Appendix A.4

TESTING AND IMPOSING MONOTICITY, CONVEXITY AND QUASI-CONVEXITY CONSTRAINTS

LAWRENCE J. LAU*

Stanford University

1. Introduction

1.1. *Statement of the Problem*

In many areas of economic analysis, functions are frequently assumed to be monotonic, convex or quasi-convex.[1] Production, profit, and utility functions are obvious examples. The natural questions that arise are first, whether one can test the hypotheses of monotonicity, convexity, or quasi-convexity of these functions statistically; and second, whether one can estimate these functions subject to monotonicity, convexity or quasi-convexity constraints.

---

*The author wishes to thank T. Amemiya, E.R. Berndt, W.E. Diewert, A.S. Goldberger, Z. Griliches, G. Hanoch, J.A. Hausman, D.W. Jorgenson, D.L. McFadden, F.C. Nold, I. Olkin, and W.H. Yang for helpful discussions. This research was supported in part by the National Science Foundation, through Grant GS-40104 at the Institute of Mathematical Studies in the Social Sciences, Stanford University and the John Simon Guggenheim Memorial Foundation. It was essentially completed during the author's tenure as a Visiting Scholar in the Department of Economics, Massachusetts Institute of Technology, 1973–74. He retains sole responsibility for any remaining errors. An earlier version of this paper was presented at the North American Winter Meeting of the Econometric Society, New York, December 1973, and issued as Technical Report no. 123, "Econometrics of monotonicity, convexity and quasi-convexity", Institute of Mathematical Studies in the Social Sciences, Stanford University, 1974.

[1] Since the negative of a convex function is concave, and the negative of a quasi-convex function is quasi-concave, all the statements made about convexity and quasi-convexity apply to concavity and quasi-concavity as well. Concavity and quasi-concavity will not be separately considered.

In the past, empirical estimation of the parameters of these functions have been limited to those of rather simple algebraic form for which the constraints of monotonicity, convexity or quasi-convexity are either automatically or readily satisfied or can be easily imposed. For example, the linear function is always convex (and concave); the Cobb–Douglas production function estimated by the factor shares method is always monotonic and concave;[2] and, more generally, estimated Cobb–Douglas production functions are automatically quasi-concave if they satisfy the monotonicity conditions. Thus there has not been any pressing need for the development of techniques to test or impose the hypotheses of monotonicity, convexity or quasi-convexity.

However, two recent developments in empirical economic analysis have made it necessary to confront the dual problems of hypothesis testing and constrained estimation. First, partly because of advances in computational technology, partly because of substantial improvements in the quality and quantity of economic data, and partly because of a general dissatisfaction with the restrictive implications of the simple functional forms, there is a proliferation of new algebraic forms in empirical work in recent years. We shall name only a few which are capable of providing a second-order numerical approximation to an arbitrary function.[3] There is the "Transcendental Logarithmic Function", proposed by Christensen, Jorgenson, and Lau (1971, 1973, 1975),

$$\ln F(\mathbf{x}) = \alpha_0 + \boldsymbol{\alpha}' \ln \mathbf{x} + \tfrac{1}{2} \ln \mathbf{x}'\mathbf{B} \ln \mathbf{x},$$

where $\ln \mathbf{x} \equiv [\ln x_1 \ln x_2 \cdots \ln x_m]'$; the generalized version of the "Generalized Linear Function" proposed by Diewert (1971),

$$F(\mathbf{x}) = \alpha_0 + \boldsymbol{\alpha}'\mathbf{x}^{1/2} + \tfrac{1}{2}\mathbf{x}^{1/2'}\mathbf{B}\mathbf{x}^{1/2},$$

where $\mathbf{x}^{1/2} \equiv [x_1^{1/2} x_2^{1/2} \cdots x_m^{1/2}]'$; and the "Quadratic Function",

$$F(\mathbf{x}) = \alpha_0 + \boldsymbol{\alpha}'\mathbf{x} + \tfrac{1}{2}\mathbf{x}'\mathbf{B}\mathbf{x},$$

where $\mathbf{x} \equiv [x_1 x_2 \cdots x_m]'$.[4]

These functions may be employed as production, profit (and normalized profit), or utility functions. For an arbitrary set of parameters, these functions do not necessarily satisfy monotonicity, convexity or quasi-convexity conditions, either locally or globally. Hence there is a need to test or maintain these hypotheses.

---

[2] The factor shares method was first proposed by Klein (1953).

[3] For a definition of a second-order numerical approximation, see Lau (1974).

[4] It is a remarkable fact that Heady and Dillon (1961) have proposed a two-factor version of the transcendental logarithmic function (p. 205), the generalized linear function (pp. 91–92 and p. 206) and the quadratic function (pp. 88–91 and p. 205); and made empirical

Second, the increasing use of duality principles which rely heavily on convexity assumptions in empirical applications makes it mandatory that the estimated dual functions be monotonic, convex or quasi-convex.[5] While one may be willing to entertain the possibility that the production function may not be convex, the normalized profit function is always convex if the output and input markets are competitive under the assumption of profit maximization.[6] Thus a non-convex normalized profit function is inconsistent with profit maximization – the basic behavioral postulate of the theory of production. Likewise an indirect utility function is always quasi-convex by virtue of its being a maximum subject to a linear constraint. Thus, a non-quasi-convex indirect utility function is inconsistent with utility maximization – the basic behavioral postulate of the theory of consumer demand. Moreover, if the estimated normalized profit function is non-convex, or the indirect utility function is non-quasi-convex, the own and cross-price supply and demand elasticities will not have the theoretically expected signs and magnitudes. Thus one should at least test the hypotheses of monotonicity, convexity, or quasi-convexity; and if one does not reject these hypotheses, impose the corresponding constraints on the estimators so as to obtain economically meaningful estimates in practical applications.

1.2. *Historical Review*

There are two principal approaches to the testing of monotonicity, convexity and quasi-convexity – the parametric approach and the non-parametric approach.

There is little previous work in the parametric approach. Judge and

---

application with the latter two functions. For other scholars who have independently proposed the transcendental logarithmic functions, see the references listed in Christensen, Jorgenson and Lau (1973). Lau (1974) appears to be the first to propose the quadratic function as a normalized profit function.

[5] For an excellent survey of applications of duality theory, see Diewert (1974a); see also the comments by Lau (1974). Jorgenson and Lau (1974a, 1974b) give an exhaustive treatment of the role of convexity in production theory.

[6] McFadden (1966) appears to be the first person to emphasize this point. The concept of a normalized profit function is introduced by Lau (1969c). See also Jorgenson and Lau (1974a, 1974b).

Takayama (1966) analyze the question of inequality constraints in regression analysis and propose an estimator based on the solution of a constrained quadratic programming problem. Liew (1976) gives another algorithm for the computation of inequality constrained least-squares estimators.

Hudson (1969) proposes a method for fitting a polynomial in x such that it is convex in a closed interval. However, the computations are not fully worked out for all cases; the function to be estimated is restricted to be defined on a subset of R; and there is no distribution theory for the estimators.

In a previous version of this paper, a method for testing and maintaining the hypothesis of convexity of an estimated function based on the eigenvalue decomposition was proposed. The method based on the eigenvalue decomposition, however, does not reduce to an unconstrained minimization problem because the requirement of orthonormality of the eigenvectors makes necessary the construction of a set of orthonormal vectors and hence a great deal of computation, although it involves no conceptual difficulties. The present method based on the Cholesky factorization requires much less computation.

The previous work in the non-parametric approach is somewhat more numerous. Hildreth (1954) is the pioneer of this approach: assuming no algebraic functional form, he proposes to estimate the values of a function $F(x)$ at given values of x such that $F(x)$ is concave. This approach is extended by Dent (1973), who proposes to approximate $F(x)$ by polygonal segmentation.

Hanoch and Rothschild (1972) provide algorithms for testing monotonicity and quasi-convexity without assuming a specific algebraic form of the function. Afriat (1967, 1968, 1972) and Diewert (1974d) give alternative methods for estimating non-parametric functions which satisfy the assumptions of monotonicity, convexity, and homogeneity.

1.3. *Proposed Solution*

The basic technique consists of a transformation of parameters constrained to be non-negative into the squares of arbitrary real parameters and may be referred to as the "method of squaring". This technique appears to have been introduced by Valentine (1937), in connection with the solution of problems of calculus of variations subject to inequality constraints.[7] Thus, if a parameter B is required to be non-negative, it can be transformed into the parameter B^{*2}, and the estimation problem becomes that of choosing a B^* such that an appropriate sum of squares of residuals is minimized.

Further variations of this theme are possible. For example, one can substitute for any positive parameter B by e^{B^*}; any parameter lying between zero and one by the substitution $B = \frac{1}{2}(1 + \sin B^*)$,[8] or alternatively by

$$B = 1/(1 + e^{-B^*}).^9$$

By transformations similar to these one can suitably restrict a parameter to be within any prescribed interval. The advantage of this transformation is that it reduces the likelihood maximization problem to an *unconstrained* nonlinear least-squares problem.

The method of squaring thus provides a straightforward solution to the problem of monotonicity or for that matter any inequality constraint. The solution to the problem of convexity and quasi-convexity makes use of the properties of the Hessian matrices of convex and quasi-convex functions. A twice differentiable real-valued function is convex on an open convex set if and only if the Hessian is positive semidefinite everywhere on the open convex set. A twice differentiable real-valued function is quasi-convex if and only if the Hessian is positive semidefinite on any set of vectors $\mathbf{y}$ such that $\nabla F(\mathbf{x})\cdot\mathbf{y} = 0$, where $\nabla F(\mathbf{x})$ is the non-zero gradient of $F(\mathbf{x})$. The task of this paper is to transform these conditions on the parameters into simple non-negativity conditions by a suitable reparametrization. The basis of the reparametrization is the Cholesky factorization of real symmetric matrices. Through this factorization, the determinantal conditions of positive semidefiniteness are transformed into non-negativity constraints. Once more, the method of squaring may be employed to convert the likelihood maximization problem into an unconstrained nonlinear least-squares problem.

Our method has been employed by Jorgenson and Lau (forthcoming) in the analysis of production and by Barten and Geyskens (1975) in the analysis of consumer demand. Jorgenson and Lau (1975b) have developed an alternative procedure for testing and imposing quasi-convexity constraints and extended it for testing monotonicity of correspondences not necessarily derivable from a single function.

---

[7] See Valentine (1937, pp. 407–409).

[8] Since the sine function is periodic, $\sin B^* = \sin(B^* + 2n\pi)$, where n is an integer. Thus any $B^* + 2n\pi$, n being an arbitrary integer, results in the same B.

[9] Alternatively, one may use the transformation $B = \sin^2 B^*$. The possibilities are limitless.

2. Hessian Matrices of Convex and Quasi-Convex Functions

2.1. *Introduction*

In this section we discuss the properties of the Hessian matrices of twice differentiable real-valued convex and quasi-convex functions. We show that these Hessian matrices may indeed be represented by positive semidefinite matrices, which as we shall show in Section 3 have convenient factorization properties which facilitate the solution of the problems of testing and constrained estimation under the hypothesis of convexity or quasi-convexity.

In Section 2.2, the Hessian matrix of a convex function is characterized, and in Section 2.3, the Hessian matrix of a quasi-convex function is characterized. In Section 2.4, we examine the Hessian matrices of approximating functions under the hypotheses of convexity and quasi-convexity.

2.2. *The Hessian of a Convex Function*

Theorem 2.1. A twice differentiable real-valued function defined on an open convex set C is convex if and only if the Hessian matrix is positive semidefinite everywhere on C.

This theorem is well known. A proof may be found in Rockafellar (1970).[10]

2.3. *The Hessian of a Quasi-convex Function*

Definition. A real-valued function $F(\mathbf{x})$ defined on a convex set C is *quasi-convex* if

$$F(\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) \leqq \max(F(\mathbf{x}_1), F(\mathbf{x}_2)), \qquad 0 \leqq \lambda \leqq 1,$$

for all $\mathbf{x}_1$, $\mathbf{x}_2$ in C.

---

[10] See Rockafellar (1970, p. 27).

We specialize our attention to consider only the class of twice differentiable functions with everywhere non-zero first partial derivatives.

