

# Edmond Malinvaud: A Tribute to his Contributions in Econometrics

Peter C. B. Phillips  
Yale University, University of Auckland,  
University of Southampton, Singapore Management University

April 23, 2015

*“Malinvaud stands as one of the enduring figures of 20th century economics. His passing is a sad and permanent loss to the French academy, to the profession at large, and to the world of econometrics where he reinstated in our discipline the rigor of the Cowles Commission researchers of the 1940s and delivered a unified thematic for the field that echoed the genius of Ragnar Frisch.”*<sup>1</sup>

Within the economics profession for several decades in the mid twentieth century the French economist Edmond Malinvaud stood virtually unchallenged in his mastery of our discipline, raising the central chambers of the modern edifice of quantitative economics to new heights of rigor and empirical relevance. His mission as an educator and researcher drew its motive power from the Ragnar Frisch mantra of unification that had established a new frontier for the discipline where ideas from economic theory conjoined with mathematical method and statistical methodology to advance our understanding of economic activity and to illumine the implications of economic policy making.

While this mantra of unification remains deeply respected, field fragmentation and research specialization in all areas of economics over recent years has considerably diminished its observance. But as a young scientist arriving at the Cowles Commission as a Visiting Research Fellow in 1950, Edmond Malinvaud was consumed with a passion for quantitative economic research in the widest sense that married perfectly with the Frisch mantra. Over his career as a thinker and writer, Malinvaud’s work became an exemplar of the strengths of unification. In education as well as research, Malinvaud commanded the high ground, leading the profession with the astonishing feat of a triple opus of advanced textbooks that trained an entire generation of economists in advanced principles of microeconomic theory, macroeconomics, and the statistical methods of econometrics.

---

<sup>1</sup>From the author’s (2015) obituary: “Memorial to Edmond Malinvaud.”

The treatise on econometrics was a masterpiece of brilliant exposition. It brought together economic ideas, statistical method, inferential tools, and probabilistic underpinnings in a multitude of stochastic models that demonstrated their utility in what Malinvaud saw as the ultimate mission of all econometric work – “the empirical determination of economic laws”. A mission that contributes to society through data collection, modeling economic activity at all levels, prediction in the wide sense that includes forecasting exercises and modern treatment effect learning about the effectiveness of economic policy interventions.

This tribute to Edmond Malinvaud briefly reviews what Malinvaud himself saw as his main contributions to the field of econometrics, all of which are embodied in his exceptional treatise. The Collège de France, where Malinvaud held the Chaire d’Analyse Économique over the years 1987-1993, posts on its website Malinvaud’s résumé, which classifies his contributions to economics under the three central pillars of the discipline: Théorie Microéconomique, Macroéconomie, and Méthodes Économétriques. Succinct as ever, an entry of a single ouvrage graces the econometric work in his Bibliographie: the original 1964 French edition of his econometric opus, *Méthodes Statistiques de l’Économétrie* (with quiet mention of its 4<sup>th</sup> edition in 1981 and its many foreign language translations), which we henceforth reference as MSE or just Malinvaud.

Malinvaud’s novel contributions to econometrics in MSE lie in two areas, areas that embodied much of the extant theory of econometrics at the time MSE first appeared in French in 1964 and in its English translation in 1966. Malinvaud identified these two contributions to econometrics in his interview<sup>2</sup> with Alan Krueger (2003) in the *Journal of Economic Perspectives* as the geometric treatment of linear models and a rigorous treatment of nonlinear regression. These contributions and some of their implications we will overview in what follows.

## Linear Estimation

*“Algebraic proofs, which in any case are often very heavy, hardly reveal the true nature of the properties.”* Malinvaud (1966, Ch. 5)

Chapter 5 of Malinvaud provides his novel geometric exposition of the Gauss Markov (or, more strictly, the Gauss Markov Aitken) theory of linear estimation. The framework is that of a general linear space containing the observation vector  $x \in \mathbb{R}^n$ , its mean  $y = \mathbb{E}(x) \in L$ , a  $p < n$  dimensional linear subspace, and the error vector  $u = x - y$  with zero mean and positive semidefinite variance matrix  $\Omega \geq 0$ , so that

$$x = y + u, \text{ with } y \in L \subset \mathbb{R}^n. \quad (1)$$

When  $\Omega$  is singular, the support of the random vector  $u$  is  $S = \mathcal{R}(\Omega)$ , the range space of  $\Omega$ , so that every  $u = \Omega v$  a.s., for some  $v \in \mathbb{R}^n$ . It is common, but not

<sup>2</sup>See Holly and Phillips (1988) for a further interview with Edmond Malinvaud that focused on econometrics.

necessary, to assume that  $L \subset \mathcal{R}(\Omega)$ , which simplifies the theory and assures that  $x \in \mathcal{R}(\Omega)$  *a.s.* also.

The geometry underlying the Gauss Markov theory is based on the relationship of the linear space  $L$ , which is known to contain the unknown  $y$ , to the concentration ellipsoid  $E = \{u = \Omega v | u' \Omega^- u \leq 1\}$  of  $u$ , in which the quadratic form  $u' \Omega^- u$  involves a generalized inverse<sup>3</sup>,  $\Omega^-$ , of  $\Omega$ . The ellipsoid  $E$  gives a geometric representation of the variability of  $u$  (and hence  $x$ ). The concept appeared in Cramér's (1946) classic treatise of mathematical statistics<sup>4</sup>. Malinvaud provided the first systematic treatment of linear estimation using the concept in econometrics.

