is doubtful though, that t-statistics of size 2 or 3 together with an R^2 = 0,20 would lead to an acceptance of the estimates if one does not know the true parameters, which is the normal case in nonexperimental situations.

But even when K6 and K7 are unknown this equation is a good example for a case in which partial a-priori information on the parameters is available and asks for consideration. Of the three parameters in eq. (11a), only the salesmen adjustment time SAT is unobservable while revenue to sales RS as well as salesman salery SS are directly observable constants. There is no reason why this a-priori information should be left out of the estimation process. With RS = 12 and SS = 2000 we get:

(11c) SH = 0,006*(1/SAT)*DRA - (1/SAT) * S

We can therefore impose the restriction

(11d) K6 + 0,006*K7 = 0

On the OLS-estimation. The ROLS-parameters (restricted ordinary least squares) are very accurate and extremely tight (Tab. 4, col. 6). This example should encourage the use of a-priori-information, held to be so frequently available by system dynamicists, within the econometric estimation of SD models.

2.4 First Order Information Delays and Smoothing

The Market Growth Model contains two first order information delays DDRC and DDRM which have not been estimated in either of the MIT studies. The problems involved in estimating the time constants of first order delays under the measurement error conditions chosen by the authors can well be understood by using an analytical formula developed for the asymptotic case of unlimited sample size. The insights gained can then be used to uncover the reason

Parameters test statistics	SENGE (1974)	SENGE (1977)1)	MORECROFT (1977)2)		SOMMER		TRUE VALUE
			a	b	a	b	
K6 |T-STAT|	0,00065 (4,0)	0,00051	0,00014 (9,4)	0,00042 (3,1)	0,00010 (6,9)	0,00024 (38,0)	0,0003
K7 |T-STAT|	-0,12972 (3,4)	-0,098	-0,012 (3,4)	-0,077 (2,4)	-0,00236 (0,7)	-0,04027 (5145,7)	-0,05
Estimation method	GLS (2)	GLS (2)	OLS	IV	OLS	ROLS	
R^2	0,17 (0,99)	? (0,99)	0,83	0,20	0,75	0,40	
DW	2,01	2,04	0,82	2,10	0,31	1,22	

1) Parameters reported to be statistically significant at a 90 p.c. confidence level

2) STDV of measurement error 5 % p.c.

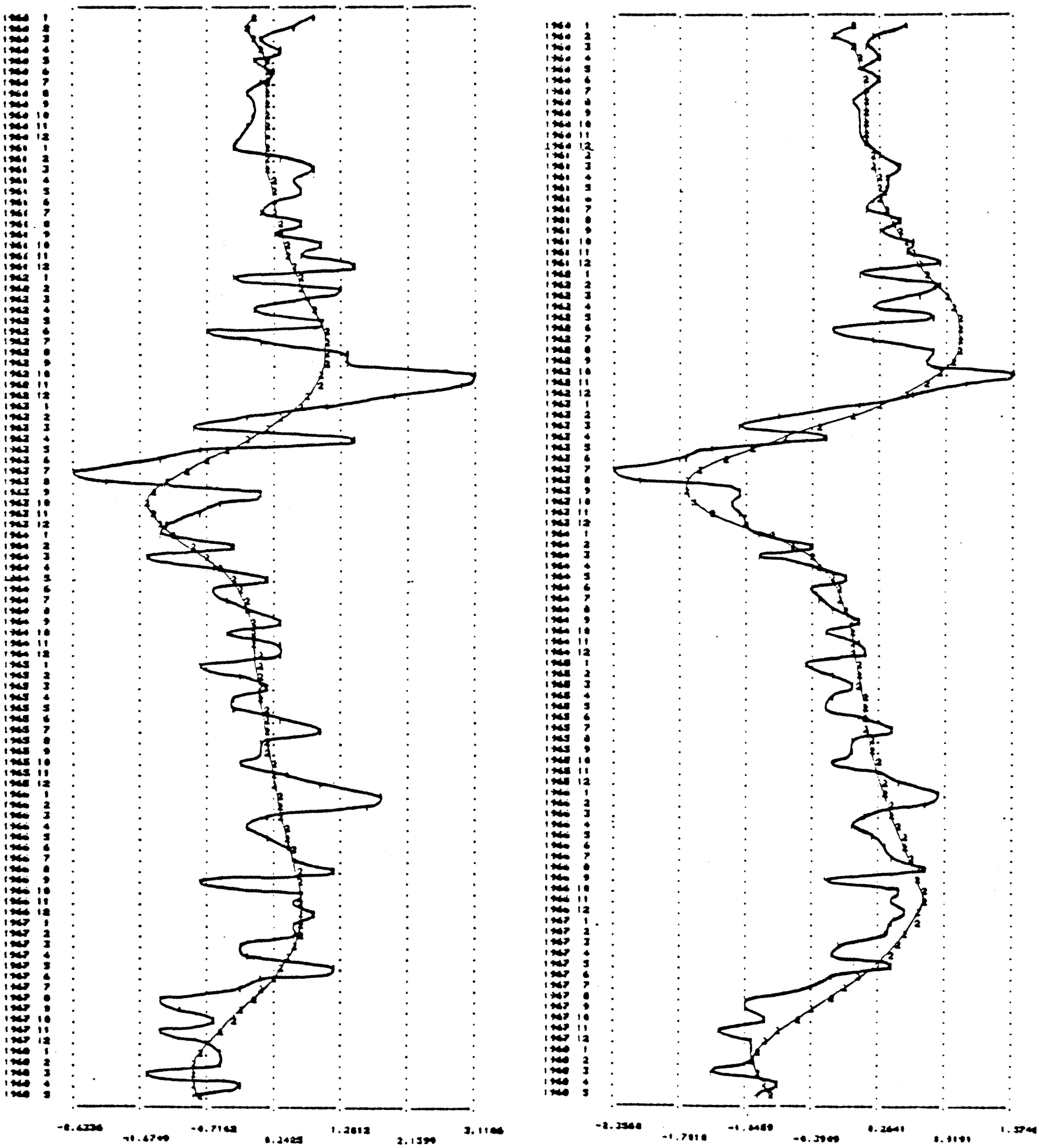
Table 4: Estimation results for SH.

