A Note on the Interpretation of Polynomial Regressions

by

A. Ronald Gallant

Institute of Statistics
North Carolina State University
P. O. Box 5457
Raleigh, N. C. 27650

Often, data which have actually been generated according to a (nonlinear) regression model

$$y_t = f(x_t, \theta) + e_t$$

are approximated by a polynomial regression, say,

$$y_t = a_0 + a_1 x_t + a_2 x_t^2 + a_3 x_t^3 + e_t$$

and the fitted function

$$\hat{f}(x) = \Sigma_{\alpha=0}^{m} \hat{a}_\alpha x^\alpha$$

is regarded as an approximation of the true response function $f(x,\theta)$.

Some unresolved issues are:

1. In what sense does $\hat{f}(x)$ approximate $f(x,\theta)$? Is it a local approximation at a point or a uniform approximation over an interval?
2. How do the coefficients a_0, a_1, a_2, ..., a_m of the approximation

   $$f(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_m x^m$$

   relate to the true response function $f(x,\theta)$?
3. Can the function $f(x,\theta)$ be approximated arbitrarily closely by taking the degree m of the polynomial suitably large?

A Fourier series approach seems to be a useful conceptual framework with which to address these issues. Also, a consideration of the problem in large samples rather than finite samples seems to contribute more to understanding.

For the moment, consider the single variable case

Fitted: $y_t = a_0 + a_1 x_t + a_2 x_t^2 + \ldots + a_m x^m + e_t$

True: $y_t = f(x_t, \theta) + e_t$.

Let the sequence x_1, x_2, x_3, ... of independent variables be stationary in the sense that the empirical distribution function

$$\hat{F}_n(x) = \frac{1}{n}(\text{the number } x_t \leq n \quad \text{for } t = 1, 2, \ldots, n)$$

converges to some distribution function $F(x)$ at every point where $F(x)$ is continuous; that is, $\lim_{n\to\infty}\hat{F}_n(x) = F(x)$ if F is continuous at x. Subject to regularity conditions stated later, it is shown that

1. The least squares estimator

   $$\hat{a} = (\hat{a}_0, \hat{a}_1, \ldots, \hat{a}_m)'$$

   which minimizes $\Sigma_{t=1}^{n}[y_t - \Sigma_{\alpha=0}^{m} a_\alpha x^\alpha]^2$ converges almost surely to a point

   $$\bar{a}(\theta) = [\bar{a}_0(\theta), \bar{a}_1(\theta), \ldots, \bar{a}_m(\theta)]'$$

   which minimizes

   $$\int_{-\infty}^{\infty}[f(x,\theta) - \Sigma_{\alpha=0}^{m} a_\alpha x^\alpha]^2 \, dF(x) .$$

2. The point $\bar{a}(\theta)$ is computed as $\bar{a}(\theta) = H^{-1}h(\theta)$ where the elements $h_{\alpha\beta}$ of H are

   $$h_{\alpha\beta} = \int_{-\infty}^{\infty} x^\alpha x^\beta \, dF(x)$$

   and those of $h(\theta)$ are

   $$h_\alpha(\theta) = \int_{-\infty}^{\infty} x^\alpha f(x,\theta) \, dF(x) \quad .$$

3. If the measure of the error of approximation is taken to be

   $$\int_{-\infty}^{\infty} [f(x,\theta) - \Sigma_{\alpha=0}^{m} a_\alpha x^\alpha]^2 \, dF(x)$$

   then it may be made arbitrarily small by taking m sufficiently large.

These facts may be related to a Fourier series expansion of $f(x,\theta)$ as follows. Let $Q_0(x)$, $Q_1(x)$, $Q_2(x)$, ... be a sequence of polynomials, the first of degree 0, the second of degree 1, and so on, which are orthonormal with respect to $F(x)$. For example, if $F(x)$ were the exponential distribution on $(0,\infty)$ then $Q_0(x)$, $Q_1(x)$, ... are essentially the Laguerre polynomials; if $F(x)$ were the unit normal distribution on $(-\infty, \infty)$ then $Q_0(x)$, $Q_1(x)$, ... are essentially the Hermite polynomials. Let $f(x,\theta)$ have the Fourier series expansion

$$f(x,\theta) = \Sigma_{\alpha=0}^{\infty} b_\alpha(\theta)\, Q_\alpha(x)$$

where

$$b_\alpha(\theta) = \int_{-\infty}^{\infty} f(x,\theta)\, Q_\alpha(x)\, dF(x) \quad .$$

If this expansion is truncated at m terms, viz.,

$$\bar{f}(x) = \Sigma_{\alpha=0}^{m} b_\alpha(\theta)\, Q_\alpha(x)$$

then

$$\bar{f}(x) = \Sigma_{\alpha=0}^{m} \bar{a}_\alpha(\theta)\, x^\alpha \quad .$$

Thus, if one approximates the true regression model

$$y = f(x_t,\theta) + e_t$$

by a polynomial regression

$$y = \Sigma_{\alpha=0}^{m} a_\alpha x^\alpha + e_t$$

the fitted function

$$\hat{f}(x) = \Sigma_{\alpha=0}^{m} \hat{a}_\alpha x^\alpha$$

is estimating the truncated Fourier series expansion

$$\bar{f}(x) = \Sigma_{\alpha=0}^{m} b_\alpha(\theta)\, Q_\alpha(x)$$

where $\{Q_\alpha(x)\}_{\alpha=0}^m$ is a system of orthonormal polynomials with respect to the limiting distribution $F(x)$ of the independent variables.