In simple regression problems where  $y = Z\beta$  for some matrix  $Z$  of observations on a vector of regressors  $z$ , the linear space  $L$  is the range space of  $Z$ , viz.,  $L = \mathcal{R}(Z)$ . When the covariance matrix  $\Omega$  is nonsingular, we have a conventional linear regression model in which least squares and Aitken (1934) generalized least squares estimators take on their usual algebraic forms

$$\hat{y} = Z(Z'Z)^{-1}Z'x, \text{ and } \tilde{y} = Z(Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}x,$$

with the respective orthogonal projector  $P_Z = Z(Z'Z)^{-1}Z'$  and non-orthogonal projector  ${}_{\Omega}P_Z = Z(Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}$ .

In the general linear space framework, we have similar orthogonal and non-orthogonal projections. Orthogonal projections are obtained via the decomposition of  $\mathbb{R}^n$  into the direct sum  $L \oplus L^\perp$  of  $L$  and its orthogonal complement  $L^\perp$ . If the corresponding decomposition of  $x$  is  $x = a + b$  with  $a \in L$  and  $b \in L^\perp$ , then by standard projection geometry the orthogonal projection operator  $P_L$  of  $\mathbb{R}^n$  into  $L$  maps  $x \mapsto P_L x = a$  and so the least squares estimator of  $y$  is given by  $\hat{y} = P_L x = a$ . The projection residual  $\hat{u} = x - \hat{y} = (I - P_L)x = b$  is then orthogonal to  $L$ . Since  $x, \hat{y} \in S$ , it follows that  $\hat{u} \in S$ . So, when  $L \subset S$ , the projection geometry occurs within the support  $S$ .

Non-orthogonal projections are obtained via the decomposition of  $S$  into the direct sum  $L \oplus K$  of  $L$  and its principal conjugate subspace  $K$  in  $S$  with respect to the ellipsoid  $E$  of  $x$ . The space  $K$  is defined as  $K = \{v \in S | v' \Omega^- u = 0, \forall u \in L\}$ . If  ${}_K P_L : S \mapsto L$  is the corresponding projection operator of  $S$  into  $L$  along the subspace  $K$  that is  $E$ -conjugate to  $L$  and the corresponding decomposition of  $x$  is  $x = c + d$ , with  $c \in L$  and  $d \in K$ , then the Gauss Markov (or, strictly

<sup>3</sup>The quadratic form  $u' \Omega^- u$  (and hence  $E$ ) is invariant to the choice of generalized inverse because  $u = \Omega v \in \mathcal{R}(\Omega)$  *a.s.*, for some vector  $v$ , where  $\mathcal{R}(\cdot)$  is the range space of the argument matrix, so that  $u' \Omega^- u = v' \Omega \Omega^- \Omega v = v' \Omega v$  and then  $E = \{\Omega v | v' \Omega v \leq 1, v \in \mathbb{R}^n\}$ . In the second and subsequent editions of MSE, Malinvaud found it convenient to use the nonsingular generalized inverse  $\Omega^- = (\Omega + RR')^{-1}$  where  $R$  is a matrix of full column rank  $r$  that spans the null space  $\mathcal{N}(\Omega)$  of  $\Omega$  where  $r = \dim\{\mathcal{N}(\Omega)\}$ . In this case,  $E$  has the alternate representation  $E = \{u | u'R = 0, u'(\Omega + RR')^{-1}u \leq 1\}$ . A fact of some interest, not noted in Malinvaud, is that the Moore-Penrose inverse of  $\Omega$  can be written in the convenient form  $\Omega^+ = (\Omega + RR')^{-1} - R(R'R)^{-2}R'$ . The definition of  $E$  given in the text was suggested in Phillips (1979) and Philoche (1971). See also Drygas (1970) and Nordstrom (1991) for further discussion of the coordinate free approach and concentration ellipsoids.

<sup>4</sup>The concept was independently introduced by Darmois (1945) around the same time.

speaking, Aitken) estimator of  $y$  is given by  $\tilde{y} = {}_K P_L x = c$ . The residual in this event is  $\tilde{u} = x - \tilde{y} = (I - {}_K P_L)x = d$ , which is conjugate to  $L$  with respect to the ellipsoid  $E$ .

Linear mappings such as  $P_L$  and  ${}_K P_L$  take the concentration ellipsoid  $E$  of  $x$  into the corresponding concentration ellipsoids of  $P_L x$  and  ${}_K P_L x$ . It is therefore immediately apparent that the smallest possible concentration ellipsoid that can be so obtained by projection is just  $E \cap L$ . By immediate consequence of the direction  $K$  of the projection  ${}_K P_L$ , this minimum  $E \cap L$  is achieved by  ${}_K P_L x$ , so that  ${}_K P_L x E = E \cap L$  is the concentration ellipsoid of  ${}_K P_L x$ , which gives the minimum variance unbiased estimator of  $y \in L$ . This elegant argument neatly and rigorously establishes the Gauss Markov theorem.

The geometry of the Gauss Markov theorem is illustrated in Figure 1 for  $n = 2, p = 1, S = \mathbb{R}^2, \Omega = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} > 0$  for some  $|\rho| < 1$ , and the linear manifold  $L = \{\lambda a | \lambda \in \mathbb{R}\}$  for some fixed vector  $a$  distinct from the minor and major axes of the ellipsoid  $E = \{u \in \mathbb{R}^2 | u' \Omega^{-1} u \leq 1\}$ . The figure shows the ellipsoid  $E$ , the linear manifold  $L$ , and the principal  $E$ -conjugate subspace  $K$ . The space  $K$  is determined by the direction vector  $du$  of the tangent to the ellipse  $u' \Omega^{-1} u = 1$  at points  $u \in L$  on  $E$  where the manifold  $L$  intersects the ellipse and so this vector satisfies the first order condition  $du' \Omega^{-1} u = 0$  for tangency.