> *Theorem* 2.2. A twice differentiable real-valued function $F(\mathbf{x})$ defined on an open convex set C with everywhere non-zero first partial derivatives is quasi-convex if and only if for all $\mathbf{x}$ in C, $\mathbf{y}'\mathbf{H}(\mathbf{x})\mathbf{y} \geqq 0$ whenever $\nabla F(\mathbf{x})'\mathbf{y} = 0$, where $\mathbf{H}(\mathbf{x})$ and $\nabla F(\mathbf{x})$ are respectively the Hessian matrix and the gradient of the function $F(\mathbf{x})$.

A version of this theorem is proved by Katzner (1970).[11] We omit the proof. Diewert (1973b) has derived a similar theorem under the weaker condition that not all of the components of $\nabla F(\mathbf{x})$ are zero.

> *Theorem* 2.3. A necessary and sufficient condition that $\mathbf{x}'\mathbf{A}\mathbf{x} \geqq 0$ whenever $\mathbf{a}'\mathbf{x} = 0$ is that $\mathbf{x}'(\mathbf{A} + \lambda\mathbf{a}\mathbf{a}')\mathbf{x} \geqq 0$ for all $\mathbf{x}$ for all sufficiently large positive scalar constants λ.

Proof: Necessity is proved by contradiction. Suppose the theorem is false, then there exists $\mathbf{x}$ such that

$$\mathbf{x}'(\mathbf{A} + \lambda\mathbf{a}\mathbf{a}')\mathbf{x} < 0,$$

for all sufficiently large positive scalar constant λ. If $\mathbf{a}'\mathbf{x} \neq 0$, this implies that $\mathbf{x}'\mathbf{A}\mathbf{x}$ is not only negative but unbounded, which is not possible. Thus, $\mathbf{a}'\mathbf{x} = 0$. However, this implies $\mathbf{x}'\mathbf{A}\mathbf{x} < 0$, a contradiction.

Sufficiency follows in a straightforward manner. If $\mathbf{x}'(\mathbf{A} + \lambda\mathbf{a}\mathbf{a}')\mathbf{x} \geqq 0$ for all $\mathbf{x}$ for some positive scalar constant λ, then $\mathbf{x}'\mathbf{A}\mathbf{x} + \lambda\mathbf{x}'\mathbf{a}\mathbf{a}'\mathbf{x} \geqq 0$ for all $\mathbf{x}$. In particular, this holds whenever $\mathbf{a}'\mathbf{x} = 0$, in which case

$$\mathbf{x}'\mathbf{A}\mathbf{x} + \lambda\mathbf{x}'\mathbf{a}\mathbf{a}'\mathbf{x} = \mathbf{x}'\mathbf{A}\mathbf{x} \geqq 0. \qquad \text{Q.E.D.}$$

The basic result here is due to Finsler (1936–37). Bellman (1970) gives another proof of Finsler's Theorem.[12] The proof here is somewhat different from Bellman's and is modified to apply to the case of positive semidefiniteness of a constrained quadratic form. For the present application $\mathbf{A}$ and $\mathbf{a}$ may be identified with the Hessian and gradient of $F(\mathbf{x})$, respectively.

---

[11] See Katzner (1970, pp. 210–211).

[12] See Bellman (1970, pp. 76–81).

> *Theorem* 2.4. A necessary condition for a twice differentiable real-valued function $F(\mathbf{x})$ defined on an open convex set C to be quasi-convex is that all the ordered principal minors of the bordered Hessian matrix be non-positive for all $\mathbf{x}$ in C.

This theorem is due to Arrow and Enthoven (1961).[13] We omit the proof.

We now show that indeed if $\mathbf{A}+\lambda\mathbf{a}\mathbf{a}'$ is positive semidefinite for sufficiently large positive λ, then all the ordered principal minors of the matrix

$$\begin{bmatrix} 0 & \mathbf{a}' \\ \mathbf{a} & \mathbf{A} \end{bmatrix},$$

which has the interpretation of a bordered Hessian in the present application, are non-positive. Positive semidefiniteness of $\mathbf{A}+\lambda\mathbf{a}\mathbf{a}'$ implies that all the principal minors of $\mathbf{A}+\lambda\mathbf{a}\mathbf{a}'$ are non-negative. Consider the following matrix identity:[14]

$$\begin{bmatrix} -1 & \mathbf{a}' \\ \lambda\mathbf{a} & \mathbf{A} \end{bmatrix} \begin{bmatrix} 1 & \mathbf{a}' \\ 0 & \\ 0 & \\ \vdots & \mathbf{I} \\ 0 & \end{bmatrix} = \begin{bmatrix} -1 & 0 & 0\cdots 0 \\ \lambda\mathbf{a} & & \mathbf{a}+\lambda\mathbf{a}\mathbf{a}' \end{bmatrix}.$$

Taking determinants of both sides, we obtain,

$$-1|\mathbf{A}|+\lambda \begin{vmatrix} 0 & \mathbf{a}' \\ \mathbf{a} & \mathbf{A} \end{vmatrix} = -1|\mathbf{A}+\lambda\mathbf{a}\mathbf{a}'| \leqq 0,$$

or

$$-|\mathbf{A}|+\lambda \begin{vmatrix} 0 & \mathbf{a}' \\ \mathbf{a} & \mathbf{A} \end{vmatrix} \leqq 0.$$

Now this inequality must hold for all sufficiently large positive λ (in fact it can be easily shown that if $\mathbf{A}+\lambda\mathbf{a}\mathbf{a}'$ is positive semidefinite for some positive $\bar{\lambda}$, then it is positive semidefinite for all $\lambda > \bar{\lambda}$). Thus, one must have

$$\begin{vmatrix} 0 & \mathbf{a}' \\ \mathbf{a} & \mathbf{A} \end{vmatrix} \leqq 0.$$

---

[13] See Arrow and Enthoven (1961, pp. 797–799).

[14] This construction follows Bellman (1970, pp. 78–80).

The same proof applies to any principal submatrix of

$$\begin{bmatrix} 0 & \mathbf{a}' \\ \mathbf{a} & \mathbf{A} \end{bmatrix}.$$

We note that if the domain of $F(\mathbf{x})$ is restricted to the positive orthant then negativity of all the ordered principal minors of the bordered Hessian matrix except the first one everywhere is sufficient for quasi-convexity.[15]

Thus we conclude that a necessary and sufficient condition for a twice differentiable real-valued function with everywhere non-zero first partial derivatives to be quasi-convex is that the matrix

$$\mathbf{H}(\mathbf{x}) + \lambda \nabla F(\mathbf{x}) \nabla F(\mathbf{x}')$$

be positive semidefinite for all $\mathbf{x}$ for all sufficiently large positive scalar constant λ.

For the purposes of testing the hypothesis of quasi-convexity, the following theorem is more useful:

> *Theorem* 2.5. A necessary and sufficient condition that there exists a vector $\mathbf{a}$ such that $\mathbf{x}'\mathbf{A}\mathbf{x} \geqq 0$ whenever $\mathbf{a}'\mathbf{x} = 0$ is that the number of non-negative eigenvalues of $\mathbf{A}$, an $n \times n$ real symmetric matrix, must be greater than or equal to $(n-1)$.

Proof: *Necessity*. Suppose $\mathbf{A}$ has two negative eigenvalues. Let $\mathbf{v}_1$ and $\mathbf{v}_2$ be the unit eigenvectors corresponding to these two negative eigenvalues. Consider any $\mathbf{a}$: if either $\mathbf{a}'\mathbf{v}_1$ or $\mathbf{a}'\mathbf{v}_2 = 0$, then let $\mathbf{x} = \mathbf{v}_1$ or $\mathbf{v}_2$ and $\mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{v}_i'\mathbf{A}\mathbf{v}_i = \rho_i < 0$; so assume $\mathbf{a}'\mathbf{v}_1 \times \mathbf{a}_1^* \neq 0$, $\mathbf{a}'\mathbf{v}_2 = \mathbf{a}_2^* \neq 0$. Let $\mathbf{x} = a_2^*\mathbf{v}_1 - a_1^*\mathbf{v}_2$, then $\mathbf{a}'\mathbf{x} = a_2^* a_1^* - a_1^* a_2^* = 0$. But $\mathbf{x}'\mathbf{A}\mathbf{x} = a_2^*\mathbf{v}_1'\mathbf{A}\mathbf{v}_1 a_2^* + a_1^*\mathbf{v}_2'\mathbf{A}\mathbf{v}_2 a_1^* - a_2^*\mathbf{v}_1'\mathbf{a}\mathbf{v}_2 a_1^* = \rho_1 a_2^{*2} + \rho_2 a_1^{*2} < 0(\mathbf{v}_1'\mathbf{v}_2 = 0)$. Thus the number of non-negative eigenvalues of $\mathbf{A}$ must be greater than or equal to $(n-1)$.

Sufficiency. If the number of non-negative eigenvalues of $\mathbf{A}$ is $(n-1)$, and suppose $\mathbf{v}_1$ is the eigenvector corresponding to the one remaining negative eigenvalue, let $\mathbf{a} = \mathbf{v}_1$. Then whenever $\mathbf{a}'\mathbf{x} = \mathbf{v}_1'\mathbf{x} = 0$, $\mathbf{x}$ is a linear combination of the remaining $(n-1)$ eigenvectors with non-negative eigenvalues. Thus $\mathbf{x}'\mathbf{A}\mathbf{x} \geqq 0$. Q.E.D.

---

[15] See Arrow and Enthoven (1961, pp. 797–799).

The practical implication of this theorem for the quasi-convexity problem is that the Hessian of a twice differentiable quasi-convex function, $\mathbf{H}(\mathbf{x})$, must have greater than or equal to $(n-1)$ non-negative eigenvalues (or alternatively at most one negative eigenvalue) everywhere on the interior of its effective domain. In other words, in order that one can find a $\nabla F(\mathbf{x})$ such that $\mathbf{y}'\mathbf{H}(\mathbf{x})\mathbf{y} \geqq 0$ whenever $\nabla F(\mathbf{x})'\mathbf{y} = 0$, it is necessary and sufficient that $\mathbf{H}(\mathbf{x})$ has at least $(n-1)$ non-negative eigenvalues.

2.4. *Hessian Matrices of Functions Approximating Convex and Quasi-Convex Functions*

In general, economic theory itself does not provide sufficient restrictions on the functional relationships used in economic analysis so that these relationships may be represented by a single parametric class of algebraic functions.[16] The restrictions derived from economic theory are almost always of a more general type – monotonicity, convexity, quasi-convexity, homogeneity, etc., which may be satisfied by many different parametric classes of algebraic functions.

However, in applied econometrics, it is frequently necessary to estimate a function, for example, a production function, parametrically and a particular parametric form of an algebraic function must be specified. Such a functional form, inasmuch as it is not directly derivable from economic theory, should be looked upon as an approximation to the unknown underlying true function. Thus, the "translog", the generalized linear, and the quadratic functions may all be considered as alternative second-order numerical approximations to the same unknown, underlying true function.

The distinction between the approximating functions and the underlying true function is not so important, were it not for the fact that not all of the properties of the underlying true function will be inherited by the approximating function.[17] In general, even if the underlying true function is convex globally, the approximating function, while providing a good approximation, may not be globally convex or even locally convex itself.

However, if we restrict our attention to those second-order approximating functions which agree with the first and second derivatives at the point of approximation, then all approximating functions to underlying monotonic, convex, or quasi-convex functions will exhibit behavior similar to the function they are approximating at the point of approximation. In particular, the approximating function to a monotonic function will have all of its partial derivatives of one sign at the point of approximation. The approximating function to a convex function will have its Hessian matrix positive semidefinite at the point of approximation. The approximating function to a quasi-convex function will have a Hessian matrix that is positive semidefinite with respect to all vectors orthogonal to its gradient at the point of approximation.

---

[16] By contrast, consider the inverse-square law of electric potential, or Boltzman's equation in statistical mechanics.

[17] See the discussion in Lau (1975).

It should be noted that convexity of the approximating function at the point of approximation does not in general guarantee that the approximating function itself will be globally convex. It also does not guarantee global convexity of the underlying true function that is being approximated. However, non-convexity of the approximating function at the point of approximation necessarily implies non-convexity of the underlying true function. Hence from the point of view of statistical inference, one may use local convexity of the approximating function as a basis for a test.[18]

Thus these local conditions are necessary conditions in the sense that if the underlying true function were to be monotonic, convex and quasi-convex, the approximating function must exhibit corresponding properties at the point of approximation. They are therefore ideally suited for hypothesis testing.

On the other hand, these properties at the point of approximation are all that one can expect to hold for an arbitrary approximating function if one wants to impose the constraints implied by these hypotheses, rather than to test them statistically.

