Staff Paper

PARAMETER IDENTIFICATION IN NONLINEAR STATISTICAL MODELS: A MONTE CARLO ANALYSIS OF CROP-PEST RESPONSE FUNCTIONS

by
Scott M. Swinton and Conrad P. Lyford

February, 1992 **No. 92-14**

Department of Agricultural Economics
MICHIGAN STATE UNIVERSITY
East Lansing , Michigan

Parameter Identification in Nonlinear Statistical Models:

A Monte Carlo Analysis of Crop-Pest Response Functions

by

Scott M. Swinton and Conrad P. Lyford*

Staff Paper 92-14

Department of Agricultural Economics
Michigan State University
East Lansing, MI 48824-1039
tel: 517-353-7218
FAX: 517-336-1800

* Scott M. Swinton is an assistant professor and Conrad P. Lyford is a graduate research assistant in the Department of Agricultural Economics, Michigan State University. They thank Lester V. Manderscheid and James F. Oehmke for helpful comments on an earlier draft.

ABSTRACT:

Parameter Identification in Nonlinear Statistical Models:

A Monte Carlo Analysis of Crop-Pest Response Functions

The form of crop yield-pest density functions influences control thresholds and risk management strategies. Monte Carlo simulation of sigmoidal yield response to weed density suggests that given empirical levels of functional curvature and data variability, large sample sizes are required to reject the null hypothesis of hyperbolic response.

Parameter Identification in Nonlinear Statistical Models:

A Monte Carlo Analysis of Crop-Pest Response Functions

Introduction

Until recently, limits on computational power have restricted most nonlinear statistical estimation to models that could be made linear by a simple transformation or by polynomial expansion (Heady and Dillon). For static linear models with independently and identically distributed (i.i.d.) normal disturbances, collinearity due to ill-conditioned design matrices is the principal impediment to parameter identification. For intrinsically nonlinear models, however, the set of identifiability problems extends beyond the design matrix to the functional form itself and the criterion function used to estimate it (Belsley; Seber and Wild).

Parameter identification is particularly important for testing hypotheses about nonlinear functional form. In the analysis of optimal pest management, the form of the crop-pest yield function is critical, especially for agents who are not risk-neutral. In weed management, this centers around the enduring debate over whether crop response to weed density is best described by a sigmoidal or a hyperbolic yield function (Cousens, Pannell 1991, Swinton). From a management perspective, the steep initial slope of a hyperbolic yield function means that at low weed densities, crop yield damage per weed is greatest. This implies a low economic threshold for weed control at given input costs and product prices. The gentle initial slope of a sigmoidal curve implies a higher weed control threshold. A hyperbola everywhere convex to the origin implies that weed control inputs can be risk-increasing when weed density or product price is uncertain (Pannell 1990). Sigmoidal yield response implies the opposite over the concave portion of the yield function's range (Swinton).

While Cousens found his hyperbolic curve to outperform a variety of other forms published in the weed literature, none of those he evaluated was sigmoidal. The only sigmoidal yield function that appears in print is a logistical curve (King et al.). However, the inherent symmetry of this form makes it unsuited to capturing nonsymmetrically curved yield response.

The question remains whether a flexible sigmoid provides a better fit to crop-weed competition data than the hyperbolic alternative. This begs a more fundamental question: Can the variability of field data support estimating the number of parameters necessary for a sigmoidal curve?

Consider the first question. We wish to fit the basic model: $Y = f(x,\theta)$. Apart from the biologically unacceptable linear model (which can predict negative yields), two classes of functional form are of interest: a) concave hyperbolic (3 parameters--e.g., Cousens), and b) sigmoid functions (3-4 parameters--e.g., King et al.). The ideal way of evaluating the degree of nonlinearity of the model is to choose a flexible sigmoidal function that nests the hyperbolic model as a special case. This admits several statistical tests, including the likelihood ratio, Lagrange multiplier, and Wald tests (Bates and Watts, Gallant, Judge et al.). A less attractive alternative is to compare non-nested models using a mean squared error criterion (Bates and Watts).