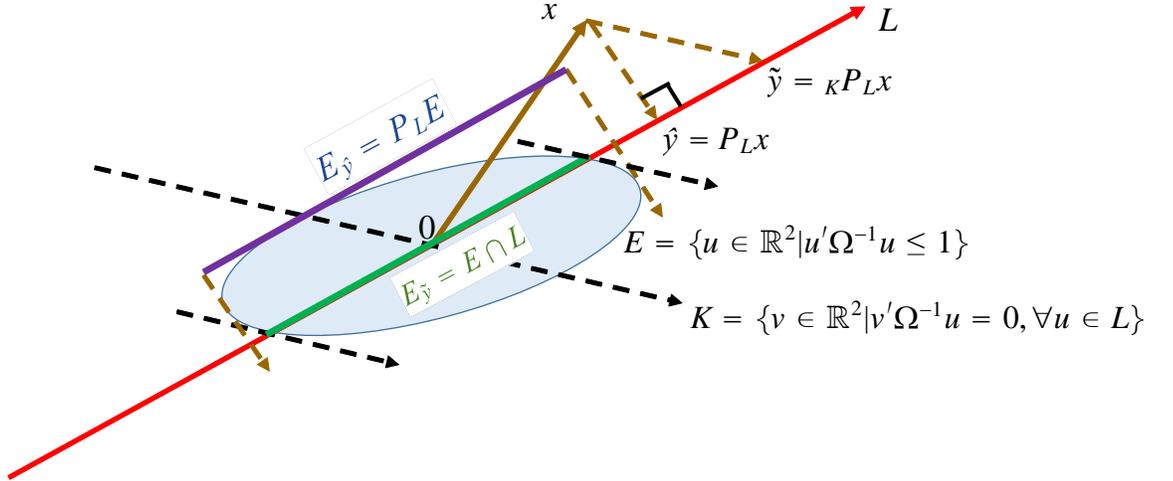


Figure 1: Least squares geometry showing the ellipsoid  $E$ , the principal conjugate subspace  $K$  of  $L$ , the corresponding projection  ${}_K P_L x$ , the least squares projection  $P_L x$ , and the ellipsoids of these two estimators.

Figure 1 shows how the minimum concentration ellipsoid  $E \cap L$  (colored green in the figure) of any estimator that projects the observation vector  $x$  onto  $L$  is obtained by projecting  $x$  in the specific direction  $K$  which collapses  $E$  parallel to the tangent spaces at the points where  $L$  intersects the ellipse

$u'\Omega^{-1}u = 1$ . By contrast the least squares estimator  $P_Lx$  is obtained by an orthogonal projection and has concentration ellipsoid  $P_LE$  (colored purple in the figure) which is obtained by projecting  $E$  orthogonally onto  $L$ , which produces a region that is always at least as great as the minimum  $E \cap L$ .

This geometric representation of the least squares (LS) estimator, the generalized least squares (GLS, Aitken) estimator, and the minimum concentration ellipsoid  $E \cap L$  of linear unbiased estimation provides an elegant and rigorous proof of the Gauss Markov Aitken theory.

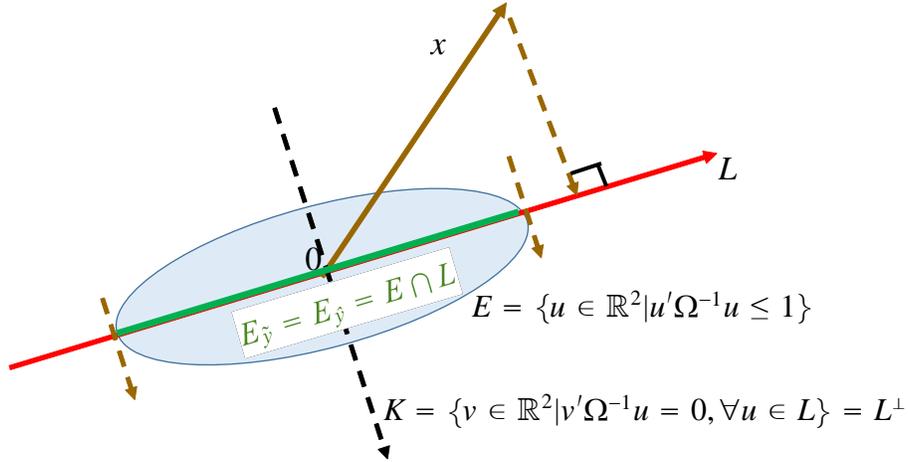


Figure 2: Least squares geometry when  $L = \Omega \times L$  and LS=GLS.

Although Malinvaud did not do so, an extension of this geometric argument proves the Kruskal (1968) theorem that least squares and generalized least squares are equivalent iff the space  $L$  is invariant under the mapping  $\Omega$ . It is convenient but not essential to assume that  $\Omega > 0$ . Sufficiency is then immediate. If  $v \in K$  we have  $v'\Omega^{-1}u = 0$  for all  $u \in L$ , and if  $L = \Omega \times L$  then take  $u = \Omega b$  for any  $b \in L$ , so that  $v'b = 0$ , giving  $K = L^\perp$ . The condition is necessary because equivalence of least squares and generalized least squares implies the conjugate subspace  $K = L^\perp$ . Hence, for all  $u \in E \cap L$ , the tangent vector  $du \in K = L^\perp$  to  $E$  that satisfies  $du'\Omega^{-1}u = 0$  also satisfies  $du'u = 0$ , so that  $du$  is orthogonal to  $u$ . It follows that  $u$  lies in the span of  $p$  of the principal axes of the ellipsoid  $u'\Omega^{-1}u = 1$ , which are defined by the property that the tangent to the ellipse at  $u$  is orthogonal to  $u$ , a property that holds when  $u$  is an eigenvector of  $\Omega$ . Let  $U = [u_1, \dots, u_p]$  be an orthonormal span of  $L$ . Then  $\Omega^{-1}U = UM$  for some diagonal matrix  $M = \text{diag}(\mu_1, \dots, \mu_p)$ , from which it follows that  $L = \mathcal{R}(U) = \mathcal{R}(\Omega U) = \Omega \times L$ . In Figure 1, simply rotate  $L$  to align with the major (or minor) axis of the ellipse and the result is immediate.