As has been pointed out elsewhere, one has the choice of imposing these hypotheses either globally or locally on the approximating functions.[19] Convexity at the point of approximation, of course, does not in general guarantee that the approximating function itself will be globally convex with the exception of special cases such as approximation by a quadratic function. For many families of approximating functions such as the generalized linear function, it is possible to find restrictions on the parameters such that a member of the family is globally convex on its effective domain. However, these restrictions will usually turn out to be more stringent than those implied by local convexity at the point of approximation. Moreover, not all families of approximating functions can be made globally convex without severely restricting the parameters. The transcendental logarithmic function, for example, can be made globally convex (for all **x**) only under rather restrictive assumptions such as a unitary elasticity of substitution between all pairs of commodities. It can be shown, however, that a sufficient condition for convexity on the set $\{\mathbf{x} | \mathbf{x} \geqq \mathbf{e}\}$, where $\mathbf{e} = [e\, e \cdots e]'$, is

---

[18] One should bear in mind here that under classical procedures for statistical inference, one can attach a confidence level to rejections, but not to "acceptances".

[19] Lau (1974).

$$\alpha_i \leqq 0 \quad \text{and} \quad B_{ij} \leqq 0 \quad \forall\, i \quad \text{and} \quad j.$$[20]

Unfortunately similar conditions do not obtain when the effective domain properly contains $\mathbf{x} = [\mathbf{1}]$, a vector of units, in which case one has to be content with local convexity. It can also be shown that a necessary and sufficient condition for a generalized version of the generalized linear function to be globally convex on its effective domain (the non-negative orthant of R^n), is $\alpha_i \leqq 0$ and $B_{ij} \leqq 0$, $i \neq j$, $\forall i$ and j.[21] Finally, it can be shown that for the quadratic function a necessary and sufficient condition for global convexity (that is, convexity on all of R^n) is that **B** is positive semidefinite.

Thus, the problem of testing the hypotheses of monotonicity, convexity and quasi-convexity becomes that of testing non-negativity and positive semidefiniteness constraints; and the problem of constrained estimation becomes that of imposition of non-negativity and positive semidefiniteness restrictions. Non-negativity constraints suffice for global convexity of the generalized linear function. However, convexity of the quadratic function, the transcendental logarithmic function, and the generalized linear function (at $\mathbf{x} = [\mathbf{1}]$), requires positive semidefiniteness constraints.

---

[20] Sufficiency follows from consideration of the convexity conditions of functions such as $B_{ij} \ln x_i \ln x_j$ and the fact that a non-negative linear combination of convex functions is convex and that an increasing convex function of a convex function is convex.

[21] Sufficiency is trivial and follows from the following facts: (i) $-y_i^{1/2} y_j^{1/2}$, $i \neq j$, is a convex function; (ii) a non-negative linear combination of convex function is convex; and finally (iii) the sum of a convex function and a linear function is convex.

Necessity follows from consideration of the diagonal elements of the Hessian matrix, which are

$$F_{ii} = -y_i^{-3/2}\Big(\alpha_i + \sum_{j \neq i} B_{ij} y_j^{1/2}\Big), \quad \forall i.$$

In order for $F_{ii} \geqq 0$ for all y in the non-negative orthant of R^n, it is necessary that $\alpha_i \leqq 0$ and $B_{ij} \leqq 0$, $i \neq j$, $\forall i,j$. Since these conditions are also sufficient for global convexity, they are therefore both necessary and sufficient.

3. The Cholesky Factorizability of Semidefinite and Indefinite Matrices

3.1. *Introduction*

In the preceding section we have seen that the hypothesis of convexity implies that the Hessian matrix of the approximating function is positive semidefinite at the point of approximation. Similarly, the hypothesis of quasi-convexity implies that the matrix $(\mathbf{H} + \lambda\, \nabla F\, \nabla F')$ where $\mathbf{H}$ and ∇F are respectively the Hessian matrix and gradient of the approximating function is positive semidefinite for a sufficiently large λ at the point of approximation. Thus, the problem of testing convexity and quasi-convexity of the underlying true functions becomes that of testing whether a given real symmetric matrix is positive semidefinite.

One solution which naturally suggests itself is based on the eigenvalue decomposition of real symmetric matrices.[22] It is well known that any real symmetric matrix $\mathbf{B}$ can be written as

$$\mathbf{B} = \mathbf{P\Lambda P}',$$

where $\mathbf{\Lambda}$ is a diagonal matrix whose elements are the eigenvalues of the matrix $\mathbf{B}$ and $\mathbf{PP}' = \mathbf{I}$. Moreover, $\mathbf{B}$ is positive semidefinite if and only if $\Lambda_{ii} \geqq 0$, $\forall i$. Thus, in principle, given an estimator of $\mathbf{B}$ and its variance–covariance matrix, one can compute an estimator of $\mathbf{\Lambda}$ and the variance–covariance matrix of $\mathbf{\Lambda}$, which may then be used to test the hypothesis that $\Lambda_{ii} \geqq 0$, $\forall i$. However, since $\mathbf{\Lambda}$ is in general a rather complicated function of $\mathbf{B}$, such computation is likely to be laborious. Further, if the estimate of $\mathbf{\Lambda}$ turns out not to be non-negative, and constrained estimation is necessary, that is, one needs to impose the constraint that $\Lambda_{ii} \geqq 0$, $\forall i$, the computational problem for the eigenvalue decomposition becomes quite complex because in the estimation of $\mathbf{B} = \mathbf{P\Lambda p}'$, it is necessary to impose not only the constraint that $\Lambda_{ii} \geqq 0$, $\forall i$, but also the orthonormality constraint of $\mathbf{PP}' = \mathbf{I}$. For $n > 2$ the computational burden becomes quite formidable.

For the practical reason, we introduce a different factorization of the matrix $\mathbf{B}$ – the Cholesky factorization – which avoids some of these difficulties.[23] It will be shown that every positive semidefinite matrix has a Cholesky factorization with non-negative Cholesky values.[24] Thus, the Hessian matrices of twice differentiable convex and quasi-convex functions may be characterized in terms of their Cholesky factorizations.

---

[22] This method was proposed in the cited earlier version of this paper and independently by Salvas-Bronsard et al. (1973).

[23] After this paper was substantially completed, Arthur Goldberger brought to my attention an article by Wiley, Schmidt and Bramble (1973) which also makes use of the Cholesky factorization in connection with imposing positive definiteness constraints.

Although not all real symmetric matrices have Cholesky factorizations, it will be shown that the set of real symmetric matrices of order n which do not have Cholesky factorizations has measure zero in the set of real symmetric matrices of order n.

3.2. *Cholesky Factorization*

Definition. A square matrix $\mathbf{A}$ is a *unit lower triangular matrix* if

$$\begin{aligned} A_{ii} &= 1, \quad \forall i, \\ A_{ij} &= 0, \quad j > i, \quad \forall i,j. \end{aligned}$$

A unit lower triangular matrix will be denoted $\mathbf{L}$.

Definition. A square matrix $\mathbf{A}$ is a *unit upper triangular matrix* if

$$\begin{aligned} A_{ii} &= 1, \quad \forall i, \\ A_{ij} &= 0, \quad j < i, \quad \forall i,j. \end{aligned}$$

A unit upper triangular matrix will be denoted by $\mathbf{R}$. The transpose of a unit lower triangular matrix is of course a unit upper triangular matrix, and vice versa.

Definition. A real symmetric square matrix $\mathbf{A}$ is said to have a *Cholesky factorization* if there exists a unit lower triangular matrix $\mathbf{L}$ and a diagonal matrix $\mathbf{D}$ such that

$$\mathbf{A} = \mathbf{LDL}',$$

where $\mathbf{L}'$ denotes the transpose of $\mathbf{L}$. The matrix $\mathbf{A}$ is also said to be *Cholesky factorizable.*

Definition. A square matrix $\mathbf{A}$ is an *upper triangular matrix* if

$$A_{ij} = 0, \quad j < i, \quad \forall i,j.$$

---

[24] The concept of Cholesky factorization is due to Cholesky. It is discussed in Householder (1964, pp. 10–17) and Wilkinson (1965, pp. 229–230).

An upper triangular matrix will be denoted **U**. We note that the product **DL**′ is an upper triangular matrix. Thus, for any matrix **A** which has a Cholesky factorization, one may write equivalently

$$\mathbf{A} = \mathbf{L}\mathbf{U}.$$

Lemma 3.1. The inverse of a unit lower triangular matrix is a unit lower triangular matrix.

Proof: The proof is by induction on the order of the matrix. The lemma is obviously true for $n = 1$. Assume that it is true for $n - 1$, we shall prove that it is true for n. An nth order unit lower triangular matrix may be written as

$$\mathbf{L}_n = \begin{bmatrix} \mathbf{L}_{n-1} & \begin{matrix} 0 \\ 0 \\ \cdot \\ \cdot \end{matrix} \\ \mathbf{l}' & 1 \end{bmatrix}$$

where $\mathbf{L}_{n-1}$ is a unit lower triangular matrix of order $n - 1$. The inverse may be directly computed as

$$\mathbf{L}_n^{-1} = \begin{bmatrix} \mathbf{L}_{n-1}^{-1} & \begin{matrix} 0 \\ 0 \\ \cdot \\ \cdot \\ 0 \end{matrix} \\ -\mathbf{l}'\mathbf{L}_{n-1}^{-1} & 1 \end{bmatrix}$$

since $\mathbf{L}_{n-1}$ is of order $n - 1$, $\mathbf{L}_{n-1}^{-1}$ is unit lower triangular by hypothesis. Hence $\mathbf{L}_n^{-1}$ is also unit lower triangular and the lemma is proved. Q.E.D.

We shall now show how a Cholesky factorization may be accomplished by way of an example. Let

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix},$$

and

$$\begin{aligned} \mathbf{L}\mathbf{U} &= \begin{bmatrix} 1 & 0 \\ L_{21} & 1 \end{bmatrix} \begin{bmatrix} U_{11} & U_{12} \\ 0 & U_{22} \end{bmatrix} \\ &= \begin{bmatrix} U_{11} & U_{12} \\ L_{21}U_{11} & L_{21}U_{12} + U_{22} \end{bmatrix}. \end{aligned}$$

By equating the matrices element by element, we have

$$\begin{aligned} U_{11} &= 1, \\ U_{12} &= 1, \\ L_{21}U_{11} &= 1, \end{aligned}$$

and

$$L_{21}U_{12} + U_{22} = 1,$$

which gives

$$\begin{aligned} U_{11} &= 1, \\ U_{12} &= 1, \\ L_{21} &= 1, \end{aligned}$$

and

$$U_{22} = 0.$$

Thus

$$\mathbf{L} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \qquad \mathbf{U} = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}.$$

Now

$$\begin{aligned} \mathbf{U} &= \mathbf{DL}', \\ \mathbf{D} &= \mathbf{UL}'^{-1}, \\ &= \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}. \end{aligned}$$

Thus, the Cholesky factorization is given by

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

In actual fact, it is never necessary to calculate the inverse of **L** explicitly. Since **D** is diagonal, one can obtain the elements of **D** from the equations

$$U_{ii} = D_{ii}L_{ii}, \quad \forall i,$$

or

$$D_{ii} = U_{ii},$$

since

$$L_{ii} = 1, \quad \forall i.$$

Notice that in the example, the matrix **A** is singular. Hence non-singularity of a matrix is not a necessary condition for the possibility of Cholesky factorization. It is not true, however, that all real symmetric matrices have Cholesky factorizations. For example, let

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix},$$

then the equations for the elements of **L** and **U** become

$$\begin{aligned} U_{11} &= 0, \\ U_{12} &= 1, \\ L_{21}U_{11} &= 1, \\ L_{21}U_{12} + U_{22} &= 1. \end{aligned}$$

But the equations $U_{11} = 0$ and $L_{21}U_{11} = 1$ are inconsistent with each other. Hence there do not exist matrices **L** and **U** such that $\mathbf{A} = \mathbf{LU}$ and therefore the matrix **A** does not have a Cholesky factorization. Notice also that **A** is non-singular. Hence non-singularity is not a sufficient condition for the possibility of Cholesky factorization. Finally, we note that by a simultaneous permutation of rows and columns,

$$\mathbf{A}^* = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix},$$

A becomes Cholesky factorizable. But this is not true of all real symmetric matrices either, as demonstrated by the example

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

which is non-singular but not Cholesky factorizable, even with simultaneous permutations of rows and columns. We shall show, however, that every positive semidefinite symmetric matrix has a Cholesky factorization in Section 3.3 and characterize all Cholesky factorizable real symmetric matrices in Section 3.4.