The Morgan-Mercer-Flodin or MMF model (Morgan et al.; Ratkowsky; Seber and Wild) is a flexible sigmoid which embodies the Cousens hyperbola as a special case. The MMF model takes the form:

$$Y = \frac{\beta\gamma + \alpha x^{\delta}}{\gamma + x^{\delta}} \tag{1}$$

The parameters have the following interpretations: α is the minimum yield asymptote as weed density (x) approaches infinity, β is the maximum (weed-free) yield, γ is a curvature measure that determines the rate at which yield reaches its lower asymptote (i.e., the lower curve of the sigmoid), and δ is a curvature measure that determines the point at which yield begins to decline at a decreasing rate (i.e., the upper curve of the sigmoid).

The Cousens hyperbola takes the form:

$$Y = Y^{0}\left[1 - \frac{Ix}{(1 + Ix/A)}\right] \tag{2}$$

where Y^0 represents maximum yield, I represents the proportion of crop yield loss per unit of weed density (x) as density approaches zero, and A represents the maximum proportion of crop yield loss asymptote as weed density approaches infinity.

Cousens' parameters can be translated into MMF coefficients as follows:

$\alpha = Y^0(1 - A)$, $\beta = Y^0$, $\gamma = A/I$, and $\delta = 1$. One test for the sigmoid model is to evaluate the null hypothesis that $\delta = 1$ in the MMF model. Acceptance would support use of Cousens' hyperbolic functional form.

An empirical example

The MMF model was estimated using several sets of corn and soybean yield data as a function of weed density. The model was fit using nonlinear maximum likelihood estimation with SHAZAM's quasi-Newton, Davidson-Fletcher-Powell algorithm (White et al.), assuming additive disturbances distributed i.i.d. normal$(0,\sigma^2)$. Following Ratkowsky, the original MMF model was reparameterized to reduce parameter effects nonlinearity, substituting e^{γ^*} for γ (so $\gamma^* = \ln(\gamma)$). Typical results came from fitting this model to Buhler et al.'s 1989-90 Minnesota soybean yield, cocklebur weed density data. Figure 1 illustrates the data along with the fitted curves for the MMF sigmoid and Cousens hyperbolic functions. As crop yield-weed density data sets go, this one gives an unusually clear picture of the functional relationship.

Parameter estimates and asymptotic t-values for the MMF model with restriction and with the restriction $\delta = 1$ are presented in Table 1. While asymptotic t-values must be interpreted with caution (Ratkowsky), it is clear that these model parameters are poorly determined. Only the β and $\ln(\gamma)$ estimates appear to be robust relative to the null hypotheses α, β and $\gamma = 0$, and $\delta = 1$. The Wald and likelihood