These results are proved with somewhat more generality as x is permitted to be multivariate. The function $f(x,\theta)$ is taken to be continuous on a compact set $\mathcal{X}$ and the independent variables x_t are restricted to $\mathcal{X}$ to permit application of the more familiar ideas of convergence in distribution in the proofs. If desired, unbounded $\mathcal{X}$ may be accommodated by applying the notion of Cesaro summable sequences (Gallant and Holly, 1980). A few very elementary facts about Hilbert spaces are used; a concise reference is Section 16 of Hewit and Stromberg (1965).

Assumptions. Let $\{x_t\}_{t=1}^\infty$ be a sequence of k-vectors from a compact set $\mathcal{X}$ such that the empirical distribution function $\hat{F}_n(x)$ of $\{x_t\}_{t=1}^n$ converges to a distribution function $F(x)$ at each $x \in \mathcal{X}$ where $F(x)$ is continuous. Let $F(x)$ possess a moment generating function in a neighborhood of zero and let $f(x,\theta)$ be a continuous function of x. Let the sequence of random variables $\{y_t\}_{t=1}^\infty$ be generated according to the regression model $y_t = f(x_t,\theta) + e_t$ where the errors $\{e_t\}_{t=1}^\infty$ are a sequence of independent and identically distributed random variables each with mean zero and finite variance.

Notation. Let x be a k-vector and let the sequence $\{z_\alpha(x)\}_{\alpha=0}^\infty$ consist of terms of the form

$$z_\alpha(x) = x_1^{j_1} x_2^{j_2} \cdots x_k^{j_k}$$

ordered such that the degree $\Sigma_{i=1}^k j_i$ of the terms is non-decreasing in α. The collection of measurable functions $g(x)$ that are square integrable with respect to the distribution $F(x)$ is denoted by $\mathcal{L}_2(\mathcal{X}, dF)$. The matrix of order $(m+1) \times (m+1)$ with typical element

$$h_{\alpha\beta} = \int_{\mathcal{X}} z_\alpha(x)\, z_\beta(x)\, dF(x) \qquad \alpha,\beta = 0,1, \ldots, m$$

is denoted by H ; $h(\theta)$ denotes the $(m+1)$ - vector with typical element

$$h_\alpha(\theta) = \int_{\mathfrak{X}} z_\alpha(x)\ f(x,\theta)\ dF(x) \qquad \alpha = 0,1,\ \ldots,\ m\ .$$

Let $\{Q_\alpha\}_{\alpha=0}^{\infty}$ be the sequence generated from $\{z_\alpha(x)\}_{\alpha=0}^{\infty}$ by the Gram-Schmidt orthonormalization process. If H^{-1} is factored as $H^{-1} = P'P$ where P is lower triangular then $Q_\alpha(x) = \Sigma_\beta p_{\alpha\beta} z_\beta(x)$ for $\alpha = 0,1,\ \ldots,\ m$. Let

$b_\alpha(\theta) = \int_{\mathfrak{X}} Q_\alpha(x)\ f(x,\theta)\ dF(x)$ $\alpha = 0,1,\ \ldots,\ m$ which are the Fourier coefficients of $f(x,\theta)$.

Theorem 1. Let the assumptions listed above hold and $\{z_\alpha(x)\}_{\alpha=0}^{m}$ be linearly independent a.e. $F(x)$. Then H is positive definite, $\bar{a}(\theta) = H^{-1}h(\theta)$ is the unique minimum of

$$S(a) = \int_{\mathfrak{X}} [f(x,\theta) - \Sigma_{\alpha=0}^{m}\ a_\alpha z_\alpha(x)]^2\ dF(x)$$

and

$$\Sigma_{\alpha=0}^{m}\ \bar{a}_\alpha(\theta)\ z_\alpha(x) = \Sigma_{\alpha=0}^{m}\ b_\alpha(\theta) Q_\alpha(x)\ .$$