3.3. *Representation Theorem for Positive Semidefinite Matrices*

First we define positive definiteness and prove a couple of useful lemmas:

> *Definition.* A real symmetric matrix $\mathbf{A}$ is positive definite if $\mathbf{x}'\mathbf{A}\mathbf{x} > 0$ for all $\mathbf{x} \neq 0$; it is positive semidefinite if $\mathbf{x}'\mathbf{A}\mathbf{x} \geqq 0$ for all $\mathbf{x}$.

An equivalent definition of positive semidefiniteness is the following:

> *Definition.* A real symmetric matrix $\mathbf{A}$ is positive semidefinite if and only if *all* the principal minors of $\mathbf{A}$ of all orders are non-negative.[25]

> *Lemma* 3.2. If a diagonal element of a positive semidefinite matrix $\mathbf{A}$ is zero, the corresponding row and column must be identically zero.

Proof: Without loss of generality, we may take the matrix to be

$$\mathbf{A} = \begin{bmatrix} 0 & \mathbf{A}_1' \\ \mathbf{A}_1 & \mathbf{A}_n \end{bmatrix},$$

where

$$\mathbf{A}_1 \equiv \begin{bmatrix} A_{12} \\ A_{13} \\ \vdots \\ A_{1n} \end{bmatrix}, \qquad \mathbf{A}_n \equiv \begin{bmatrix} A_{22} A_{23} \cdots A_{2n} \\ A_{23} A_{33} \cdots A_{3n} \\ \vdots \\ A_{2n} A_{3n} \cdots A_{nn} \end{bmatrix}.$$

Positive semidefiniteness of $\mathbf{A}$ implies that for all $\mathbf{x}, \mathbf{x}'\mathbf{A}\mathbf{x} \geqq 0$. Let

$$\mathbf{x} = \begin{bmatrix} x_1 \\ A_1 \end{bmatrix},$$

where $x_1 < 0$. Then

$$\begin{aligned} \mathbf{x}'\mathbf{A}\mathbf{x} &= [x_1 \mathbf{A}_1'] \begin{bmatrix} 0 & \mathbf{A}_1' \\ \mathbf{A}_1 & \mathbf{A}_n \end{bmatrix} \begin{bmatrix} x_1 \\ \mathbf{A}_1 \end{bmatrix} \\ &= \mathbf{A}_1'\mathbf{A}_n\mathbf{A}_1 + 2x_1\mathbf{A}_1'\mathbf{A}_1. \end{aligned}$$

---

[25] Note that non-negativity of the *ordered* principal minors alone is not sufficient (although positivity is). For example, consider the matrix

$$\mathbf{A} = \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix}.$$

If $A_1'A_1 > 0$, then by choosing x_1 sufficiently large in magnitude, $\mathbf{x}'\mathbf{A}\mathbf{x}$ will become negative. Thus in order for $\mathbf{A}$ to be positive semidefinite, one must have $\mathbf{A}_1'\mathbf{A}_1 = 0$ which implies that $\mathbf{A}_1 = \mathbf{0}$. Q.E.D.

Lemma 3.3. If a matrix $\begin{bmatrix} A_{11} & \mathbf{A}_1' \\ \mathbf{A}_1 & \mathbf{A}_n \end{bmatrix}$ is positive semidefinite, and $A_{11} \neq 0$ then $\mathbf{A}_n - A_{11}^{-1}\mathbf{A}_1\mathbf{A}_1'$ is also positive semidefinite.

Proof: $\begin{bmatrix} A_{11} & \mathbf{A}_1' \\ \mathbf{A}_1 & \mathbf{A}_n \end{bmatrix}$ positive semidefinite implies that

$$[x_1\mathbf{x}']\begin{bmatrix} A_{11} & \mathbf{A}_1' \\ \mathbf{A}_1 & \mathbf{A}_n \end{bmatrix}\begin{bmatrix} x_1 \\ \mathbf{x} \end{bmatrix} \geqq 0, \quad \forall x_1, \mathbf{x}.$$

Thus, $x_1A_{11}x_1 + \mathbf{x}'\mathbf{A}_1x_1 + x_1\mathbf{A}_1'\mathbf{x} + \mathbf{x}'\mathbf{A}_n\mathbf{x} \geqq 0, \forall x_1, \mathbf{x}$. One may choose

$$x_1 = -A_{11}^{-1}\mathbf{A}_1'\mathbf{x}.$$

The inequality becomes

$$A_{11}^{-1}\mathbf{x}'\mathbf{A}_1\mathbf{A}_1'\mathbf{x} - \mathbf{x}'\mathbf{A}_1A_{11}^{-1}\mathbf{A}_1'\mathbf{x} - \mathbf{x}'\mathbf{A}_1A_{11}^{-1}\mathbf{A}_1'\mathbf{x} + \mathbf{x}'\mathbf{A}_n\mathbf{x} \geqq 0,$$

or

$$\mathbf{x}'(\mathbf{A}_n - A_{11}^{-1}\mathbf{A}_1\mathbf{A}_1')\mathbf{x} \geqq 0.$$

Since this holds for all $\mathbf{x}$, the matrix

$(\mathbf{A}_n - A_{11}^{-1}\mathbf{A}_1\mathbf{A}_1')$ is positive semidefinite. Q.E.D.

A corollary of this lemma is that if $\mathbf{A}$ is positive definite, then $\mathbf{A}_n - A_{11}^{-1}\mathbf{A}_1\mathbf{A}_1'$ is also positive definite.

Now we can prove the following representation theorem:

Theorem 3.1. Every positive semidefinite matrix $\mathbf{A}$ has a Cholesky factorization.

Proof: The proof is by induction on the order of the matrix. Clearly for $n = 1$,

$$\mathbf{A} = [A_{11}], \qquad A_{11} \geqq 0,$$

one can choose

$$\mathbf{L} = [1], \qquad \mathbf{D} = [A_{11}].$$

Now suppose that the theorem is true for positive semidefinite matrices or order less than or equal to $(n-1)$, we shall prove that it is true for positive semidefinite matrices of order n.

Let

$$\mathbf{A}=\begin{bmatrix} A_{11} & \mathbf{A}_1' \\ \mathbf{A}_1 & \mathbf{A}_n \end{bmatrix}.$$

Since $\mathbf{A}$ is positive semidefinite, $A_{11} \geqq 0$. If $A_{11}=0$, then by Lemma 3.2,

$$\mathbf{A}=\begin{bmatrix} 0 & 0\cdots 0 \\ 0 & \\ \vdots & \mathbf{A}_n \\ 0 & \end{bmatrix}.$$

$\mathbf{A}_n$ is a positive semidefinite matrix of order $(n-1)$ and hence has a Cholesky factorization. One may therefore choose

$$\mathbf{A}=\begin{bmatrix} 1 & 0 & 0\cdots 0 \\ 0 & & \\ 0 & & \\ \vdots & & \mathbf{L}_n \\ 0 & & \end{bmatrix}\begin{bmatrix} 0 & & 0\cdots 0 \\ 0 & & \\ \vdots & & \\ 0 & & \mathbf{D}_n \\ 0 & & \end{bmatrix}\begin{bmatrix} 1 & 0 & 0\cdots 0 \\ 0 & & \\ 0 & & \\ \vdots & & \mathbf{L}_n' \\ 0 & & \end{bmatrix},$$

where $\mathbf{A}_n=\mathbf{L}_n\mathbf{D}_n\mathbf{L}_n'$ is the Cholesky factorization of $\mathbf{A}_n$. If $A_{11}>0$, consider the following matrix identity:

$$\begin{bmatrix} A_{11} & \mathbf{A}_1' \\ \mathbf{A}_1 & \mathbf{A}_n \end{bmatrix}=\begin{bmatrix} 1 & 0 & 0\cdots 0 \\ A_{11}^{-1}\mathbf{A}_1 & & \mathbf{I} \end{bmatrix}\begin{bmatrix} A_{11} & 0 & 0\cdots 0 \\ 0 & & \\ 0 & & \mathbf{A}_n-A_{11}^{-1}\mathbf{A}_1\mathbf{A}_1' \\ \vdots & & \\ 0 & & \end{bmatrix}\begin{bmatrix} 1 & A_{11}^{-1}\mathbf{A}_1' \\ 0 & \\ 0 & \\ \vdots & \mathbf{I} \\ 0 & \end{bmatrix},$$

which may be directly verified by computation. Now $\mathbf{A}_n-A_{11}^{-1}\mathbf{A}_1\mathbf{A}_1'$ is a positive semidefinite matrix of order $(n-1)$, by Lemma 3.3. Thus it has a Cholesky factorization,

$$\mathbf{A}_n-A_{11}^{-1}\mathbf{A}_1\mathbf{A}_1'=\mathbf{L}_n^*\mathbf{D}_n^*\mathbf{L}_n^{*\prime}.$$

Therefore $\mathbf{A}$ has a Cholesky factorization given by

$$\mathbf{A} = \begin{bmatrix} A_{11} & \mathbf{A}_1' \\ \mathbf{A}_1 & \mathbf{A}_n \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 0\cdots 0 \\ A_{11}^{-1}\mathbf{A}_1 & \mathbf{L}_n^* & \end{bmatrix} \begin{bmatrix} A_{11} & 0 & 0\cdots 0 \\ 0 & & \\ 0 & & \\ \vdots & & \mathbf{D}_n^* \\ 0 & & \end{bmatrix} \begin{bmatrix} 1 & A_{11}^{-1}\mathbf{A}_1' \\ 0 & \\ 0 & \\ \vdots & \mathbf{L}_n^{*\prime} \\ 0 & \end{bmatrix}. \qquad \text{Q.E.D.}$$

The proof given here follows essentially the argument given by Householder (1964).[26] Note that this theorem also implies that all negative semidefinite matrices have Cholesky factorizations.

The elements of the diagonal matrix $\mathbf{D}$, that is, the D_{ii}'s will be referred to as *Cholesky values*. The following theorem establishes the properties of the Cholesky values of positive definite and semidefinite matrices.

> *Theorem* 3.2. A real symmetric matrix $\mathbf{A}$ is positive definite (semidefinite) if and only if its Cholesky values are positive (nonnegative).

Proof: A positive definite matrix $\mathbf{A}$ can be written as $\mathbf{LDL}'$. Thus

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{x}'\mathbf{LDL}'\mathbf{x} > 0, \quad \forall \mathbf{x}, \quad \mathbf{x} \neq 0.$$

Writing $\mathbf{y} = \mathbf{L}'\mathbf{x}$, we have

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{y}'\mathbf{D}\mathbf{y} = \sum_i D_{ii}y_i^2 > 0, \quad \forall \mathbf{y}, \quad \mathbf{y} \neq 0.$$

Thus all $D_{ii} > 0, \forall i$. Conversely, if $D_{ii} > 0$, $\forall i, \mathbf{y}'\mathbf{D}\mathbf{y} > 0$, $\forall \mathbf{y}, \mathbf{y} \neq 0$. Buy $\mathbf{y} = \mathbf{0}$ implies $\mathbf{x} = \mathbf{L}'^{-1}\mathbf{y} = \mathbf{0}$. Thus

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{y}'\mathbf{D}\mathbf{y} > 0, \quad \forall \mathbf{x}; \quad \mathbf{x} \neq 0.$$

A similar proof goes through for a positive semidefinite matrix $\mathbf{A}$. Q.E.D.