As indicated at the outset, it is not necessary to assume that  $L \subset \mathcal{R}(\Omega)$ . If  $L \not\subset \mathcal{R}(\Omega)$  we may decompose the space  $L$  into the direct sum  $L = L_1 \oplus L_2$  where  $L_1 = L \cap \mathcal{R}(\Omega)$  and  $L_2 = L \cap \mathcal{R}(\Omega)^\perp$ . Now  $\mathcal{R}(\Omega)^\perp = \mathcal{N}(\Omega) = \mathcal{R}(R)$  where  $\Omega R = 0$ , as above, so we may project the model  $x = y + u$  onto  $\mathcal{R}(\Omega)$  and  $\mathcal{R}(\Omega)^\perp$

using the transformations  $Q_R = I - R(R'R)^{-1}R'$  and  $P_R = R(R'R)^{-1}R'$ , leading to the system

$$x_1 = y_1 + u, \text{ with } x_1 = Q_R x \text{ and } y_1 = Q_R y, \quad (2)$$

$$x_2 = y_2, \text{ a.s., with } x_2 = P_R x \text{ and } y_2 = P_R y. \quad (3)$$

The component  $y_2$  of  $y = y_1 + y_2$  is known almost surely from the observation  $x$  since the (known) projection  $x_2 = P_R x = y_2$ , *a.s.* Hence, part (3) of the system is a known equation relating  $y_2$  directly to the data. The primary part (2) of this system falls into the framework analyzed above because  $y_2 \in Q_R \times L = L \cap \mathcal{R}(\Omega) \subset S$ .

On the other hand, in a regression context where  $L = \mathcal{R}(Z)$  for some matrix of observations of regressors and  $y = Z\beta$  for some  $p$ -vector of parameters  $\beta$ , the component (3) leads to the linear restriction  $R'x = R'Z\beta$ , *a.s.*, which will involve a restriction on the parameters of  $\beta$  when  $R'Z \neq 0$ . The best linear unbiased estimator of  $\beta$  then needs to take account of these restrictions. This case is easily accommodated in practice by using a restricted version of generalized least squares.

With this background of geometric theory, Malinvaud goes on to study estimable functions in unidentified cases, practical considerations of computation, algebraic formulae, estimation under Gaussian assumptions, and a selection of illustrative examples from econometrics. Hypothesis testing is also conducted using a geometric approach by considering a linear hypothesis  $\mathbb{H}_0 : y \in M \subset L$  where  $M$  is a linear subspace of  $L$ . Likelihood ratio tests are derived under Gaussianity and for various cases depending on the extent of the a priori knowledge concerning the covariance matrix  $\Omega = \omega^2\Omega_0$ . The central and non-central chi-square and F distribution theory is given under both null and fixed alternatives. The chapter ends with a discussion of statistical decision theory, admissability, and non-linear shrinkage estimation of the Stein-James type. Chapter 6 provides a very detailed implementation of this general theory of linear estimation in the context of multivariate regression. Later chapters extend the analysis further by considering linear multivariate models with cross equation parameter restrictions that include seemingly unrelated regressions as a special case.

The scope of this treatment of linear estimation is at once innovative and comprehensive. It covers substantial ground, is succinct yet eminently well exposed, and takes readers directly to the frontier of research in this field as it stood at the end of the 1960s. As Malinvaud intimates in the header to this section, the geometric approach helps to reveal the essence of least squares estimation at a level of generality that avoids much of the heavy lifting required in a purely algebraic treatment.

## Nonlinear Regression

Chapter 9 of Malinvaud is devoted to estimation and inference in multivariate nonlinear regression models belonging to the reduced form variety

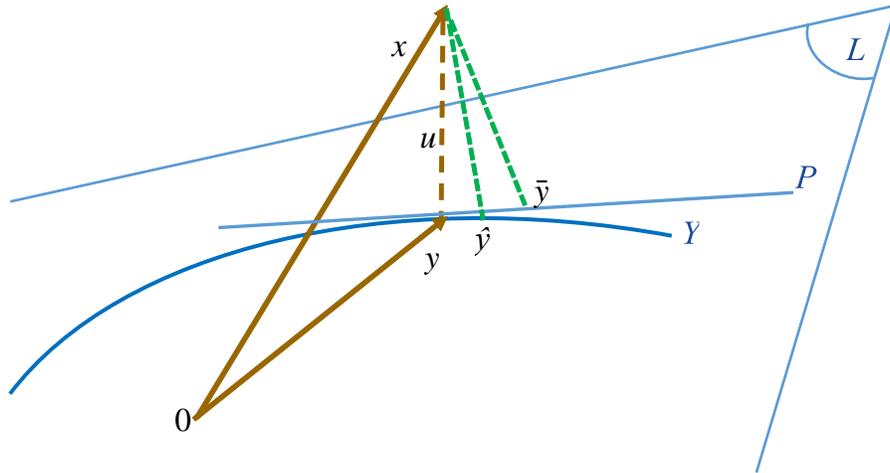
$$x_t = g(z_t; \theta) + u_t, \quad u_t \sim_{iid} (0, \Omega), \quad \Omega > 0, \quad t = 1, \dots, T, \quad (4)$$