for the poor estimates the authors received for the delivery rate average DRA which is formally equivalent to a first order information delay. We will limit the discussion to DDRC and not touch upon DDRM, because the first equation presents the worse estimates.

$$(14)\quad DDRC(t) = DDRC(t\text{-}1) + DT * RDDRC(t\text{-}1)$$

$$(15)\quad RDDRC = (1/TDDRC) * (DDI - DDRC)$$

$$(16)\quad DDI = BL/DRA$$



(a) pure random measurement error *(b) mixed random systematic error*

Figure 3: Time series of error-free and error-corrupted variable, (BL/DRA-DDRC).

DDI is a definitional equation of the theoretical delivery delay while DDRC is the delivery delay as recognized by the company. The parameter to be estimates is the time constant TDDRC of this recognition delay:

$$\text{(15a) RDDRC} = \underset{(8,8)}{0{,}107}*(\text{BL/DRA} - \text{DDRC})$$

DW=0,64 R^2 = 0,42 Goldfeld-Quandt-test: F=2,14

This implies an estimates value for TDDRC=9,3 months which is more than twice the true value of 4 months. Denote x = BL/DRA–DDRC for the right-hand-side variable of (15) and let v be the measurement error of x, then the probability limit for (1/TDDRC) in infinite samples is[15] (p.293):

$$\text{(17)}\quad \text{plim}\,(\widehat{1/\text{TDDRC}}) = \frac{(1/\text{TDDRC})}{1+\dfrac{\hat{C}^2(v)}{\hat{C}^2(x)}} = \frac{0{,}25}{1+\dfrac{0{,}475}{0{,}351}} = 0{,}1062$$

Thus, the probability limit for (1/TDDRC) is very close to the experimental finite sample size estimate (0,107). Formula (17) tells that the degree of underestimation depends on the ratio of the measurement error variance to the variance of the error free time series:

$$\text{(18)}\quad \hat{C}^2(v)/\hat{C}^2(x) = 0{,}457/0{,}351 = 1{,}353$$

$$\text{(19)}\quad \hat{C}^2(\text{VBL-BL})/\hat{C}^2(\text{BL}) = 31377417/1241055738 = 0{,}025$$

$$\text{(20)}\quad \hat{C}^2(\text{VDRA-DRA})/\hat{C}^2(\text{DRA}) = 2443571/59892751 = 0{,}041$$

$$\text{(21)}\quad \hat{C}^2(\text{VDDRC-DDRC})/\hat{C}^2(\text{DDRC}) = 0{,}157/1{,}398 = 0{,}112$$

$$\text{(22)}\quad \hat{C}^2(\text{VDDI-DDI})/\hat{C}^2(\text{DDI}) = 0{,}330/1{,}573 = 0{,}210$$

Of course, (18) is the ratio which leads to the severe distortion of the TDDRC estimate in (17). But neither the relative measurement error variances (19) to (21) nor the division of BL/DRA in (22) are the main source of the bias but rather a too small variance of the error free difference variable x = DDI-DDRC (0,351). It must be doubtful, though, that realistic economic time series contain measurement errors whose variances are of the same size as or even exceed the variances of the "true" time series, as it is the case in the experimental situation discussed here. Fig. 3a illustrates these conditions: observed time series would have to be highly erratic because random measurement errors are superimposed on "true" time series with much smoother dynamics. This seems to be a very unlikely situation. Nevertheless it might be useful to plot the data against time before estimation. If such erratic plots really occur, the question arises whether they reflect a true dynamic process with very short periodicity or are rather a result of random measurement errors.