It is useful to give a determinantal interpretation to the Cholesky values. Consider the following partitioned matrix identity:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{L}_{11} & 0 \\ \mathbf{L}_{21} & \mathbf{L}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{D}_{11} & 0 \\ 0 & \mathbf{D}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{L}_{11}' & \mathbf{L}_{21}' \\ 0 & \mathbf{L}_{22}' \end{bmatrix},$$

---

[26] See Householder (1964, pp. 12–13).

where $\mathbf{L}_{11}$ and $\mathbf{L}_{22}$ are unit lower triangular matrices, and $\mathbf{D}_{11}$ and $\mathbf{D}_{22}$ are diagonal matrices. By direct computation

$$\begin{bmatrix}\mathbf{A}_{11} & \mathbf{A}_{12}\\ \mathbf{A}_{21} & \mathbf{A}_{22}\end{bmatrix} = \begin{bmatrix}\mathbf{L}_{11}\mathbf{D}_{11}\mathbf{L}'_{11} & \mathbf{L}_{11}\mathbf{D}_{11}\mathbf{L}'_{21}\\ \mathbf{L}_{21}\mathbf{D}_{11}\mathbf{L}'_{11} & \mathbf{L}_{21}\mathbf{D}_{11}\mathbf{L}'_{21}+\mathbf{L}_{22}\mathbf{D}_{22}\mathbf{L}'_{22}\end{bmatrix}.$$

First, since the determinants of unit lower triangular matrices are identically one, we have

$$|\mathbf{A}| = \begin{vmatrix}\mathbf{D}_{11} & 0\\ 0 & \mathbf{D}_{22}\end{vmatrix} = \prod_i D_{ii}.$$

The determinant of $\mathbf{A}$ is thus the product of all of its Cholesky values. Second, take any principal minor of $\mathbf{A}$, say $|\mathbf{A}_{11}|$,

$$|\mathbf{A}_{11}| = |\mathbf{L}_{11}||\mathbf{D}_{11}||\mathbf{L}'_{11}| = |\mathbf{D}_{11}| = \prod_i D_{ii},$$

where the product is taken over the Cholesky values corresponding to the $\mathbf{A}_{11}$ block. Thus, we have the following system of determinantal equalities:

$$|A_{11}| = D_{11},$$

$$\begin{vmatrix}A_{11} & A_{12}\\ A_{12} & A_{22}\end{vmatrix} = D_{11}D_{22},$$

$$\begin{vmatrix}A_{11} & A_{12} & A_{13}\\ A_{12} & A_{22} & A_{23}\\ A_{13} & A_{23} & A_{33}\end{vmatrix} = D_{11}D_{22}D_{33},$$

$$\vdots$$

$$\begin{vmatrix}A_{11} & A_{12}\cdots A_{1n}\\ A_{12} & A_{22}\cdots A_{2n}\\ \vdots & \vdots\\ A_{1n} & A_{2n}\cdots A_{nn}\end{vmatrix} = D_{11}D_{22}D_{33}\cdots D_{nn}.$$

This interpretation is valid, however, if and only if a Cholesky factorization of the matrix $\mathbf{A}$ exists.

> *Theorem* 3.3. If $\mathbf{A}$ is positive definite, then the Cholesky factorization is unique.

Proof: The proof is by direct computation. Suppose $\mathbf{A} = \mathbf{LDL}' = \mathbf{L}^*\mathbf{D}^*\mathbf{L}^{*\prime}$, and $\mathbf{A}$ is positive definite. This implies that $D_{ii} > 0$ and

$D_{ii}^* > 0, \forall i$, by Theorem 3.2. By equating $\mathbf{LDL}'$ and $\mathbf{L}^*\mathbf{D}^*\mathbf{L}^{*\prime}$ element by element, it can be shown that indeed $\mathbf{L} = \mathbf{L}^*$ and $\mathbf{D} = \mathbf{D}^*$. Q.E.D.

Uniqueness is also discussed by Householder (1964) in terms of determinantal conditions.[27]

> *Theorem* 3.4. If a real symmetric matrix **A** has a Cholesky factorization, then the number of positive, negative, and zero Cholesky values is the same as the number of positive, negative, and zero eigenvalues.

This theorem follows from Sylvester's Law of Inertia, a proof of which may be found in Gantmacher (1959),[28] which implies conservation of the signature of a real symmetric matrix. From this theorem we can also deduce immediately that if **A** is positive definite, all the Cholesky values are positive; if **A** is positive semidefinite, all the Cholesky values are nonnegative.

We conclude that every positive semidefinite matrix **A** has a Cholesky factorization $\mathbf{LDL}'$ with all the elements of **D** non-negative. Thus, to check whether a real symmetric matrix is positive semidefinite, one needs only check its Cholesky values. And to impose the condition that a real symmetric matrix is positive semidefinite, one needs only constrain the Cholesky values to be non-negative.

3.4. *Representation Theorem for Arbitrary Real Symmetric Matrices*

We have shown that all semidefinite matrices are Cholesky factorizable in the previous subsection. However, semidefiniteness is by no means necessary for Cholesky factorizability as the last example in subsection 3.2 illustrates.

Let

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}.$$

The Cholesky factorization of **A** is

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

---

[27] See Householder (1964, pp. 10–12).

[28] See Gantmacher (1959, pp. 296–298).

Since the Cholesky values are 1 and -1, $\mathbf{A}$ is an indefinite matrix. It is also clear that non-singularity of $\mathbf{A}$ is neither necessary (because of semidefinite matrices) nor sufficient – because of example

$$A=\begin{bmatrix}0 & 1\\ 1 & 1\end{bmatrix},$$

for Cholesky factorizability. The purpose of this subsection is to characterize the set of all real symmetric matrices which are Cholesky factorizable and to show that the set of all real symmetric matrices which are not Cholesky factorizable is a subset of measure zero in the set of all real symmetric matrices.

We first prove two simple but useful lemmas:

Lemma 3.4. A real symmetric matrix $\mathbf{A}$ with $A_{11}=0$ is Cholesky factorizable only if the first row and first column are identically zero.

Proof: Let

$$\mathbf{A}=\begin{bmatrix}0 & \mathbf{A}_1'\\ \mathbf{A}_1 & \mathbf{A}_n\end{bmatrix}=\mathbf{LDL}'.$$

The first row of $\mathbf{LDL}'$ may be directly computed as

$$\begin{aligned}D_{11}&=A_{11}=0,\\ L_{21}D_{11}&=A_{12},\\ &\vdots\\ L_{n1}D_{11}&=A_{1n}.\end{aligned}$$

Since $D_{11}=0$, $A_{1i}=0$, $i=2\cdots n$. Q.E.D.

Lemma 3.5. A real symmetric Cholesky factorizable square matrix is non-singular if and only if all the Cholesky values are non-zero.

Proof: Referring back to the determinantal interpretation of Cholesky values in Section 3.3, if one or more of the Cholesky values is zero, then the determinant of the matrix is zero, and therefore the matrix is singular. If the matrix is non-singular, then its determinant is non-zero and none of the Cholesky values can be zero. Q.E.D.

Our first theorem characterizes all real symmetric matrices which are Cholesky factorizable.

Theorem 3.5. Let a real symmetric matrix $\mathbf{A}$ be partitioned conformably as

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}'_{1n} \\ \mathbf{A}_{1n} & \mathbf{A}_n \end{bmatrix},$$

where $\mathbf{A}_1$ and $\mathbf{A}_n$ are both real symmetric square matrices. In addition, suppose that $\mathbf{A}_1$ is Cholesky factorizable and non-singular. Then $\mathbf{A}$ is Cholesky factorizable if and only if $\mathbf{A}_n - \mathbf{A}_{1n}\mathbf{A}_1^{-1}\mathbf{A}'_{1n}$ is Cholesky factorizable.

Proof: If $\mathbf{A}$ were Cholesky factorizable then there exist $\mathbf{L}_1$, $\mathbf{L}_2$, $\mathbf{L}_{21}$, $\mathbf{D}_1$, $\mathbf{D}_2$ such that

$$\begin{bmatrix} \mathbf{L}_1 & 0 \\ \mathbf{L}_{21} & \mathbf{L}_2 \end{bmatrix} \begin{bmatrix} \mathbf{D}_1 & 0 \\ 0 & \mathbf{D}_2 \end{bmatrix} \begin{bmatrix} \mathbf{L}'_1 & \mathbf{L}'_{21} \\ 0 & \mathbf{L}'_2 \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}'_{1n} \\ \mathbf{A}_{1n} & \mathbf{A}_n \end{bmatrix},$$

which implies

$$\begin{aligned} \mathbf{L}_1\mathbf{D}_1\mathbf{L}'_1 &= \mathbf{A}_1, \\ \mathbf{L}_1\mathbf{D}_1\mathbf{L}'_{21} &= \mathbf{A}'_{1n}, \\ \mathbf{L}_{21}\mathbf{D}_1\mathbf{L}'_{21} + \mathbf{L}_2\mathbf{D}_2\mathbf{L}'_2 &= \mathbf{A}_n. \end{aligned}$$

But $\mathbf{A}_1$ is non-singular, then by Lemma 3.5, $\mathbf{D}^{-1}$ exists. Therefore, $\mathbf{L}'_{21} = \mathbf{D}_1^{-1}\mathbf{L}_1^{-1}\mathbf{A}'_{1n}$ and $\mathbf{L}_{21}\mathbf{D}_1\mathbf{L}'_{21} = \mathbf{A}_{1n}\mathbf{A}_1^{-1}\mathbf{A}'_{1n}$. Hence $\mathbf{A}_n - \mathbf{A}_{1n}\mathbf{A}_1^{-1}\mathbf{A}'_{1n} = \mathbf{L}_2\mathbf{D}_2\mathbf{L}'_2$ and is Cholesky factorizable.

Conversely, consider the matrix identity

$$\begin{bmatrix} \mathbf{A}_1 & \mathbf{A}'_{1n} \\ \mathbf{A}_{1n} & \mathbf{A}_n \end{bmatrix} = \begin{bmatrix} \mathbf{I} & 0 \\ \mathbf{A}_{1n}\mathbf{A}_1^{-1} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{A}_1 & 0 \\ 0 & \mathbf{A}_n - \mathbf{A}_{1n}\mathbf{A}_1^{-1}\mathbf{A}'_{1n} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{A}_1^{-1}\mathbf{A}'_{1n} \\ \mathbf{0} & \mathbf{I} \end{bmatrix},$$

which may be verified by direct multiplication. If

$$\mathbf{A}_1 = \mathbf{L}_1\mathbf{D}_1\mathbf{L}'_1,$$

and

$$\mathbf{A}_n - \mathbf{A}_{1n}\mathbf{A}_1^{-1}\mathbf{A}'_{1n} = \mathbf{L}_2\mathbf{D}_2\mathbf{L}'_2,$$

then the matrix identity becomes

$$\begin{bmatrix} \mathbf{I} & 0 \\ \mathbf{A}_{1n}\mathbf{A}_1^{-1} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{L}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_2 \end{bmatrix} \begin{bmatrix} \mathbf{D}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_2 \end{bmatrix} \begin{bmatrix} \mathbf{L}'_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{L}'_2 \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{A}_1^{-1}\mathbf{A}'_{1n} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{L}_1 & \mathbf{0} \\ \mathbf{A}_{1n}\mathbf{A}_1^{-1} & \mathbf{L}_2 \end{bmatrix} \begin{bmatrix} \mathbf{D}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_2 \end{bmatrix} \begin{bmatrix} \mathbf{L}'_1 & \mathbf{A}_1^{-1}\mathbf{A}'_{1n} \\ \mathbf{0} & \mathbf{L}'_2 \end{bmatrix},$$

which implies that $\mathbf{A}$ is Cholesky factorizable. Q.E.D.

Corollary 5.1. Let the real symmetric matrix be

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}'_{1n} \\ \mathbf{A}_{1n} & \mathbf{A}_n \end{bmatrix},$$

where $\mathbf{A}_n$ is a scalar, then $\mathbf{A}_1$ is Cholesky factorizable and nonsingular implies that $\mathbf{A}$ is Cholesky factorizable.

Proof: $A_n - \mathbf{A}_{1n}\mathbf{A}_1^{-1}\mathbf{A}'_{1n}$ is a scalar matrix which is always Cholesky factorizable. Q.E.D.

This corollary implies that any 2×2 real symmetric matrix $\mathbf{A}$ is Cholesky factorizable if A_{11} is different from zero.

This theorem provides a constructive way of verifying whether a given real symmetric matrix $\mathbf{A}$ is Cholesky factorizable. If $A_{11} = 0$, we know that the first row and column must be identically zero for Cholesky factorizability. Then $\mathbf{A}$ is Cholesky factorizable if and only if $\mathbf{A}_n$, the submatrix of $\mathbf{A}$ with the first row and first column deleted is Cholesky factorizable. If $A_{11} \neq 0$, then $\mathbf{A}$ is Cholesky factorizable if and only if $\mathbf{A}_n - \mathbf{A}_{1n}A_{11}^{-1}\mathbf{A}'_{1n}$ is Cholesky factorizable. One can continue in this way until some $\mathbf{A}^*_n - \mathbf{A}^*_{1n}A^{*-1}_{11}\mathbf{A}^{*\prime}_{1n}$ becomes a scalar matrix or is shown to be not Cholesky factorizable.