Proof. Now $\{z_\alpha(x)\}_{\alpha=0}^{\infty} \in \mathcal{L}_2(\mathfrak{X},dF)$ because $F(x)$ has a moment generating function in a neighborhood of zero. Thus H and $h(\theta)$ are well defined. Since $t'Ht = \int_{\mathfrak{X}} [\Sigma_{\alpha=0}^{m}\ t_\alpha\ z_\alpha(x)]^2\ dF(x) \geq 0$ and $\{z_\alpha(x)\}_{\alpha=0}^{m}$ are linearly independent, H is positive definite. Now

$$\begin{aligned} S(a) = {} & \int_{\mathfrak{X}} [f(x,\theta) - \Sigma_{\alpha=0}^{m}\ \bar{a}_\alpha(\theta)\ z_\alpha(x)]^2\ dF(x) \\ & + [a - \bar{a}(\theta)]' H [a - \bar{a}(\theta)] \\ & + 2\Sigma_{\beta=0}^{m} \int_{\mathfrak{X}} [f(x,\theta) - \Sigma_{\alpha=0}^{m}\ \bar{a}_\alpha(\theta)\ z_\alpha(x)] z_\beta(x) [a_\beta - \bar{a}_\beta(\theta)]\, dF(x)\ . \end{aligned}$$

The last term may be rewritten as

$$\begin{aligned} & 2\Sigma_{\beta=0}^{m} [h_\beta(\theta) - \Sigma_{\alpha=0}^{m} \bar{a}_\alpha(\theta)\ h_{\alpha\beta}][a_\beta - \bar{a}_\beta(\theta)] \\ & = 2[h(\theta) - H\ \bar{a}(\theta)]'[a - \bar{a}(\theta)] \end{aligned}$$

which is zero since $\bar{a}(\theta) = H^{-1}h(\theta)$. Thus S(a) is a constant plus a positive definite quadratic form in $a - \bar{a}(\theta)$ and has unique minimum at $a - \bar{a}(\theta) = 0$.

Now

$$\Sigma_{\alpha=0}^{m} b_{\alpha}(\theta) Q_{\alpha}(\theta) = \Sigma_{\alpha=0}^{m} \Sigma_{\beta=0}^{m} p_{\alpha\beta} h_{\beta}(\theta) Q_{\alpha}(\theta)$$

$$= \Sigma_{\alpha=0}^{m} \Sigma_{\beta=0}^{m} \Sigma_{\gamma=0}^{m} p_{\alpha\beta} h_{\beta}(\theta) p_{\alpha\gamma} z_{\gamma}(x)$$

$$= \Sigma_{\gamma=0}^{m} \bar{a}_{\gamma}(\theta) z_{\gamma}(x) \quad . \quad \square$$

Theorem 2. Let the assumptions listed above hold and let $\{z_{\alpha}(x)\}_{\alpha=0}^{m}$ be linearly independent a.e. $F(x)$. Then the least squares estimator

$$\hat{a}_n = (\hat{a}_0, \hat{a}_1, \ldots, \hat{a}_m)'$$

which minimizes

$$\Sigma_{t=1}^{n}[y_t - \Sigma_{\alpha=0}^{m} a_{\alpha} z_{\alpha}(x)]^2$$

converges almost surely to $\bar{a}(\theta)$ as n tends to infinity.

Proof. Let

$$z_t = [z_1(x_t), z_2(x_t), \ldots, z_m(x_t)] \ .$$

Then

$$\hat{a} = (\Sigma_{t=1}^{m} z_t z_t')^{-1}(\Sigma_{t=1}^{m} f(x_t,\theta) z_t + \Sigma_{t=1}^{n} e_t z_t) \ .$$

By the Helly-Bray theorem we have

$$\ell im_{n\to\infty}(1/n)\Sigma_{t=1}^{n} z_t z_t' = H$$

$$\ell im_{n\to\infty}(1/n)\Sigma_{t=1}^{n} f(x_t,\theta) z_t = h(\theta)$$

and by Theorem 3 of Jennrich (1969)

$$\ell im_{n\to\infty}(1/n)\Sigma_{t=1}^{n} e_t z_t = 0 \quad \text{a.s.} \quad \square$$

Theorem 3. Let the assumptions listed above hold. Then

$$\ell im_{m\to\infty} \int_{-\infty}^{\infty} [f(x,\theta) - \Sigma_{\alpha=0}^{m} a_{\alpha} z_{\alpha}(x)]^2 \, dF(x) = 0 \quad .$$

Proof. A direct consequence of Theorem 3 of Gallant (1980) which shows that the polynomials in R^k are complete with respect to $\mathcal{L}_2(\mathcal{X}, dF)$ provided $F(x)$ possesses a moment generating function. ▯

References

Gallant, A. Ronald and Alberto Holly (1980). "Statistical Inference in an Implicit, Nonlinear, Simultaneous Equation Model in the Context of Maximum Likelihood Estimation," Econometrica (in press).

Hewitt, Edwin and Karl Stromberg (1965). Real and Abstract Analysis, New York: Springer-Verlag, Inc.

Gallant, A. Ronald (1980). "Explicit Estimators of Parametric Functions in Nonlinear Regression," Journal of the American Statistical Association (in press).

Jennrich, Robert I. (1969). "Asymptotic Properties of Nonlinear Least Squares Estimators," Annals of Mathematical Statistics 40, 633-643.