SENGE and MASS have referred to MORGENSTERN's classical work "On the Accuracy of Economic Observation"[20] to justify a measurement error of 10% employed in their data production process. But neither MORGENSTERN's inquiry nor the voluminous modern literature on that topic (e.g.[1,2,3,23,31]) suggest that errors in economic statistics are purely random. Comparisons of preliminary and revised data as well as comparisons of data compiled by different institutions usually suggest a strong systematic error component. This is quite plausible since the methods of data compilation which produce these errors are not changed every period but used for some years until "better" methods have been developed. This is not the place for an in-depth discussion of the kind and magnitude of measurement errors in economic and social statistics, but even a cursory contemplation confirms our reluctance to accept the 10% random measurement error of Fig. 3a as realistic. In short, MADDALA can be approved when stating that "the errors occuring in economic data are systematic rather than random"[15] (p. 292).

Just for illustrating the relevance of the disput we have alternatively produced data where one part of the measurement errors in systematic (5% of the "true" variables mean values) and the other part is random (standard deviation of 5% of the variables current values). Fig. 3b displays the resulting time series for the right-hand-side variable BL/DRA–DDRC, which still contains a remarkable measurement error but of different composition. OLS estimation of (15) based on these data leads to:

$$\text{(15b) RDDRC} = \underset{(15,7)}{0{,}192} * (\text{BL/DRA–DDRC})$$

DW = 0,82 R^2 = 0,71 Goldfeld-Quant-test: F = 1,71

The substantial improvement of (15b) over (15a) by simply making half of the 10% measurement error systematic is beyond discussion. Since the results obtained for DDRC also hold for DDRM we can skip the latter equation. The last equation deserving attention explains the delivery rate average DRA, because the problems of estimating first order delays or smoothing processes are overlayed here by the neccesity to watch for parameter restrictions across equations.

$$\text{(23)}\quad \text{DRA(t)} = \text{DRA(t-1)} + \text{DT} * \text{RDRA(t-1)}$$

$$\text{(24)}\quad \text{RDRA} = (1/\text{DRAT}) * (\text{DR} - \text{DRA})$$

Since DR is explained by eqs. (5) to (7), eq. (24) has to be case into the following form for estimation:

$$\text{(25)}\quad \text{RDRA} = \text{K8*BL} + \text{K9*BL}^2/\text{PC} + \text{K10*BL}^3/\text{PC}^2 + \text{K11*DRA}$$

The results of our OLS estimation as well as the GLS estimates published in SENGE[26] are very poor (Tab. 5). GLS estimates reported earlier in SENGE[25] are much better but we did not find a way to reproduce them under the conditions described there. Using the a-priori knowledge K11 = (-1/DRAT), eq.(25) with SENGE's 1974 parameters becomes

$$\text{(25a) RDRA} = 0{,}3740*(0{,}6643*\text{BL} + 0{,}1484*\text{BL}^2/\text{PC} + 0{,}01144*\text{BL}^3/\text{PC}^2 - \text{DRA})$$

Thus the three parameters of the DR component are quite close to their true values and only the time constant (1/DRAT) is heavily biased implying a delivery rate avering time of 2,7 instead of one month. This is no surprise after the previous

Parameters, test statistics	SENGE (1974)	SENGE (1977)[1)]	SOMMER			TRUE VALUE
			a	b	b*	
K8 |T-STAT|	0,2485 (5,6)	0,0876	0,0124 (2,2)	0,2438 (48,4)	K8 * = -K3 = 0,6476	0,6178
K9 |T-STAT|	-0,0565 (5,2)	-0,0246	-0,0054 (3,3)	-0,0502 (28,8)	K9 * = -K4 = -0,1332	-0,1324
K10 |T-STAT|	0,0043 (4,5)	0,0023	0,0005 (3,1)	0,0035 (16,5)	K10 * = -K5 = 0,00916	0,00975
K11 |T-STAT|	-0,3740 (5,2)	-0,1024	0,0163 (1,8)	-0,3764 (52,4)	K11 * = -1	-1
Estimation method	GLS (2)	GLS (2)	OLS	R3SLS		
R^2	0,24 (0,99)	? (0,99)	0,41			
DW	2,15	2,13	0,32			
test on homosced.	n.r.	n.r.	0,55			

1) All parameters reported to be significant at 90 p.c. confidence level (two-sided test) → t > 1,28

Table 5: Estimation results for the smoothing equation DRA.

discussion of the problems with the estimation of the RDDRC equation. Divering from SENGE's estimation procedure the following a-priori parameter restrictions should be taken into account:

(26) K8 = K3

(27) K9 = K4

(28) K10 = K5

These parameter restirctions across equations can be captured by R3SLS (Restricted Three Stage Least Squares) estimation of a submodel existing of the OB, DR, BL, RDRA, and DRA equations. After again taking account for:

(29) K11 = –(1/DRAT)

we obtain the following equations

(5b) $DR = -0{,}6476*BL + 0{,}1332*BL^2/PC - 0{,}00916*BL^3/PC^2$

(25b) $RDRA = 0{,}3746*(0{,}6476*BL - 0{,}1332*BL^2/PC + 0{,}00916*BL^3/PC^2 - DRA)$

The consistency between DR and RDRA is now guaranteed. What remains is the unsatisfactory estimation of DRAT. With the same mixed systematic and random measurement error already used for RDDRC this problem vanishes too:

(5c) $DR = 0{,}5620*BL + 0{,}1030*BL^2/PC - 0{,}00619*BL^3/PC^2$

(25c) $RDRA = 0{,}8889*(0{,}5620*BL - 0{,}1030*BL^2/PC + 0{,}00619*BL^3/PC^2 - DRA)$

The slightly worse parameters in the DR-equation are over-compensated by the impressing improvement for DRAT, which is now very close to its true value.

3. INVESTIGATION OF THE RELEVANCE OF DIFFERENT DATA PRODUCTION MODES FOR PARAMETER ESTIMATION

At the beginning of section 2. we have described the data production conditions chosen for the following estimation experiments. They were made to come close to the experimental designs of the MIT studies. Thus our critical reinvestigation and divering interpretations do not leave the platform of assumptions underlying these experiments. This does not mean that we accept these assumptions as useful methods for the production of realistic data. In section 2.4., purely random measurement errors have already been debated as one example which deserves attention. In Tab. 6 the estimation results of section 2 are taken as the reference schema (DATMOD1) for diverging modes of generating synthetic data. The indices which describe these conditions are explained in Fig. 1.

Compared with DATMOD1, neither data sampling(DATMOD3) nor reduced length of the time series (DATMOD8) is of major importance for the parameter estimates. The same holds for DATMOD6, where the standard deviation of the measurement error has been linked to the mean instead of the current value of the time series and for the stochasization of gross instead of net rate equations (DATMOD2), which must not be confused with the important alternative of estimating net or gross rate equations as discussed in section 2.1. DATMOD4 proves that reduced measurement error variance is very favourable especially to the estimates of time constants

Parameters	DATMOD 1	DATMOD 2	DATMOD 3	DATMOD 4	DATMOD 5	DATMOD 6	DATMOD 7	DATMOD 8	DATMOD 9 a	DATMOD 9 b	DATMOD 9 c	DATMOD 9 d	DATMOD 9 e	mean
K1	463,4	459,4	450,6	472,1	428,8	443,8	475,0	473,0	422,8	430,3	427,1	436,3	422,5	433,2
K2	-54,6	-53,6	-53,5	-59,3	-50,5	-53,1	-61,5	-55,9	-48,7	-51,3	-50,0	-53,9	-49,6	50,7
K3	0,6476	0,6372	0,5890	0,6086	0,5620	0,6716	0,6197	0,5965	0,5640	0,5751	0,5760	0,5754	0,5922	-0,5765
K4	-0,1332	-0,1316	-0,1200	-0,1232	-0,1030	-0,1505	-0,1259	-0,1204	-0,1084	-0,1154	-0,1127	-0,1139	-0,1269	0,1157
K5	0,00916	0,00918	0,00846	0,00846	0,00619	0,0115	0,00847	0,00803	0,00702	0,00789	0,00731	0,00758	0,00953	0,00787
K6	0,00024	0,00024	0,00028	0,00028	0,00027	0,00025	0,0003	0,00025	0,00028	0,00028	0,00028	0,00028	0,00029	0,00028
K7	-0,0403	-0,0403	-0,0466	-0,0464	-0,0447	-0,0423	-0,05	-0,0419	-0,0462	-0,0466	-0,0469	-0,0465	-0,0477	0,0468
K8*	0,6476	0,6372	0,5890	0,6086	0,5620	0,6716	0,6197	0,5965	0,5640	0,5751	0,5760	0,5754	0,5922	0,5765
K9*	-0,1332	-0,1316	-0,1200	-0,1232	-0,1030	-0,1505	-0,1259	-0,1204	-0,1084	-0,1154	-0,1127	-0,1139	-0,1269	-0,1157
K10*	0,00916	0,00918	0,00846	0,00846	0,00619	0,0115	0,00847	0,00803	0,00702	0,00789	0,00731	0,00758	0,00953	0,00787
(1 DRAT)	0,3764	0,5249	0,6727	0,9077	0,8889	0,3874	0,9761	0,8694	0,8816	0,8906	0,9167	0,9143	0,9244	0,9071
K12	-0,0347	-0,0340	-0,0473	-0,0236	-0,0292	-0,0206	-0,0451	-0,0381	-0,0354	-0,0373	-0,0443	-0,0359	-0,0333	-0,0372
K13	0,0306	0,0298	0,0489	0,0149	0,0198	0,0128 1)	0,0415	0,0325	0,0267	0,0299	0,0380	0,0284	0,0257	0,0297
K15	0,0052	0,0053	0,0013 1)	0,0080	0,0063	0,0078	0,0051	0,0054	0,0054	0,0047	0,0035	0,0051	0,0050	0,0047
K16	0,1070	0,1023	0,1081	0,1877	0,1921	0,0608	0,2250	0,1114	0,1864	0,1999	0,1888	0,1899	0,1844	0,1899
K17	0,1036	0,1010	0,0982	0,1422	0,1429	0,1128	0,1522	0,1115	0,1368	0,1509	0,1456	0,1498	0,1458	0,1458