where  $g(z_t; \theta)$  is a vector of  $n$  nonlinear functions  $g_i$  of  $m$  dimensional exogenous or predetermined variables  $z_t$  and a  $p$  dimensional parameter vector  $\theta \in \Theta \subset \mathbb{R}^p$ . Writing  $g_t(\theta) = g(z_t; \theta)$  and stacking the  $T$  observations we may write (4) in the same form as the linear system (1), namely as  $x = y + u$ , but now the mean vector  $y$  is restricted to lie in a nonlinear manifold  $Y$  determined by the functional form of  $g(\theta) = (g_1(\theta), \dots, g_T(\theta))$ . Figure 3 gives the geometric configuration.

The system (4) includes a large number of important special cases, enough to cover most of the remaining econometric models considered in MSE<sup>5</sup>. A particularly important member of this class is a multivariate system that is linear in variables but nonlinear in parameters, notably the model

$$x_t = A(\theta) z_t + u_t, \quad (5)$$

where  $A(\theta)$  is a matrix whose elements  $a_{ij}(\theta)$  depend on  $\theta$ . The model (5) itself includes the special class of structural systems of linear simultaneous equations of the form  $B(\theta) x_t = C(\theta) z_t + u_t$ , where the coefficient matrices  $(B(\theta), C(\theta))$  have structural elements that depend on a subset of parameters contained in  $\theta$ . In this case, the coefficient matrix  $A(\theta) = B(\theta)^{-1} C(\theta)$ , so that (5) is simply the reduced form of the simultaneous equations system.



<sup>5</sup>In the fourth edition, Malinvaud (1981) included a new chapter dealing with the added complication of systems that are nonlinear in both the endogenous variables  $x_t$  and the exogenous variables  $z_t$ . This chapter embodied the research conducted in the 1970s on such systems, which were important in practical work with macroeconomic models where such nonlinearities frequently arise naturally from the presence of linear in levels identities coupled with linear behavioral equations formulated in logarithms.

Figure 3: The Geometry of Nonlinear Regression

Importantly, (5) is embedded in the linear class (1) whose mean vector  $y \in L = \mathcal{R}(Z)$  with  $Z = [Z'_1, \dots, Z'_T]'$  and  $Z_t = (I \otimes z'_t)$ , since we may write the model in the following form

$$x_t = (I \otimes z'_t) a(\theta) + u_t = Z_t a(\theta) + u_t,$$

using row vectorization  $a(\theta) = \text{vec}(A(\theta))$  of the coefficient matrix and stacking columns so that  $x = y + u$  with  $y = Za(\theta) \in Y \subset L$ . The nonlinear manifold  $Y$  then lies within the linear subspace  $L$ , as shown in Figure 3.

Malinvaud's novel contributions to the theory of nonlinear regression were to furnish a rigorous proof of the consistency of nonlinear regression under high level conditions, to provide a set of primitive conditions that assured their validity, develop a rigorous limit distribution theory suitable for inference, and provide a large number of examples relevant to econometric work. The first full development of this subject broke new ground in mathematical statistics and appeared in Malinvaud's original French (1964) edition, which was followed by a major article (1970) in the *Annals of Mathematical Statistics* that was largely concerned with conditions for consistency. More complete treatments using the 1970 paper appeared in later editions of MSE.

Both minimum distance estimators and Gaussian maximum likelihood estimators were considered. The minimum distance estimator of  $\theta$  was viewed as an extended version of least squares and solved the following extremum problem

$$\hat{\theta} = \arg \min_{\theta \in \Theta} \sum_{t=1}^T (x_t - g_t(\theta))' S_T (x_t - g_t(\theta)) \quad (6)$$

where  $S_T > 0$  is a weight matrix for which  $S_T \rightarrow_p S > 0$  as  $T \rightarrow \infty$ . Malinvaud's proof of the consistency of  $\hat{\theta}_T$  used a lemma that relied on two high level conditions involving the quantities

$$\begin{aligned} Q_T(\theta) &= \sum_{t=1}^T (g_t(\theta) - g_t(\theta^0))' S_T (g_t(\theta) - g_t(\theta^0)), \\ U_T(\theta) &= \frac{\sum_{t=1}^T u'_t S_T \{g_t(\theta) - g_t(\theta^0)\}}{Q_T(\theta)}, \end{aligned}$$

where  $\theta^0$  denotes the true value of the parameter  $\theta$  in (4). In view of the extremum property (6) and the fact that  $x_t = g_t(\theta^0) + u_t$ , the following inequality holds

$$\sum_{t=1}^T u'_t S_T u_t \geq \sum_{t=1}^T (x_t - g_t(\hat{\theta}))' S_T (x_t - g_t(\hat{\theta})) = \sum_{t=1}^T u'_t S_T u_t + Q_T(\hat{\theta}) [1 - 2U_T(\hat{\theta})],$$

which implies that

$$Q_T(\hat{\theta}) [2U_T(\hat{\theta}) - 1] \geq 0, \quad (7)$$

an inequality that in turn implies either  $Q_T(\hat{\theta}) = 0$  or  $U_T(\hat{\theta}) \geq \frac{1}{2}$ . If the following two conditions hold as  $T \rightarrow \infty$

$$(i) \mathbb{P} \left\{ \inf_{\theta \in \omega} Q_T(\theta) = 0 \right\} \rightarrow 0, \quad \text{and (ii) } \mathbb{P} \left\{ \sup_{\theta \in \omega} U_T(\theta) \geq \frac{1}{2} \right\} \rightarrow 0 \quad (8)$$