Of special interest, of course, is the case in which the 1,1-th element of $\mathbf{A}_n - \mathbf{A}_{1n}\mathbf{A}_1^{-1}\mathbf{A}'_{1n}$ is zero. In that case, by Lemma 3.4, the first row and the first column of $\mathbf{A}_n - \mathbf{A}_{1n}\mathbf{A}_1^{-1}\mathbf{A}'_{1n}$ is identically zero. To be more specific let us write:

$$\mathbf{A}_1 = \begin{bmatrix} A_{11} & A_{12} \cdots A_{1m} \\ A_{12} & A_{22} \cdots A_{2m} \\ \vdots & \vdots \quad \vdots \\ A_{1m} & A_{2m} \cdots A_{mm} \end{bmatrix},$$

$$\mathbf{A}'_{1n} = \begin{bmatrix} A_{1,m+1} & A_{1,m+2} \cdots A_{1n} \\ \vdots & \vdots \quad \vdots \\ A_{m,m+1} & A_{m,m+2} \cdots A_{m,n} \end{bmatrix},$$

$$\mathbf{A}_n = \begin{bmatrix} A_{m+1,m+1} & A_{m+1,m+2} \cdots A_{m+1,n} \\ \vdots & \vdots \quad \vdots \\ A_{m+1,n} & A_{m+2,n} \cdots A_{nn} \end{bmatrix},$$

then the first row and column of $\mathbf{A}_n - \mathbf{A}_{1n}\mathbf{A}_1^{-1}\mathbf{A}'_{1n}$ being zero implies

$$\begin{bmatrix} A_{m+1,m+1} \\ A_{m+1,m+2} \\ \vdots \\ A_{m+1,n} \end{bmatrix} = \mathbf{A}_{1n}\mathbf{A}_1^{-1} \begin{bmatrix} A_{1,m+1} \\ A_{2,m+1} \\ \vdots \\ A_{m,m+1} \end{bmatrix},$$

which in turn implies that

$$\begin{bmatrix} A_{m+1,m+2} \\ A_{m+1,m+2} \\ \vdots \\ A_{m+1,n} \end{bmatrix} = \sum_{i=1}^{m} \alpha_i \begin{bmatrix} A_{i,m+1} \\ A_{i,m+2} \\ \vdots \\ A_{i,n} \end{bmatrix},$$

that is, the first column of $\mathbf{A}_n$ can be expressed as a linear combination of the columns of $\mathbf{A}_{1n}$.

Next, we give a set of sufficient conditions for Cholesky factorizability which do not depend on semidefiniteness.

> *Theorem* 3.6. If the ordered principal submatrices of all orders up to $n-1$ of a real symmetric matrix $\mathbf{A}$ of order n are non-singular, then $\mathbf{A}$ is Cholesky factorizable.

Proof: The proof is by induction on the order of the matrix. For $n = 1$, the theorem is trivially true. For $n = 2$, if $A_{11} \neq 0$, $\mathbf{A}$ is Cholesky factorizable by Corollary 5.1. Now suppose the theorem is true for $(n-1)$, consider an nth order real symmetric matrix, partitioned into

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}'_{1n} \\ \mathbf{A}_{1n} & \mathbf{A}_{nn} \end{bmatrix},$$

where A_1 is $(n-1)\times(n-1)$ and all the ordered principal submatrices of $\mathbf{A}$ are non-singular. Since all the ordered principal submatrices of $\mathbf{A}_1$ are non-singular, $\mathbf{A}_1$ is Cholesky factorizable by hypothesis. But $\mathbf{A}_1$ is also non-singular, because it is a principal submatrix of $\mathbf{A}$. Then by Theorem 3.5, $\mathbf{A}$ is Cholesky factorizable. Q.E.D.

Finally, we want to show that although not all real symmetric matrices are Cholesky factorizable, those which are not constitute a subset of measure zero in the set of all real symmetric matrices.

Let A_n be the set of all real symmetric matrices of order n. Let A_n^0 be the subset of all such matrices which are singular. Let A_n^* be the subset of all such matrices which are non-singular. Then, by definition

$$\boldsymbol{A}_n^0 \bigcup \boldsymbol{A}_n^* = \boldsymbol{A}_n, \qquad \boldsymbol{A}_n^0 \bigcap \boldsymbol{A}_n^* = \phi.$$

Lemma 3.6. The set of singular matrices $\boldsymbol{A}_n^0$ in $\boldsymbol{A}_n$ is of measure zero.

Proof: Consider each element A of $\boldsymbol{A}_n$ as an element in $R^{n(n+1)/2}$. For each $\{A_{12}A_{13}\cdots A_{nn}\}$ in $R^{n(n+1)/2-1}$, the set of A_{11}'s such that

$$|\mathbf{A}| = A_{11}\begin{vmatrix} A_{22}\cdots A_{2n} \\ A_{23}\cdots A_{3n} \\ \vdots \quad \vdots \\ A_{2n}\cdots A_{nn}\end{vmatrix} + \begin{vmatrix} 0 & A_{12} & A_{13}\cdots A_{1n} \\ A_{12} & A_{22} & \cdots A_{2n} \\ \vdots & & \vdots \\ A_{1n} & A_{2n} & \cdots A_{nn}\end{vmatrix} = 0$$

is a set of measure zero. Hence by Fubini's theorem, the measure of the set of A's for which $|\mathbf{A}| = 0$ is of measure zero. Q.E.D.

Lemma 3.7. The set of matrices $\boldsymbol{A}_{n,k}^0$ in $\boldsymbol{A}_n$ for which the kth order principal submatrix, that is, $[A_{ij};i,j = 1,\ldots,k;k < n]$, is singular is of measure zero.

Proof: Again consider each element A of $\boldsymbol{A}_n$ as an element in $R^{n(n+1)/2}$. For each $\{A_{12}A_{13}\cdots A_{nn}\}$ in $R^{n(n+1)/2-1}$, the set of A_{11} is such that

$$|\mathbf{A}_{n,k}| = A_{11}\begin{vmatrix} A_{22}\cdots A_{2k} \\ A_{23}\cdots A_{3k} \\ \\ A_{2k}\cdots A_{kk}\end{vmatrix} + \begin{vmatrix} 0 & A_{12} & A_{13}\cdots A_{1k} \\ A_{12} & A_{22} & \cdots A_{2k} \\ \vdots & & \vdots \\ A_{1k} & A_{2k} & \cdots A_{kk}\end{vmatrix} = 0$$

is a set of measure zero. Hence by Fubini's theorem, the set of A's for which $|\mathbf{A}_{n,k}| = 0$ is of measure zero. Q.E.D.

Lemma 3.8. The set of matrices $\boldsymbol{A}_{n,n}^*$ for which the naturally ordered sequence of principal submatrices are all non-singular is of measure one in $\boldsymbol{A}_n$.

Proof: Any element of $\boldsymbol{A}_{n,n}^*$ cannot be contained in $\boldsymbol{A}_{n,k}^0$ for any k, otherwise at least one of the principal submatrices will be singular. Hence it must be contained in $\boldsymbol{A}_n - \bigcup_k \boldsymbol{A}_{n,k}^0$. But any element of $\boldsymbol{A}_n - \bigcup_k \boldsymbol{A}_{n,k}^0$ must have its naturally ordered principal submatrices all non-singular, and is hence contained in $\boldsymbol{A}^*_{n,n}$. Thus,

$$\boldsymbol{A}^*_{n,n} = \boldsymbol{A}_n - \bigcup_k \boldsymbol{A}^0_{n,k},$$

and

$$\boldsymbol{A}^*_{n,n} \bigcap \left[\bigcup_k \boldsymbol{A}^0_{n,k} \right] = \phi.$$

Therefore,

$$\mu(\boldsymbol{A}^*_{n,n}) = \mu(\boldsymbol{A}_n) - \mu\left(\bigcup_k \boldsymbol{A}^0_{n,k} \right),$$

but the measure of a finite union of sets of measure zero is zero. Thus

$$\mu(\boldsymbol{A}^*_{n,n}) = 1. \qquad \text{Q.E.D.}$$

Theorem 3.7. The set of real symmetric matrices of order n which are not Cholesky factorizable is of measure zero within the set of all real symmetric matrices of order n.

Proof:[29] The set of real symmetric matrices which are Cholesky factorizable contains the set of real symmetric matrices with all ordered principal submatrices which are non-singular, by Theorem 3.6. But the latter set has measure one by Lemma 3.8. Hence the set of Cholesky factorizable real symmetric matrices is of measure one, and its complement, the set of matrices which are not Cholesky factorizable, is of measure zero. Q.E.D.

4. Estimation

4.1. *Introduction*

The proposed method of estimation is maximum likelihood, which is known to have certain optimal properties. For the sake of expositional convenience, we shall focus our attention on the following model:[30]

$$Y_t = \alpha_0 + \boldsymbol{\alpha}'\mathbf{Z}_t + \tfrac{1}{2}\mathbf{Z}_t'\mathbf{B}\mathbf{Z}_t + \epsilon_t, \quad t = 1,\dots,T,$$

---

[29] The proof of this theorem is due to Daniel McFadden.

[30] In most models of producer or consumer behavior, there will be more than one equation as well as parametric restrictions across equations. We abstract from these complications so as to keep the exposition simple.

where $Z_{ti} = g(X_{ti})$ with $g(\cdot)$ a known algebraic function of a single variable, $\mathbf{X}_t$ is a vector of independent variables, and $\boldsymbol{\epsilon}$ is distributed as $N(\mathbf{0},\sigma^2\mathbf{I})$, $\mathbf{B}$ is a real symmetric matrix and our objective is to test the hypotheses that the gradient is non-negative and/or the Hessian matrix is positive semidefinite and to find an estimator of $\boldsymbol{\alpha}$ and $\mathbf{B}$ that are consistent with a non-negative gradient and/or a positive semidefinite Hessian matrix at the point of approximation.

Since the Hessian matrix of any convex function has a Cholesky factorization, it is possible to transform the elements of the Hessian matrix in terms of the elements of its Cholesky factorization $\mathbf{L}$ and $\mathbf{D}$. Then a test of the convexity hypothesis consists of a simultaneous test that all the D_{ii}'s of the Cholesky factorization are non-negative. If constrained estimates are needed, they can be obtained by setting each D_{ii} equal to the square of a new parameter, say, D_{ii}^{*2}.

In terms of the parameters of the quadratic function, this implies

$$\begin{bmatrix} B_{11} & B_{12}\cdots B_{1n} \\ B_{12} & B_{22}\cdots B_{2n} \\ \vdots & \\ B_{1n} & B_{2n}\cdots B_{nn} \end{bmatrix}$$

$$= \begin{bmatrix} D_{11} & L_{21}D_{11} & \cdots L_{n1}D_{11} \\ L_{21}D_{11} & L_{21}^2D_{11} + D_{22} & \cdots L_{21}L_{n1}D_{11} + L_{n2}D_{22} \\ \vdots & & \\ L_{n1}D_{11} & L_{21}L_{n1}D_{11} + L_{n2}D_{22} & \cdots L_{n1}^2D_{11} + L_{n2}^2D_{22} + \cdots + D_{nn} \end{bmatrix}.$$

We note that this condition is global, as the Hessian matrix is constant for the quadratic function, and is consistent with local convexity of the unknown, underlying true function of which the quadratic function is an approximation.

In terms of the parameters of the transcendental logarithmic function at $\mathbf{x} = [\mathbf{1}]$, a vector of units, this implies

$$\begin{bmatrix} B_{11} + \alpha_1(\alpha_1 - 1) & B_{12} + \alpha_1\alpha_2 & \cdots B_{1n} + \alpha_1\alpha_n \\ B_{12} + \alpha_1\alpha_2 & B_{22} + \alpha_2(\alpha_2 - 1) & \cdots B_{2n} + \alpha_2\alpha_n \\ B_{1n} + \alpha_1\alpha_n & B_{2n} + \alpha_2\alpha_n & \cdots B_{nn} + \alpha_n(\alpha_n - 1) \end{bmatrix} = \mathbf{LDL}'.$$

Satisfaction of this equality with $D_{ii} \geqq 0$, $\forall i$, does not imply global convexity of the transcendental logarithmic function, but is consistent with convexity in a neighborhood of $\mathbf{x} = [\mathbf{1}]$, and with local convexity of the unknown, underlying true function of which the transcendental logarithmic function is an approximation.