1) Parameter insignificant (|T-STAT| <2)

DATMOD 1: i=1; j=1; k=1; l=1; m: 1% for u's, 10% for v's; n=1; o=100; p=1
DATMOD 2: i=2
DATMOD 3: j=2 (DT=0,5)
DATMOD 4: m: 1% for u's, 5% for v's
DATMOD 5: k=3 m: 5% for u's, 5% systematic and 5% random for v's
DATMOD 6: l=2
DATMOD 7: n=2
DATMOD 8: o=50
DATMOD 9: —— like DATMOD 5 —— p=5

* Indices defined in Figure 1

Table 6: Influence of different data production modes on estimation results.

of delay's (1/DRAT; K16 = 1/TDDRC; K17 = 1/TDDRM) while a greater errors-in-equation-variance as well as a systematic measurement error component (DATMOD5) are not harmful.

DATMOD9 replicates DATMOD5 five times with different sequences of random numbers to give an impression of the Monte Carlo aspects. Finally, parameter estimates become perfect when the error-corrupted level variables are fed back into the rate equation during data generating (DATMOD7). This is very mich in line with econometric theory which states, that "if decision makers respond to measured data then measurement error is irrelevant and our previous least-squares techniques will be valid."[11] (p. 283).

It seems legitimate to conclude that the proper use of common least-squares methods under data production conditions, which can still be judged as realistic (e.g. DATMOD5), leads to very satisfactory parameter estimates. Furthermore the statistical and econometric criteria are very helpful and hardly misleading yardsticks in the search for adequately specified equations.

4. PREDICTIVE EVALUATION OF ECONOMETRICALLY ESTIMATED SD MODELS

MASS and SENGE, who ended up with rather devestating judgements on least-squares estimation of SD models, have favored model behavior tests as a powerful alternative. To support their view, they compared the standard run of the Market Growth Model (Fig. 4) with a simulation in which the insignificant parameter K2 of the RBL equation (see Tab. 1, col. 3) was set at zero (Fig. 5). They correctly observe that the behavior of the Market Growth Model is altered markedly, and they conclude that "the outcomes of the alternative tests highlight the point that, of the two tests, only model behavior tests reliably measure the importance of an hypothized impact of one variable on the other".[17] (p.457). This conclusion is not defendable.

First, we have shown in section 2.1. that a separate estimation of OB instead of RBL renders a statistically significant influence of delivery delay recognized by market on sales effectiveness. Second, it is wrong to set an insignificant parameter at zero in following simulations. Third, hanging on to their RBL equation, MASS and SENGE should have re-estimated the RBL equation without the variable DDRM attached to the insignificant parameter K2 before model behavior testing. The perception to be made is striking.

$$\text{(2b)} \quad RBL = 69{,}2*S + 0{,}2466*BL - 0{,}0739*BL^2/PC + 0{,}0076*BL^3/PC^2$$

$$DW = 0{,}23 \quad R^2 = 0{,}06$$

The R^2 drops from 0,85 (Tab. 1, col. 3) to 0,06 when the term S * DDRM is eliminated. Thus a model-behavior test including eq. (2b) is unnecessary. The R^2 already tells that DDRM has to remain in the equation and may not be dropped.