for any closed set  $\omega \subset \Theta \subset \mathbb{R}^p$  that does not contain  $\theta^0$ , then  $\hat{\theta} \rightarrow_p \theta^0$ . The reason is simply that either  $Q_T(\hat{\theta}) = 0$  or  $U_T(\hat{\theta}) \geq \frac{1}{2}$  must hold, so that the event  $\{\hat{\theta} \in \omega\}$  implies either  $\inf_{\theta \in \omega} Q_T(\theta) = 0$  or  $\sup_{\theta \in \omega} U_T(\theta) \geq \frac{1}{2}$ . It follows that

$$\mathbb{P} \left\{ \hat{\theta} \in \omega \right\} \leq \mathbb{P} \left\{ \inf_{\theta \in \omega} Q_T(\theta) = 0 \right\} + \mathbb{P} \left\{ \sup_{\theta \in \omega} U_T(\theta) \geq \frac{1}{2} \right\} \rightarrow 0$$

which ensures that  $\hat{\theta} \rightarrow_p \theta^0$ .

Modern proofs of consistency in nonlinear extremum estimation problems rely on similar arguments but typically employ uniform strong law of large number methods.<sup>6</sup> For instance, we may replace  $Q_T(\theta)$  by the standardized quantity  $\bar{Q}_T(\theta)$  and require that  $\bar{Q}_T(\theta) \rightarrow_{a.s.} \bar{Q}(\theta)$ , uniformly over  $\theta \in \Theta$ , where the limit function  $\bar{Q}(\theta)$  is continuous and satisfies the identification condition that  $\bar{Q}(\theta) > 0$  for all  $\theta \neq \theta^0$ . Then  $\inf_{\theta \in \omega} \bar{Q}_T(\theta) \rightarrow \inf_{\theta \in \omega} \bar{Q}(\theta) > 0$  and (i) holds. Similarly, by a uniform strong law, continuous mapping, and  $\inf_{\theta \in \omega} \bar{Q}(\theta) > 0$  we have, uniformly in  $\theta \in \omega$ ,

$$U_T(\theta) = \frac{\frac{1}{T} \sum_{t=1}^T u_t' S_T \{g_t(\theta) - g_t(\theta^0)\}}{\bar{Q}_T(\theta)} \rightarrow_{a.s.} 0,$$

giving (ii).

An important strength of Malinvaud's approach is that it does not directly rely on uniform laws of large number arguments or use any particular normalization. The conditions (i) and (ii) may therefore be used in the context of observations where nonstationarity is present or trends occur in the time series. Indeed, Malinvaud's (1970) article and MSE give various examples, one involving an evaporating trend model for which the conditions fail and there is no consistent estimator. Recent research in this field now includes nonstationary regression models such as (4) in which the regressors are variables with unit roots (Park and Phillips, 2001) and may even be nonparametric functions of stochastic trends (Wang and Phillips, 2009).

With consistency in hand, Malinvaud provided a limit distribution theory for nonlinear regression estimators such as  $\hat{\theta}$  that gave a rigorous basis for inference. Continuing the geometric approach, Malinvaud used a linear pseudo-model approximation to (4) which enabled an elegant derivation of the asymptotic theory and provided links with the earlier work on linear estimation.

<sup>6</sup>This work began systematically with Jennrich (1969), which appeared around the same time as Malinvaud's (1970) article, and was taken further by Wu (1981) before substantial additional work appeared in the econometric literature on extremum estimation.

The approach is based on the idea that in an immediate neighborhood of the true value  $\theta^0$ , the nonlinear variety  $Y$  is approximately a linear manifold determined by the tangent space at  $\theta^0$ . The form of the tangent space is a  $p$  dimensional hyperplane and its analytic form follows from the first order Taylor expansion of the nonlinear function

$$g(z_t; \theta) \sim g(z_t; \theta^0) + Z_t(\theta - \theta^0), \text{ with } Z_t = \frac{\partial g(z_t; \theta^0)}{\partial \theta}.$$

Setting  $\check{x}_t = x_t - g(z_t; \theta^0) + Z_t\theta^0$ , we have the approximate linear pseudo-model

$$\check{x}_t = Z_t\theta + u_t, \tag{9}$$

which falls into the linear model class explored in Chapter 5. Then, stacking the system, we have  $\check{x} = Z\theta + u$  and  $\mathbb{E}(\check{x}) \sim Z\theta \in \mathcal{R}(Z) = P$ , as shown in Figure 3. Since  $\hat{\theta} \rightarrow_p \theta^0$ , the linear pseudo model representation is accurate asymptotically with a small enough error to ensure that the limit distribution of  $\hat{\theta}$  follows directly from that of the corresponding (infeasible) linear estimator

$$\check{\theta} = \left( \sum_{t=1}^T Z_t S_T Z_t' \right)^{-1} \left( \sum_{t=1}^T Z_t S_T \check{x}_t \right) = \arg \min_{\theta \in \Theta} \sum_{t=1}^T (\check{x}_t - Z_t\theta)' S_T (\check{x}_t - Z_t\theta)$$

of  $\theta$  in (9). According to the linear estimation theory, the mean of  $x$  in the pseudo model (9) lies in the hyperplane  $P$  in Figure 3, and its value is estimated by the projection,  $\check{y}$ , of  $x$  on this hyperplane, giving  $\check{y} = y + Z(\check{\theta} - \theta^0)$ . In the nonlinear model (4), the mean of  $x$  is  $y = g(\theta^0)$ , which is estimated by the projection,  $\hat{y}$ , of  $x$  onto the nonlinear variety  $Y$ . In large samples we can expect, due to the consistency of  $\hat{\theta}$  (and  $\check{\theta}$ ) that  $\check{y}$  will be close to  $\hat{y}$ . The geometric configuration is shown in Figure 3.