And in terms of the parameters of the generalized linear function at $\mathbf{x} = [\mathbf{1}]$, a vector of units, this implies

$$\frac{1}{4}\begin{bmatrix} -\left(\alpha_1 + \sum\limits_{j\neq 1} B_{1j}\right) & B_{12} & B_{13}\cdots & B_{1n} \\ B_{12} & -\left(\alpha_2 + \sum\limits_{j\neq 2} B_{2j}\right) & \cdots & B_{2n} \\ \vdots & \vdots & & \vdots \\ B_{1n} & B_{2n} & & -\left(\alpha_n + \sum\limits_{j\neq n} B_{nj}\right) \end{bmatrix} = \mathbf{LDL'}.$$

Again this does not imply global convexity. Estimation of the original, untransformed model will be referred to as "Problem 0". For the purposes of hypotheses testing and constrained estimation, two additional, separate estimation problems may be distinguished. For the quadratic function, these are[31]

Problem 1:

$$Y_t = \alpha_0 + \alpha'\mathbf{Z}_t + \tfrac{1}{2}\mathbf{Z}_t'\mathbf{LDL}'\mathbf{Z}_t + \epsilon_t, \qquad t = 1,\ldots,T.$$

In this case we seek parameters α_i's, L_{ij}'s $(j < i)$ and D_{ii}'s, without constraints.

Problem 2:

$$Y_t = \alpha_0 + \alpha^{*^2\prime}\mathbf{Z}_t + \tfrac{1}{2}\mathbf{Z}_t'\mathbf{LD}^{*^2}\mathbf{L}'\mathbf{Z}_t + \epsilon_t, \qquad t = 1,\ldots,T.$$

In the second case we seek parameters α_i^*'s, L_{ij}'s $(j < i)$ and D_{ii}^*'s, where

$$\alpha_i^{*^2} \equiv \alpha_i,$$

and

$$D_{ii}^{*^2} \equiv D_{ii}, \quad \forall i.$$

Again the problem is unconstrained. But the resultant estimates of α_i and D_{ii} from Problem 2 will be non-negative, hence $\hat{\alpha}_i \geqq 0$, $\forall i$, and $\hat{\mathbf{B}} \equiv \mathbf{LDL}' \equiv \mathbf{LD}^{*^2}\mathbf{L}'$ will be positive semidefinite. Under our specification, the likelihood maximization problem is equivalent to an unconstrained nonlinear least-squares problem.

---

[31] Similar problems 1 and 2 may be set up for the "translog" and generalized linear functions. One should note that it is the Hessian matrix, which may be different from the matrix of second order coefficients, which should be set equal to $\mathbf{LDL}'$.

Here, the computational advantage of the Cholesky factorization over the eigenvalue decomposition is most easily seen. If the Cholesky factorization $\mathbf{B} = \mathbf{LDL}'$ is used, then the number of independent unknown parameters of $\mathbf{B}$, $n(n+1)/2$, is precisely equal to the number of independent unknown parameters of $\mathbf{L}$ and $\mathbf{D}$, $n(n-1)/2 + n = n(n+1)/2$. Hence no additional constraints are required. On the other hand, if the eigenvalue decomposition $\mathbf{B} = \mathbf{P\Lambda P}'$ is used, the number of unknown parameters of $\mathbf{P}$ and $\mathbf{\Lambda}$ are $(n^2 + n)$. The parameters of $\mathbf{P}$ are subject to additional orthonormality restrictions such that $\mathbf{P'P} = \mathbf{I}$. One can either solve out for the parameters of $\mathbf{P}$ in terms of a minimal set of independent parameters or alternatively impose $\mathbf{P'P} = \mathbf{I}$ as $n(n+1)/2$ additional side conditions. In any event, substantial computations are involved. All these are avoided by using the Cholesky factorization.

The basic model for quasiconvexity is again

$$Y_t = \alpha_0 + \boldsymbol{\alpha}'\mathbf{Z}_t + \tfrac{1}{2}\mathbf{Z}_t'\mathbf{B}\mathbf{Z}_t + \epsilon_t, \qquad t = 1,\ldots,T,$$

where $\boldsymbol{\epsilon}$ is $N(\mathbf{0},\sigma^2\mathbf{I})$ and $\nabla F \geqq 0$, $(H + \lambda \nabla F \nabla F')$ positive semidefinite for all sufficiently large positive scalar constant λ. In terms of the parameters of the quadratic function, at $\mathbf{x} = [\mathbf{0}]$, the vector of zeroes,

$$\mathbf{H}(\mathbf{x}) + \lambda \nabla F(\mathbf{x}) \nabla F(\mathbf{x})'$$
$$= \begin{bmatrix} B_{11} + \lambda\alpha_1^2 & B_{12} + \lambda\alpha_1\alpha_2 & B_{1n} + \lambda\alpha_1\alpha_n \\ B_{12} + \lambda\alpha_1\alpha_2 & B_{22} + \lambda\alpha_2^2 & B_{2n} + \lambda\alpha_2\alpha_n \\ \vdots & & \vdots \\ B_{1n} + \lambda\alpha_1\alpha_n & B_{2n} + \lambda\alpha_2\alpha_n & B_{nn} + \lambda\alpha_n^2 \end{bmatrix}.$$

In terms of the parameters of the transcendental logarithmic function, at $\mathbf{x} = [\mathbf{1}]$, a vector of units,

$$\mathbf{H}(\mathbf{x}) + \lambda \nabla F(\mathbf{x}) \nabla F(\mathbf{x})'$$
$$= \begin{bmatrix} B_{11} - \alpha_1 + \lambda\alpha_1^2 & B_{12} + \lambda\alpha_1\alpha_2 & \cdots B_{1n} + \lambda\alpha_1\alpha_n \\ B_{12} + \lambda\alpha_1\alpha_2 & B_{22} - \alpha_2 + \lambda\alpha_2^2 \cdots & B_{2n} + \lambda\alpha_2\alpha_n \\ \vdots & \vdots & \\ B_{1n} + \lambda\alpha_1\alpha_n & B_{2n} + \lambda\alpha_2\alpha_n & \cdots B_{nn} - \alpha_n + \lambda\alpha_n^2 \end{bmatrix}.$$

Here we have made use of the fact that a monotonic transformation (in this case ln) of a quasi-convex function is quasi-convex.

Besides the Problem 0′ referred to earlier, three additional separate estimation problems are distinguished for the quadratic function:[32]

Problem 1′:

$$Yt = \alpha_0 + \boldsymbol{\alpha}'\mathbf{Z}_t + \mathbf{Z}_t'(\mathbf{LDL}' - \boldsymbol{\alpha\alpha}')\mathbf{Z}_t + \epsilon_t, \qquad t = 1,\dots,T.$$

In this case, we seek parameters α_i's, L_{ij}'s $(j < i)$ and D_{ii}'s, without constraints.

Problem 2′:

$$Y_t = \alpha_0 + \boldsymbol{\alpha}'\mathbf{Z}_t + \mathbf{Z}_t'(\mathbf{LDL}' - \bar{\lambda}\boldsymbol{\alpha\alpha}')\mathbf{Z}_t + \epsilon_t, \qquad t = 1,\dots,T,$$

where $\bar{\lambda}$ is the largest positive scalar constant that the computer can recognize. In the second case we seek parameter α_i's, L_{ij}'s $(j < i)$ and D_{ii}'s, again without constraints.

Problem 3′:

$$Y_t = \alpha_0 + \boldsymbol{\alpha}^{*2\prime}\mathbf{Z}_t + \mathbf{Z}_t'(\mathbf{LD}^{*2}\mathbf{L}' - \lambda^{*2}\boldsymbol{\alpha\alpha}')\mathbf{Z}_t + \boldsymbol{\epsilon}_t, \qquad t = 1,\dots,T,$$

where $\alpha_i^{*^2} \equiv \alpha_i$ and $D_{ii}^{*^2} \equiv D_{ii}$, $\forall i$. We seek parameters α_i^*'s L_{ij}'s $(j < i)$, D_{ii}^*'s and λ^*. Again the problem is unconstrained. But the resultant estimates of α_i, D_{ii}, and λ will be non-negative. Hence $\hat{\alpha}_i \geqq 0$, and $(\hat{B} + \hat{\boldsymbol{\alpha}}\hat{\boldsymbol{\alpha}}')$ is positive semidefinite for a non-negative $\hat{\lambda}$. Under our specification the likelihood maximization problem is again equivalent to an unconstrained nonlinear least-squares problem.

4.2. *The Method of Maximum Likelihood*

The method of maximum likelihood is well known.[33] In order to contain a paper that is already too long, maximum likelihood equations will not be reproduced here. We do want to point out several possible approaches. First, if **x** is entirely exogenous and there is only one stochastic equation in the system, the nonlinear least-squares estimate is efficient for both Problems 1 and 2. Second, the estimator obtained by stopping after the first iteration of the Newton–Raphson process corresponds to the linearized maximum likelihood estimator proposed by Rothenberg and Leenders (1964), which is also an efficient estimator, provided that the initial estimator is consistent. Since consistent estimators are easy to obtain – for example, they can be obtained in this case from the unconstrained ordinary least-squares estimator of $\boldsymbol{\alpha}$ and $\mathbf{B}$ – of Problem 0 – computationally the linearized maximum likelihood estimator is quite attractive. Third, in the case that there are other equations in the system, $\mathbf{x}$ being still entirely exogenous, full information maximum likelihood will yield an efficient estimator. The full information maximum likelihood procedure is equivalent to an iterative nonlinear weighted least-squares procedure. As before, ordinary least-squares applied to Problem 0 will provide a set of consistent estimates which may be used to initialize the Newton–Raphson process, yielding the linearized maximum likelihood estimator, which is known to be efficient, after the first iteration. Finally, in the case that some or all of $\mathbf{x}$ is endogenous, one can either do the full maximum likelihood calculation for the system, or alternatively, one can first apply Amemiya's (1974) procedure to obtain a set of consistent estimates, and then using these as initial estimators, compute the linearized full information maximum likelihood estimator. The latter estimator has the same asymptotic distribution properties as full information maximum likelihood estimator but is much easier to compute since convergence is no longer required.

---

[32] Again, similar problems may be set up for other functional forms.

[33] See, for instance, Kendall and Stuart (1967, Vol. 2, pp. 35–74).

On a more practical level the following observations may be relevant. First, it can be verified by direct inspection that the maximum likelihood estimator of $\mathbf{B}$ obtained from substitution of an estimator of $\mathbf{LDL}'$ is independent of the ordering of the variables. For example: $B_{11}=0$ implies $L_{21}=0$ and hence $B_{12}=0$; but likewise $B_{22}=0$ implies $L_{21}=0$ and hence $B_{12}=0$. On the other hand $L_{21}=0$ implies $B_{12}=0$ but has no effect on B_{11} or B_{22}. Thus the ordering of $\mathbf{x}$ imposes no special restrictions on the form that $\mathbf{B}$ can assume other than that it is Cholesky factorizable. Second, it is not in general possible to test or impose the hypothesis of strict convexity of $F(\mathbf{x})$ or positive definiteness of the Hessian, say $\mathbf{B}$. The difficulty lies in that the maximum likelihood problem in which $D_{ii}>0$, $\forall i$, has only a least upper bound, which necessarily cannot be attained. In actual implementation, however, if one is interested in imposing strict convexity, one may set $D_{ii}\equiv D_{ii}^{*2}+\bar{\epsilon}$, where $\bar{\epsilon}$ is the smallest possible positive number that a given computer can recognize. Third, it is clear that if the Hessian of a twice differentiable function is positive definite at some $\mathbf{x}$, by continuity it will be positive semidefinite in an open neighborhood of $\mathbf{x}$. Thus, one can expand the region on which the Hessian is positive semidefinite, and possibly to include all of the data points, by the choice of a sufficiently large $\bar{\epsilon}$.[34] Finally, a special word should be said with regard to estimation subject to quasi-convexity constraints. Recall that quasi-convexity implies that

$$\mathbf{H}(\mathbf{x}) + \lambda \nabla F(\mathbf{x}) \nabla F(\mathbf{x})'$$

is positive semidefinite. For Problem 1′, it is evident that λ can be set equal to one without loss of generality. However, if one or more of the constraints turns out to be violated, that is, the estimated $D_{ii} < 0$ for at least one i, then one should set $\lambda = \lambda^{*2}$ as is done in Problem 3′ and estimate λ^* as an additional parameter.[35]

4.3. *Efficiency and Asymptotic Distribution Theory*

First, it is well known that under mild regularity conditions the maximum likelihood estimators are asymptotically efficient with the asymptotic variance–covariance matrix given by $-E([\partial^2 \ln L/\partial \mathbf{l}^2])^{-1}$ where $\ln L$ is the logarithm of the likelihood function. This matrix may be consistently estimated by $-[\partial^2 \ln L/\partial \mathbf{l}^2]^{-1}_{\mathbf{l}=\mathbf{l}^*}$, where $\mathbf{l}^*$ is the maximum likelihood estimator of $\mathbf{l}$.[36] For our problem, the regularity conditions are satisfied. Thus for both Problem 1 and Problem 2, the estimated asymptotic variance–covariance matrix may be directly computed.