Simulation with the parameters estimated from data corrupted by realistic measurement errors (DATMOD5 in Tab. 6) shows a surprisingly weak change in model behavior, as can be seen by comparison of Figs. 4 and 6. This can well be attributed to the insensitivity of the Market Growth Model to the precision of parameter values. Bit it is incorrect to infer that "for acceptance simulation of the complete estimated model, the econometric model-builder must ignore all indications of statistical insignificance in estimation results"[25] (p.57), because we have demonstrated that a proper use of econo-

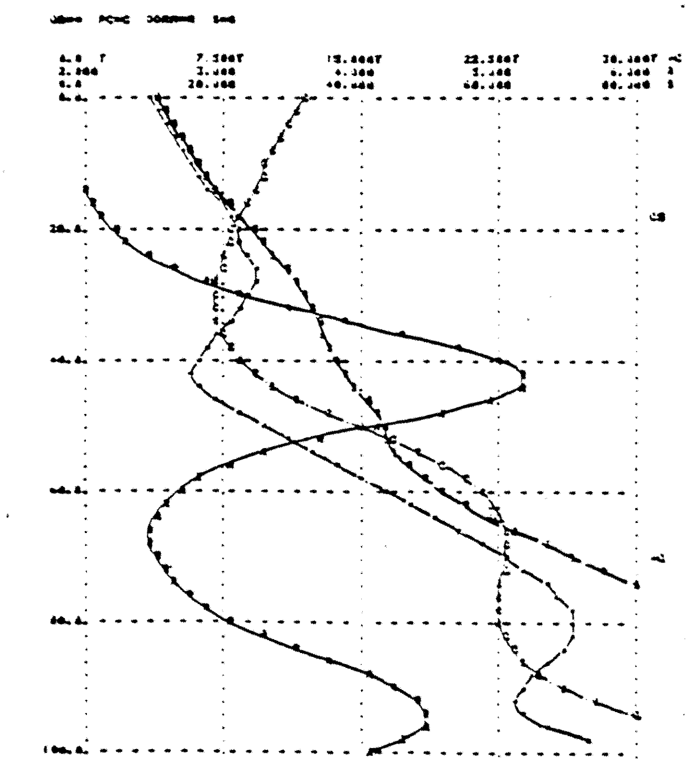


Figure 4: Simulation with true parameters (standard run).

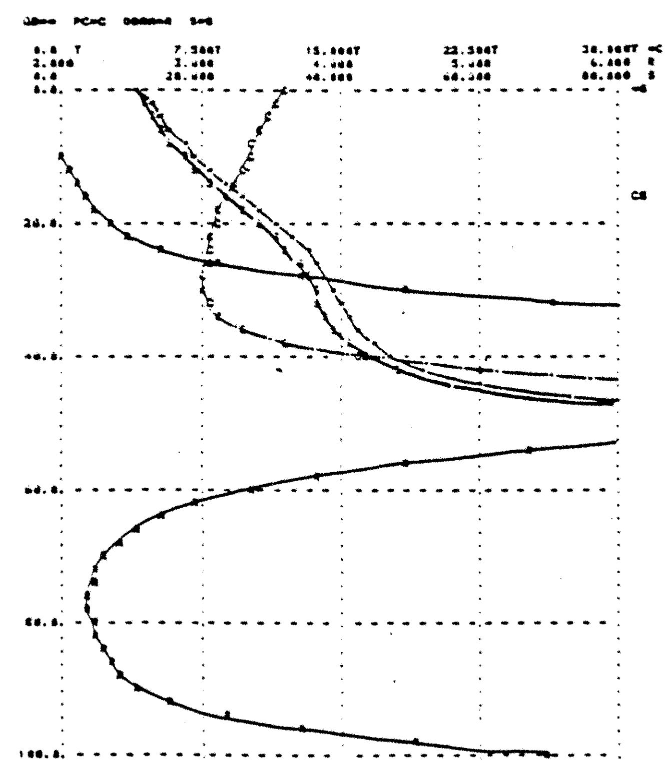


Figure 5: Simulation with insignificant parameter K2 = 0 (Tab. 1, col. 3).

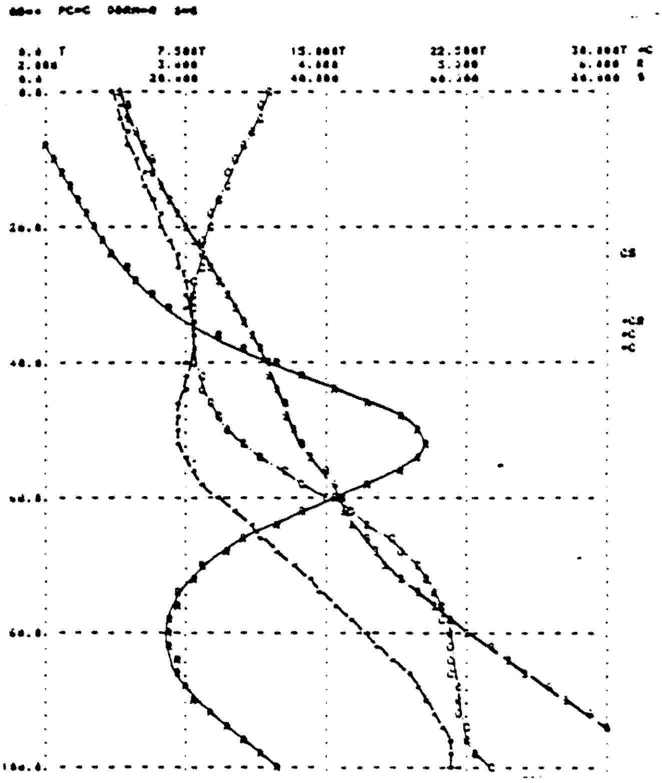


Figure 6: Simulations with parameters estimated from more realistic data.

metric techniques leads to significant estimates. Econometric and model-behavior tests do not contradict but corroborate each other.