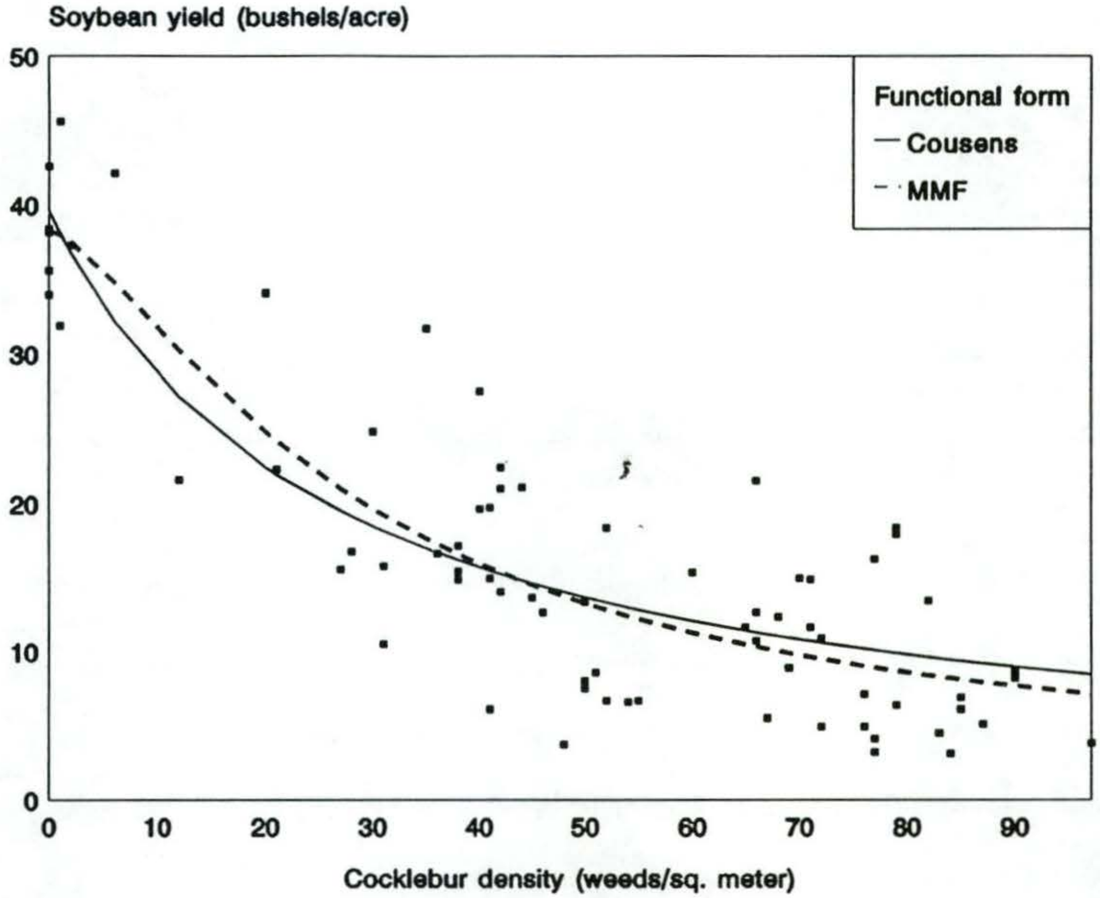

Figure 1: Soybean yield as a function of cocklebur weed density.

ratio test statistics (Judge et al., p. 545) are 0.54 and 0.63, respectively. Since both are far inferior to the $\chi^2(1,.05)$ test value of 3.84, the null hypothesis $\delta = 1$ cannot be rejected. These data do not support estimation of the 4-parameter MMF sigmoid model.

The asymptotic coefficient correlation matrix (Table 2) suggests a possible source of the problem: almost perfect positive correlation of the γ and δ parameters.

Table 1. Parameter estimate statistics from MMF regression of soybean yields on cocklebur weed density.

	Unrestricted Model		Restricted Model	
Parameter	Estimate	Asymptotic t-value	Estimate	Asymptotic t-value
Alpha	1.4	0.2	- 7.7	- 1.2
Beta	38.4	21.1	39.0	20.3
LN(Gamma)	4.8	2.9	3.8	10.8
Delta	1.4	2.5	1.0 by restriction	

Log-likelihood ratio: Unrestricted: -233.1331; Restricted: -233.4493
Asymptotic standard error of estimate (SEE): Unrestricted: 5.37; Restricted: 5.39.
R^2 between observed and predicted: Unrestricted: 0.75; Restricted: 0.75.

In fact, in the Jacobian matrix used to identify the gradient vector for the solution algorithm, γ and δ figure in every term, and both figure quadratically in the $dY/d\gamma$ and $dY/d\delta$ terms. More intuitively, since these parameters determine segments of the sigmoid with opposite convexity, both must be adjusted to compensate for data in a given range. Yet the γ and δ coefficients of the MMF model are not intrinsically unidentified in the sense that neither can be substituted for a simple function of the other.

Table 2. Asymptotic correlation matrix from unrestricted MMF regression of soybean yields on cocklebur weed density (original parameterization).

Parameter	Alpha	Beta	Gamma	Delta
Alpha	1.000			
Beta	-.219	1.000		
Gamma	.834	-.399	1.000	
Delta	.920	-.326	.982	1.000

Monte Carlo test of parameter identification with the MMF function

The example above illustrates the difficulty of fitting the MMF function to empirical data. This leads to the second question: Can routine field data ever support estimation of a sigmoidal curve? A Monte Carlo experiment was designed to test conditions under which the Cousens hyperbola model could not be rejected in spite of a true, underlying MMF model with an additive, i.i.d. normal disturbance. Factors likely to affect the ability to obtain reliable coefficient estimates include the inherent curvature of the true model, the sample size, random variability in the data, and both the domain and distribution of observations on the independent variable. Parameter identifiability was expected to improve with increases in inherent curvature, sample size and breadth of functional domain, and with decreases in random variability. The Monte Carlo experimental design varied all factors but functional domain, setting the weed density sufficiently broadly to facilitate estimation of the asymptote in all cases.