Under some regularity conditions, Malinvaud shows that

$$\sqrt{T}(\check{\theta} - \theta^0), \sqrt{T}(\hat{\theta} - \theta^0) \rightarrow_d N\left(0, M(S)^{-1} M(S\Omega S) M(S)^{-1}\right),$$

where

$$M(S) = \text{p} \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T Z_t S_T Z_t', \quad M(S\Omega S) = \text{p} \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T Z_t S_T \Omega S_T Z_t',$$

involving what is now commonly called the sandwich formula. Well known arguments then show that, under certain additional conditions, the asymptotically efficient estimator in this class involves a weight function for which  $S_T \rightarrow_p \Omega^{-1}$ , in which case the limit variance matrix is  $M(\Omega^{-1})^{-1}$ . Malinvaud shows that the Gaussian maximum likelihood estimator achieves this bound, so that appropriately constructed minimum distance estimators<sup>7</sup> are asymptotically efficient in the class of all regular consistent estimators.

<sup>7</sup>Such estimates may be constructed using a two step or iterated step procedure in which  $S_T$  is constructed from the moment matrix of residuals in an earlier step of the iteration that ensures  $S_T \rightarrow_p \Omega^{-1}$ .

With consistency and asymptotic normality results in hand for general non-linear regressions, Malinvaud systematically applies this theory to models that arise commonly in econometric applications, such as model (5) where the nonlinearities occur via analytic constraints on the coefficients of a linear model. The theory is used by Malinvaud in the later chapters of MSE to provide rigorous proofs of consistency and asymptotic normality for simultaneous equations estimators.

Chapter 9 also considers hypothesis testing, confidence interval construction in the case of analytic constraints, numerical optimization methods, and, in its final subsection, models with inequality constraints – a topic that has begun to attract considerable attention over the last decade.

This highly original chapter contains fundamental theory that lays a rigorous foundation for much subsequent theoretical work in statistics and econometrics. By reworking the limit theory of structural equation estimation in terms of his rigorous asymptotic development of nonlinear regression, Malinvaud anticipated a new generation of extremum estimation limit theory in econometrics that began to emerge in the 1980s and now dominates much econometric theory and practice.

## Conclusion

*"A thorough reading of the book demands a good mathematical background and a knowledge of general theories of probability calculus and mathematical statistics. Some chapters are particularly difficult, requiring sustained effort from the reader who seeks complete mastery of the proofs"* Malinvaud (1966, Preface)

Perhaps because of its advanced nature or its extensive use of geometric rather than purely algebraic arguments, Malinvaud's text had rather less penetration than might be expected in North American graduate programs, even in the top US schools. Instead, mechanical algebraic approaches to the development of econometrics continued as the teaching norm well into the 1970s and 1980s. Base courses in econometrics often emphasized a 'mostly harmless' approach largely to avoid intimidating students. Students emerged from these courses thinking of econometrics in terms of  $X'X$  matrices, simple models based on conditional expectations, and a variety of specialized methods useful in handling some of the challenges of panel data and simple dynamic models. Such courses often lacked unifying principles of statistical inference and probabilistic underpinnings, provided no hint of sophistication in asymptotic theory, and offered little understanding of the subtleties of finite sample econometrics. These inadequacies are amply demonstrated in other textbooks of econometrics that competed with Malinvaud's opus during this period.

By contrast, instructors and students who persevered with Malinvaud rapidly recouped the investment made in mastering his work. By the mid 1980s new textbooks in econometrics began to appear that followed Malinvaud's example

of using probabilistically sound arguments in developing asymptotic theory for econometric estimators and test statistics. But none of these adopted the mantle of Malinvaud's geometric approach that so beautifully embodied all the key results of the Gauss Markov theory, captured the essence of the asymptotic properties of nonlinear estimation, and enabled an intuitive understanding of least squares efficiency. Malinvaud's work remains distinctive in at least this respect, as well as its rigor in developing the asymptotic theory of simultaneous equations estimation and inference using principles of nonlinear regression.

Modern approaches to asymptotic statistical theory involve an ever widening array of tools, that include extremum estimation techniques, uniform laws of large numbers, locally asymptotic normal likelihood ratio methods, martingale convergence theorems, martingale central limit theory, functional limit theory, and weak convergence to processes that involve stochastic integrals. These methods, very largely derived from advances in probability calculus and statistical theory over the past half century, have now largely displaced Malinvaud's path to asymptotic theory for econometricians. Yet his geometric ideas of linear pseudo-models and linear models with analytic constraints still have wide applicability in empirical work. In the econometrician's relentless search for generality wherever possible in the development of limit theory, empirical applications in time series, microeconometrics, and panel econometrics are still heavily dependent on basic models that are often little more than inspired extensions of linear systems, where the intuition provided by Malinvaud's methods remains as useful as ever.

Malinvaud was acutely aware of the difficulties and subtleties of asymptotics and the immense intellectual challenges presented by finite sample theory. Indeed, in 1971 Malinvaud took the unusual step of writing to the Editor of the *International Economic Review* drawing attention to the absence of rigor in the asymptotic arguments that were commonly used in econometric articles and calling for higher standards in the execution of asymptotic methods in econometrics.