5. CONCLUSIONS AND GUIDELINES

Our critical reappraisal of diverse MIT studies on the adequacy of econometric methods for SD models has led to quite encouraging results. While these studies – especially those by SENGE and MASS – ended up in blaming econometrics for misguiding the model-builder practically in all but "ideal" situations, we have boiled down the problems to two areas of real concern. The first is the bias of time-constant estimates of exponential delay elements (shown for first order delays but most probably carried over to higher order delays by cascading), which is in line with asymptotic econometric theory. The degree of bias mainly depends on the standard deviation of the measurement error's random component, what points to the second area of general concern: the kind and size of measurement errors. We have argued that "really realistic" measurement conditions probably lead to more accurate parameters than those obtained from the rather extreme data production devices in the MIT studies. Nevertheless, even these very bad measurement errors rendered parameter estimates which could very well face behavior mode tests such as predictive or policy simulation.

The most important conclusions to be drawn from our experimental reinvestigation is the refutation of FORRESTER's hypothesis that least-squares methods yield "misleading indications from the internal measures of validity". We cannot recognise a need for "a basic reorientation of model building and theory testing in the social sciences"[17] (p. 459). What the MIT authors call for – namely model behavior tests – is long known and practiced in EC model building which relies both on single-equation evaluation by traditional test-statistics as well as on model simulation. Thirty years of simulation experience with hundreds of macro-econometric models disprove the assertation that in "econometric model building, where system models are common, the single-equation philosophy of testing still predominates"[17] (p. 459).

For a sound and careful usage of EC methods within SD model building we suggest to observe the following guidelines:

1. Estimate gross rather than net rate equation.
2. Specify nonlinear functions based on data plots of dependent against independent variables; this is especially applicable in cases where SD normally employs table functions.
3. Be careful with the estimation of exponentially distributed lags (delays) because they are sensitive to strong measurement errors. Plot data against the time axis and consider well, if measurement errors might contribute to an erratic time pattern; there is no chance to directly investigate the measurement error in an nonexperimental research with real world data!
4. Make use of all available a-priori-informations like restrictions on and between parameters.
5. Start with OLS before eventually changing to more sophisticated methods.

This list is by no means ment to be complete and should be ammended by further practical experience. But this requires the readiness to actually employ EC methods in SD modeling, which would be more helpful than pure calls for future research "on possibilities for intergrating the two testing approaches"[16] (p. 834). The possibilities are at hand. Knowledge in this field only grows when we turn this possibility into reality.

ABBREVIATIONS

SD	:	System Dynamics
EC	:	Econometrics
OLS	:	Ordinary Least Squares
ROLS	:	Restricted Ordinary Least Squares
GLS	:	Generalized Least Squares
R3SLS	:	Restricted Three Stage Least Squares
IV	:	Instrumental Variables
FIMLOF	:	Full Information Maximum Likelihood via Optimal Filtering

ACKNOWLEDGEMENTS

Dr. H. HAAS kindly let me use his Econometric Program System (EPS). F.U. HOLZKAMP and R. MEYER arranged the linkage of EPS and DYNAMO necessary for the integration of synthetic data production, estimation and simulation. Professors Dr. G. UEBE and Dr. E ZWICKER as well as Dr. R. FRIEDMANN and an anonymous referee of this journal provided helpful comments on an earlier version.

BIBLIOGRAPHY

1. DENTON, F.T., KUIPER, J. The effect of measurement errors on parameter estimates and forecasts. A case study based on the canadian preliminary national accounts, in: Review of Economics and Statistics, 1965, pp. 198-206.

2. ENGELMANN, M and co-workers: Revision der Volkswirtschaftlichen Gesamtrechnungen 1960 bis 1981, in: Wirtschaft und Statistik, 1982, pp. 551-572.

3. ESENWEIN-ROTHE, I. Genauigkeitsanspruch und Fehlertoleranz gegenüber den Größen der volkswirtschaftlichen Gesamtrechnung, in: Klatt, S., Willms, M. (eds.): Strucktur-wandel und makrookonomische Steuerung, Berlin, 1975, pp. 99-117.