Two "true" models were designed to be similar to the soybean-cocklebur yield function illustrated in Figure 1. Both had maximum yield (β) parameters of 40 and minimum yield (α) parameters of 4. One had weak sigmoidal curvature ($\delta = 1.5$ and $\ln(\gamma) = 6.4$) and the other strong sigmoidal curvature ($\delta = 3$ and $\ln(\gamma) = 14.2$). These are illustrated in Figure 2, along with a hyperbolic model ($\delta = 1$, $\ln(\gamma) = 3.8$). The weed densities required to induce a 2% yield loss in the hyperbolic, weakly sigmoid, and strongly sigmoid models, are 1, 6 and 32 weeds/m^2, respectively.

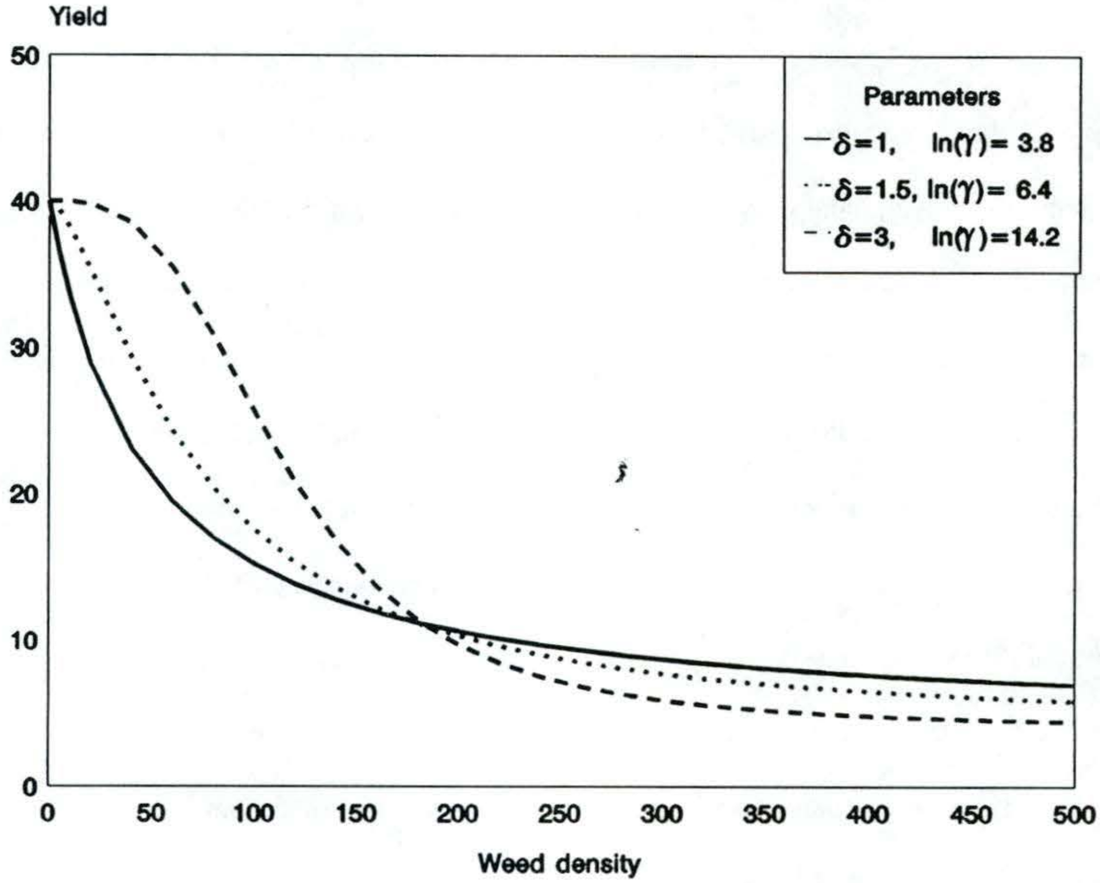

Figure 2: Yield functions synthesized (α=4 and β=40 in all cases).