Since the last edition of MSE in 1981, the subject of econometrics has moved on not only in terms of basic theory but also in the vast range of its applications. These cover every area of economics, including some higher levels of economic theory, and extend well beyond the subject area of economics into the business, financial, and social sciences, the medical sciences, and natural sciences. In view of the subject's enormous extent, it is no longer possible to attempt an encompassing pedagogical work such as Malinvaud's MSE. But it remains a desirable, if somewhat elusive, goal to capture the essence of the subject of econometrics and its intimate links and foundations within the discipline of economics. Malinvaud's book with its own uniquely broad perspective still provides a useful entry point to this ever widening universe of research and it offers readers a clear view of the goalpost of econometrics as the essential vehicle for evidence-based economic analysis and policy formation.

Malinvaud did more than write a brilliant textbook with many original elements. MSE brought the econometric universe together, married linear and nonlinear estimation, system regression, panel modeling, time series, and simul-

taneous equations estimation and inference in a unified rhetoric of econometrics that remained faithful to the subject's roots and its *raison d'être* in the discipline of economics.

## A Personal Tribute

The author learnt econometrics in the 1960s at the University of Auckland.<sup>8</sup> Located, as we were, far from the major centres of learning in North America and Europe, the primary knowledge source to guide us as students in the upper reaches of a discipline came from reading advanced texts and journal articles. In this regard, Malinvaud had no peer. It defined a new equivalence class for textbooks in econometrics.

The book's content and rhetoric reminded me in terms of its reach and its exposition of Harald Cramér's classic *Mathematical Methods of Statistics*, whose title Malinvaud adroitly massaged into the *Statistical Methods of Econometrics*, and whose division into multiple component parts provided a natural skeletal framework for the content of Malinvaud's own treatise. With these advanced texts on the desk and two extraordinary teachers forever at their sides, students located far from the centres of learning in the Western hemisphere were excited and privileged to feel advantaged, and never handicapped, by the separation.

## The Final Word

The final word of this tribute must come from Malinvaud. The following two paragraphs from Malinvaud's Conclusion provide a clear directive to the econometrics profession. Its value and relevance remain undiminished in the half century since its composition.

*“The art of the econometrician consists as much in defining a good model as in finding a good statistical procedure. Indeed, this is why he cannot be purely a statistician, but must have a solid grounding in economics. Only if this is so, will he be aware of the mass of accumulated knowledge which relates to the particular question under study and must find expression in the model.*

*Finally, we must never forget that our progress in understanding economic laws depends strictly on the quality and abundance of statistical data. Nothing can take the place of the painstaking work of objective observation of the facts. All improvements in methodology would be in vain if they had to be applied to mediocre data.” Malinvaud (1966)*

---

<sup>8</sup>The interested reader may refer to the author's (2014) account of these studies.

## References

- Aitken, A.C. (1934). On least squares and linear combinations of observations. *Proceedings of the Royal Society of Edinburgh*, 55, 42–48.
- Cramér, H. (1946). *Mathematical Methods of Statistics*. Princeton University Press.
- Darmois, G. (1945). “Sur les limites de la dispersion de certaines estimations,” *Revue de l’Institut International de Statistique*, 13, 9-15.
- Drygas, H. (1970). *The Coordinate-Free Approach to Gauss-Markov Estimation*. Berlin: Springer.
- Holly, A. and P. C. B. Phillips (1987). “The ET Interview: Professor Edmond Malinvaud,” *Econometric Theory*, 3, 273-295.
- Jennrich, R.I. (1969). “Asymptotic properties of non-linear least squares estimation,” *Annals of Mathematical Statistics* 40, 633–643.
- Krueger, A. (2003). “An Interview with Edmond Malinvaud”, *Journal of Economic Perspectives*, 17, 181-198.
- Kruskal, W. (1968). “When are Gauss-Markov and least squares estimators identical? A coordinate free approach”, *Annals of Mathematical Statistics*, 39, 70-75.
- Malinvaud, E. (1964). *Méthodes Statistiques de l’Économétrie*. Paris: Dunod.
- Malinvaud, E. (1966). *The Statistical Methods of Econometrics*. Amsterdam: North Holland. (First English Edition)
- Malinvaud, E. (1970). *The Statistical Methods of Econometrics*. Amsterdam: North Holland. (Second English Edition)
- Malinvaud, E. (1970). “The consistency of nonlinear regressions,” *Annals of Mathematical Statistics*, 41, 956–969.
- Malinvaud, E. (1971). “Letter to the Editor,” *International Economic Review*, 12, 344-345.
- Malinvaud, E. (1980). *Statistical Methods of Econometrics* . Amsterdam: North-Holland. (Third English Edition)
- Malinvaud, E. (1981). *Méthodes Statistiques de l’Économétrie*. Paris: Dunod. (Quatrième Édition).
- Park, J.Y. and P. C. B. Phillips (2001). “Nonlinear regressions with integrated time series,” *Econometrica*, 69, 117-161.
- Phillips, P. C. B. (1979), “The concentration ellipsoid of a random vector,” *Journal of Econometrics*, 11, 363-365.

- Phillips, P. C. B. (2014). "Unit Roots in Life: A Graduate Student Story," *Econometric Theory*, 30, 2014, 719-736.
- Phillips, P. C. B. (2015). "Memorial to Edmond Malinvaud," *Econometric Theory*, 31 (to appear).
- Philoche, J.-L. (1971). A propos du théorème de Gauss-Markov. *Annales de l'Institut Henri Poincaré, Section B: Calcul des Probabilités et Statistique*, VII, 271-281.
- Wang, Q. and P. C. B. Phillips (2009). "Structural nonparametric cointegrating regression," *Econometrica*, 77, 1901-1948.
- Wu, C\_F. (1981). "Asymptotic theory of nonlinear least squares estimation". *Annals of Statistics* 9: 501-513.