4. FORRESTER, J.W. Industrial Dynamics, Cambridge, Mass., 1961.

5. FORRESTER, J.W. Market growth as influenced by capital investment, in: Industrial Management Review, 1968, pp. 83-105, reprinted in: Forrester, J.W. Collected papers of J.W. Forrester, Cambridge, Mass., 1975, pp. 111-132.

6. FORRESTER, J.W. A national model for understanding social and economic change, in: Simulation Today, 1975, pp. 125-132.

7. FORRESTER, J.W. Educational implications of responses to system dynamics models, in: Churchman, C.W., Mason, R.O.(eds.): World modeling: a dialogue, Amsterdam – Oxford, 1976, pp. 27-35.

8. GOLDBERGER, A.S. Econometric theory, New York – London – Sydney, 1964.

9. GRAHAM, A.K. Parameter formulation and estimation in system dynamics models, in: Randers, J., Ervik, L.K. (eds.): The system dynamics method. The proceedings of the 1976 international conference on system dynamics at Geilo (Norway), Oslo 1977, pp. 541-580.

10. GRAHAM, A.K. Parameter estimation in system dynamics modeling, in: Legasto, A.A., Forrester, J.W., Lyneis, J.M., (eds.): System dynamics, Amsterdam – New York – Oxford, 1980, pp. 125-142.

11. JOHNSTON, J: Econometric methods, Tokyo, 1972.

12. KLEIN, L.R., GOLDBERGER, A.S. An econometric model of the United States 1929-1952, Amsterdam, 1965.

13. KOUTSOYIANNIS, A. Theory of econometrics. An introductory exposition of econometric methods, London – Basingstoke, 1977.

14. LEHMANN, G. Wirtschaftswachstum im Gleichgewicht. Eine System-Dynamics-Studie sozio-ökonomischer Entwicklungen in der Bundesrepublik Deutschland bis zum Jahr 2000, Stuttgart, 1975.

15. MADDALA, G.S. Econometrics, New York, 1977.

16. MASS, N.J., SENGE, P.M. Alternative tests for the selection of model variables, in: Randers, J., Ervik, L.K. (eds.). The system dynamics method. The proceedings of the 1976 international conference on system dynamics at Geilo (Norway), Oslo, 1977, pp. 805-840.

17. MASS, N.J., SENGE, P.M. Alternative tests for the selection of model variables, in: IEEE Transactions on Systems, Man, and Cybernetics, 1978, pp. 450-460.

18. MEISSNER, W., APEL, H. Der systemdynamische Ansatz in der Ökonometrie, in: WISU, 1976, pp. 17-21.

19. MORECROFT, J.D.W. A comparison of OLS and instrumental variable estimation of a dynamic growth model. System Dynamics Group Memorandum D-2729, Cambridge, Mass., 1977.

20. MORGENSTERN, O. On the accuracy of economic observations, Princeton (N.J.), 1963.

21. NAILL, R. Parameterizing social systems models, in: Meadows, D.L. (ed.). Methodological aspects of social system simulation, Hanover (N.H.), 1974, pp. 50-67.

22. PETERSON, D.W. Statistical tools for system dynamics, in: Randers, J., Ervik, L.K. (eds.). The system dynamics method. The proceedings of the 1976 international conference on system dynamics at Geilo (Norway), Oslo, 1977, pp. 841-872.

23. RINNE, H. Die Verläßlichkeit von Sozialproduktsdaten auf der Basis ihrer Revisionen. Eine ökonometrische Studie, dargestellt am Beispiel der Bundesrepublik Deutschland, Meisenheim am Glan, 1969.

24. SCHLEICHER, S. Erweiterungen der Ökonometrie, in: Bruckmann, G. (ed.). Langfristige Prognosen, Würzburg–Wien, 1977, pp. 245-254.

25. SENGE, P.M. An experimental evaluation of Generalized Least Squares estimation. System Dynamics Group Memorandum D-1944-6, Cambridge, Mass., 1974.

26. SENGE, P.M. Statistical estimation of feedback models, in: Simulation, 1977, pp. 177-184.

27. SMITH, V.K. Monte Carlo Methods. Their role in econometrics, Lexington, Mass., 1973.

28. SOMMER, M. System Dynamics und Makroökonometrie. Dynamische makroökonomische Model-lierung in multimethodologischer Sicht, Bern – Stuttgart, 1981.

29. SOMMER, M. Predictive and non-predictive validation strategies for system dynamics models, in: Applied Systems Analysis, 1981, pp. 108-115.

30. SOMMER, M. The econometric challenge to system dynamics and vice versa: some future perspectives, in: Technological Forecasting and Social Change, 1984, pp. 263-280.

31. YOUNG, A.H. Reliability of the quarterly national income and product accounts of the United States, 1947-1971, in: The Review of Income and Wealth, 1974, pp. 1-39.