Yield data were synthesized from a uniformly distributed population of 500 weed densities over the domain [0, 499] weeds/m^2. To these were added i.i.d. normal disturbances with mean 0 and standard deviations σ = 1, 3, and 5, subject to the constraint that yield not fall below zero. One hundred data sets were generated for each of three sample sizes (n = 50, 200, 500). The MMF model was estimated for each of these cases using coefficient initial values equal to the known true values.

Results support expectations that identifiability of the upper curvature parameter, δ, declines with increases in variance and decreases in sample size. Identifiability increases with the degree of inherent curvature in the underlying model. These points are clearly illustrated in Table 3, which reports the percentage of cases in which a Wald χ^2 test of the null hypothesis $\delta = 1$ would correctly be rejected. Of particular interest is the $\delta = 1.5$ model with $\sigma = 5$, since this approximates parameters from the soybean-cocklebur data set. Only at a sample size of 500 is the null hypothesis rejected on a regular basis.

Table 3. Percent of cases that Wald test at 5% significance level led to correct rejection of hypothesis that $\delta=1$ in 100 MMF regressions, by model curvature, standard deviation and sample size.

	Degree of curvature (δ) and sample size					
	$\delta = 3$			$\delta = 1.5$		
σ	500	200	50	500	200	50
	- - - - - - - - - - - - - percent - - - - - - - - - - - - -					
1	100	100	100	100	100	93
3	100	100	99	100	84	33
5	100	100	88	90	51	14

Another approach to evaluating parameter identifiability is to examine the significance of coefficient asymptotic t-statistics. Ratkowsky (p. 34) suggests that while a high asymptotic t-value may not insure that a parameter is well determined, a low one indicates that it is not. The variability in coefficient asymptotic t-values is summarized in Table 4. It reports t-tests of the null hypothesis that mean coefficient

Table 4. T-tests that coefficient asymptotic t-values exceed 1.96 in 100 MMF regressions.

Size of original regression samples and parameter name	Model and Standard Deviation $\delta = 3$			$\delta = 1.5$		
	σ=1	σ=3	σ=5	σ=1	σ=3	σ=5
N = 500						
Alpha	22.50	7.88	5.14	10.78	3.09	1.76
Beta	33.50	25.58	19.49	23.16	14.05	9.12
Gamma	31.07	18.95	15.60	25.45	16.14	9.83
Delta	30.18	16.02	11.69	13.89	3.68	1.41
N = 200						
Alpha	9.11	3.39	2.17	4.66	1.17	0.58
Beta	8.63	7.96	7.16	4.93	4.19	3.26
Gamma	14.09	12.21	8.52	7.78	5.53	3.58
Delta	13.68	9.78	5.99	5.32	1.03	-0.02
N = 50						
Alpha	4.81	1.31	0.67	1.86	0.32	0.19
Beta	3.63	3.27	2.87	1.90	1.53	1.21
Gamma	5.30	3.69	2.17	2.50	1.23	0.37
Delta	5.28	2.84	1.12	1.44	-0.46	-0.87

[1]T-tests were constructed as follows: $t = (|t_r| - 1.96)/s_r$, where t_r is the mean of 100 regression coefficient asymptotic t-values (t_r's) and s_r is the standard deviation of those 100 t_r's.

The one-tailed critical value at the .05 significance level is 1.66.

[2]Since the null hypothesis is $\delta = 1$, asymptotic t-values were calculated as $(\hat{\delta}-1)/\hat{s}$, where $\hat{s}$ is the asymptotic standard error.

---

asymptotic t-statistics from the 100 regressions are "significant" in the sense of exceeding 1.96. Since 1.96 is the 95% confidence level critical value for a two-tailed t-test with large sample size, higher t-values imply that the estimated coefficient indicates an identifiable parameter. For parameters α, β and γ, identifiable means non-zero; for δ, it means different from 1. T-test statistics in Table 4 whose absolute values exceed the one-tailed critical value 1.66 indicate that asymptotic t-values would be "significant" in more than 95% of cases. When underlying sigmoidal curvature is weak ($\delta=1.5$) and data are highly variable $\sigma=5$, Table 4 suggests that even 500 observations are insufficient to determine δ at that confidence level. With $\sigma=1$, 200 observations will reliably suffice to determine the curvature parameters, but 50 observations will not be adequate to determine δ. When sigmoidal curvature is strong ($\delta=3$), 200 observations are adequate at the $\sigma=5$ level or 50 observations at the $\sigma=3$ level of variability.

The δ and γ coefficients in the Monte Carlo experiments had correlation coefficients in the 0.98 - 0.99 range. In spite of the very high correlation, parameter determination did not suffer noticeably. This may owe, in part, to use of true parameter values to initialize the estimation procedure.

Conclusion

These results highlight the importance of inherent curvature, data variability, and sample size for parameter identification in inherently nonlinear models. In particular, even if the crop yield-weed density response function is truly sigmoidal, the null hypothesis of hyperbolic response is hard to reject under likely levels of inherent curvature and ordinary levels of field data variability. Yet acceptance of a hyperbolic response function tends to imply a lower weed control threshold than does the alternative sigmoid response function.

Of the determinants of parameter identification examined here, sample size is the only one that can be manipulated by researchers. The results suggest that identification of a 4-parameter sigmoid crop yield-weed density response function requires a much larger data set than is the norm for agronomic experiments. A definitive study to test the sigmoidal yield response hypothesis will also require an experiment specifically designed to optimize parameter identifiability. Apart from large sample size, this will likely call for a non-uniform distribution of weed densities and variance reduction through inclusion of additional independent variables.

REFERENCES

Bates, Douglas M. and Donald G. Watts. Nonlinear Regression Analysis and Its Applications. New York: Wiley, 1988.

Belsley, David A. Conditioning Diagnostics: Collinearity and Weak Data in Regression. New York: Wiley, 1991.

Buhler, D., J. Gunsolus and Ralston. "Reduced bantazon and cultivation for cocklebur control." Unpublished research protocol. Department of Agronomy and Plant Genetics, University of Minnesota, 1989.

Cousens, Roger. "A Simple Model Relating Yield Loss to Weed Density." Annals of Applied Biology 107(1985): 239-252.

Deen, W., A. Weersink, C.G. Turvey, and S. Weaver. "Economics of Weed Control Strategies under Uncertainty." Unpublished paper. Department of Agricultural Economics and Business, University of Guelph, Ontario (Sep. 1991).

Heady, Earl O. and John L. Dillon. Agricultural Production Functions. Ames, IA: Iowa State University Press, 1961.

Judge, G.G., R.C. Hill, W.E. Griffiths, H. Lutkepohl, and T.C. Lee. Introduction to the Theory and Practice of Econometrics. 2nd edition. New York: Wiley, 1988.

King, R.P., D.W. Lybecker, E.E. Schweizer, and R.L. Zimdahl. "Bioeconomic Modeling to Simulate Weed Control Strategies for Continuous Corn (Zea mays)." Weed Science 34(1986): 972-979.

Morgan, P.H., L.P. Mercer, and N.W. Flodin. "General Model for Nutritional Responses of Higher Organisms." Proceedings of the National Academy of Science U.S.A. 72(1975): 4327-4331.

Pannell, David J. "Responses to Risk in Weed Control Decisions under Expected Profit Maximization." Journal of Agricultural Economics 41(1990): 391-403.

--------. "Responses to Risk in Weed Control Decisions under Expected Profit Maximization: Reply." Journal of Agricultural Economics 42(1991): 407-408.

Ratkowsky, David A. Nonlinear Regression Modeling: A Unified Practical Approach. New York: Dekker, 1983.

Seber, G.A.F. and C.J. Wild. Nonlinear Regression. New York: Wiley, 1989.

Swinton, S.M. "Responses to Risk in Weed Control Decisions Under Expected Profit Maximisation: Comment." Journal of Agricultural Economics 42(1991): 404-406.

White, Kenneth J., S. Donna Wong, Diana Whistler and Shirley A. Haun. SHAZAM Econometrics Computer Program: User's Reference Manual Version 6.2. Montreal: McGraw-Hill, 1990.