However, it can be shown that, asymptotically, the estimated variance–covariance matrices of Problem 1 and Problem 2 converge to the same matrix, $-E([\partial^2 \ln L/\partial \mathbf{l}^2])^{-1}$. This is because, as Rothenberg (1966) has shown, the use of inequality constraints in maximum likelihood estimation does not increase the efficiency of the estimators.[37] Thus, for

[34] Thus in cases involving functions such as the transcendental logarithmic production function, which can only be made locally convex at some specific $\mathbf{x} = \mathbf{x}_0$, one can use this technique to enlarge the region of convexity. This idea is due to Mr. Yoichi Okita. However, there does not seem to be any optimality properties associated with this procedure. This basis of the procedure is related to the fact that given any real symmetric matrix $\mathbf{B}$ and positive scalar λ, $\mathbf{B} + \lambda\mathbf{I}$ can be made positive definite for a sufficiently large λ.

[35] The reason λ^{*2} should be re-introduced at this point is to offset partially the loss of estimable parameters as a result of a binding positive semidefiniteness constraint. If λ^{*2} is not re-introduced, one will have, in effect, constrained $\mathbf{B}$ itself to be positive semidefinite. It is possible that $\partial^2 \ln L/\partial \mathbf{l}^2$, where $\mathbf{l}$ now includes λ^* may become singular, in which case a generalized inverse should be used. See Section 4.4 below, especially Lemma 4.1.

[36] As noted below, $\partial^2 \ln L/\partial \mathbf{l}^2$ may become singular if the positive semidefiniteness constraint becomes binding. In that case the variance–covariance matrix of the estimable parameters consists of appropriate parts of the generalized inverse of $\partial^2 \ln L/\partial \mathbf{l}^2$.

[37] See Rothenberg (1966, pp. 51–55).

the purposes of obtaining efficient estimators, monotonicity and convexity constraints may be ignored; and any consistent estimator of $-[\partial^2 \ln L/\partial \boldsymbol{\Gamma}^2]^{-1}$ will do for the asymptotic variance–covariance matrix.

This is not to say that the monotonicity, convexity or quasi-convexity constraints are worthless. In most econometric applications, we require that the results are reasonable, that is, consistent with economic theory. Very often, the only way to ensure reasonableness is through the imposition of these constraints. Moreover, it is possible that the constrained estimators do better in finite samples. And a different criterion of optimality of the estimator, such as minimum expected mean squared error, may favor the constrained estimators over the unconstrained ones.

4.4. *Computational Notes*

Under our transformations the maximum likelihood problem becomes that of finding an unconstrained maximum for a nonlinear programming problem – in fact, a quadric programming problem. It is also equivalent to a nonlinear least-squares problem.

Eisenpress and Greenstadt (1966) and Eisenpress (1968) provide a maximum likelihood program that will accommodate both nonlinearities in variables and in parameters. Basically, a Marquardt–Levenberg algorithm is used, with the Newton–Raphson algorithm available as a special case. The latter is of particular interest because the first iteration of the Newton–Raphson algorithm produces the linearized maximum likelihood estimator proposed by Rothenberg and Leenders (1964).

Many other methods for unconstrained maximization are available. In general, however, a second derivative method should be used because the asymptotic variance–covariance matrix must be estimated by the negative of the Hessian matrix of the natural logarithm of the likelihood function at the point of convergence. Murray (1972) provides a comprehensive survey of alternative unconstrained maximization methods.[38]

A special feature of the positive semidefiniteness constraints as indicated by the following lemma imposes additional requirements on the algorithm:

---

[38] See Murray (1972, especially Chs. 3 and 4).

Lemma 4.1. The matrix product $\mathbf{LDL}'$ where $\mathbf{L}$ is unit lower triangular and $\mathbf{D}$ is diagonal is independent of the ith column of $\mathbf{L}$ if D_{ii} is zero.

Proof: The proof is by direct computation.

$$\begin{aligned}
\mathbf{LDL}' &= \begin{bmatrix} 1 & & \\ L_{21} & 1 & 0 \\ L_{31} & L_{32} & \\ \vdots & & \ddots \\ L_{n1} & L_{n2} & 1 \end{bmatrix} \begin{bmatrix} 0 & D_{11} & 0 & \\ & D_{22} & & \\ & 0 & \ddots & \\ & & & D_{nn} \end{bmatrix} \begin{bmatrix} 1 & L_{21} & L_{31} & L_{n1} \\ & 1 & L_{32} & L_{n2} \\ & 0 & \ddots & \\ & & & 1 \end{bmatrix} \\
&= \begin{bmatrix} D_{11} & & \cdots 0 \\ L_{21}D_{11} & D_{22} & \cdots 0 \\ L_{31}D_{11} & L_{32}D_{22} & D_{33}\cdots 0 \\ L_{n1}D_{11} & L_{n2}D_{22} & \cdots D_{nn} \end{bmatrix} \begin{bmatrix} 1 & L_{21} & L_{31} & L_{n1} \\ & 1 & L_{32} & L_{n2} \\ 0 & & \ddots & \\ & & 1 & \end{bmatrix} \\
&= \begin{bmatrix} D_{11} & L_{21}D_{11} & L_{31}D_{11} & \cdots L_{n1}D_{11} \\ L_{21}D_{11} & L_{21}^2D_{11} + D_{22} & L_{21}L_{31}D_{11} + L_{32}D_{22} & L_{21}L_{n1}D_{11} + L_{n2}D_{22} \\ \vdots & \vdots & \vdots & \\ L_{n1}D_{11} & L_{21}L_{n1}D_{11} + L_{n2}D_{22} & & L_{n1}^2D_{11} + L_{n2}^2D_{22} + \cdots + D_{nn} \end{bmatrix}
\end{aligned}$$

It is evident that if $D_{ii} = 0$, the values of $L_{ji}, \forall j$, do not matter at all, because all the L_{ji} entries are multiplied by D_{ii}. Q.E.D.

Lemma 4.1 implies that if one knows $D_{ii} = 0$, one can, without loss, replace the ith column of $\mathbf{L}$ by zeroes everywhere off the diagonal. By definition, $L_{ii} = 1$.

What this means is that if one of the positive semidefiniteness constraints turns out to be binding, the matrix $[\partial^2 \ln L/\partial \mathbf{l}^2]$ will always become singular. Hence the usual Newton–Raphson iterative method or any one of the second-derivative modified Newton methods fails. There are various ways to remedy this situation. The most elegant method is to continue the Newton–Raphson process using a generalized inverse of $[\partial^2 \ln L/\partial \mathbf{l}^2]$ until the process converges.[39] Conditions for convergence of the Newton–Raphson process using a generalized inverse are given by Ben-Israel (1966).

---

[39] For discussion of generalized inverses, see Rao and Mitra (1971). Rao and Mitra also propose the use of the generalized inverse when the information matrix becomes singular (pp. 201–203).

5. Testing of Hypotheses

5.1. *Introduction*

The basic statistical problem in testing the hypotheses of monotonicity, and/or convexity can be reduced to the following: the null hypothesis to be considered is of the type

$$H_0: \boldsymbol{\alpha} \geqq 0,$$

and the alternative hypothesis is

$$H_a: \boldsymbol{\alpha} \ngeqq 0,$$

that is, at least one $\alpha_i < 0$. If α is one-dimensional, the usual one-tailed t-test will work. However, the usual likelihood ratio test procedure based on asymptotic distribution theory breaks down because under the null hypothesis there is no reduction in the dimensionality of the space of possible parameters of the likelihood function. An alternative test procedure is needed. We shall examine three alternatives: Bonferroni t-statistics; distribution of extreme values; and finally a test procedure based on the likelihood ratio. The Bonferroni t-statistics is recommended because of its easy computability and the ready availability of tables.

To test these hypotheses, one needs to use the estimates from Problem 1, that is, the unconstrained estimates. An alternative approach is possible: let $\mathbf{B}$ be a maximum likelihood estimate of B of the original untransformed model, that is, Problem 0. For convenience we shall stack the parameters of $\mathbf{B}$ into one vector, of dimension $n(n+1)/2$,

$$\boldsymbol{\beta} \equiv [B_{11}B_{12}\cdots B_{1n}B_{22}B_{23}\cdots B_{nn}]',$$

similarly,

$$\mathbf{l} \equiv [D_{11}L_{21}L_{31}\cdots L_{n1}D_{22}L_{32}\cdots D_{nn}]'.$$

Defining the matrix

$$\frac{\partial \boldsymbol{\beta}}{\partial \mathbf{l}} = \begin{bmatrix} \frac{\partial \beta_1}{\partial l_1} & \frac{\partial \beta_2}{\partial l_1} \cdots & \frac{\partial \beta_{n^*}}{\partial l_1} \\ \frac{\partial \beta_1}{\partial l_2} & \frac{\partial \beta_2}{\partial l_2} \cdots & \frac{\partial \beta_{n^*}}{\partial l_2} \\ \vdots & \vdots & \vdots \\ \frac{\partial \beta_1}{\partial l_{n^*}} & \frac{\partial \beta_2}{\partial l_{n^*}} \cdots & \frac{\partial \beta_{n^*}}{\partial l_{n^*}} \end{bmatrix},$$

that is, the Jacobian of the transformation from $\boldsymbol{\beta}$ into $\mathbf{l}$, we have, at the point of maximum likelihood:

$$\left[\frac{\partial^2 \ln L}{\partial \mathbf{l} \partial \mathbf{l}'}\right] = \left[\frac{\partial \boldsymbol{\beta}}{\partial \mathbf{l}}\right]\left[\frac{\partial^2 \ln L}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}\right]\left[\frac{\partial \boldsymbol{\beta}}{\partial \mathbf{l}}\right]'.$$

Thus an estimator of $V(\hat{\mathbf{l}})$, which is given by

$$-E\left(\left[\frac{\partial^2 \ln L}{\partial \mathbf{l} \partial \mathbf{l}'}\right]\right)^{-1},$$

the variance–covariance matrix of $\hat{\mathbf{l}}$, may be computed given a knowledge of

$$\left[\frac{\partial^2 \ln L}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}\right].$$

One can then extract from $V(\hat{\mathbf{l}})$ those components which correspond to the D_{ii}'s and use them as a basis for statistical inference. The disadvantage of this approach, however, is that there may be cases in which one may not be able to obtain D_{ii}'s from $\hat{\mathbf{B}}$, that is, $\hat{\mathbf{B}}$ may not be Cholesky factorizable. In that case, one will have to fall back on the solution given by Problem 1.[40]

The testing of the hypotheses of monotonicity and quasiconvexity is slightly more complicated. The difficulty lies in the fact that it is sometimes not possible to reject conclusively that there may exist a sufficiently large positive scalar constant λ such that $(\mathbf{B} + \lambda \boldsymbol{\alpha}\boldsymbol{\alpha}')$ is positive semidefinite. The null hypothesis to be considered is: $\exists \lambda, \lambda > 0$, such that corresponding to that λ,

$$\alpha_i \geqq 0, \qquad D_{ii} \geqq 0, \qquad \forall i.$$

Obviously, for each λ, one can carry out a test of this type; however, each rejection is not conclusive, because there always remains the possibility (only a possibility) that for some larger λ, the hypotheses may not be rejected.

[40] Since not all real symmetric matrices are Cholesky factorizable, a legitimate question at this point is whether the power of our test of positive semidefiniteness of $\mathbf{B}$ is reduced by restricting our consideration to the class of real symmetric matrices that are Cholesky factorizable. The answer is no if the distribution function of the errors in the equation is continuous and smooth everywhere, as in the case of the normal distribution since the set of real symmetric matrices which are not Cholesky factorizable is of measure zero. Heuristically, the situation is comparable to that of maximizing a likelihood function with respect to a single parameter α and subject to the restriction that $\alpha \neq 0$. The resultant distribution of $\hat{\alpha}$ is essentially the same as that of an estimator derived without the restriction $\alpha \neq 0$. Constrained estimation, on the other hand, requires no such justification since all positive semidefinite real symmetric matrices are Cholesky factorizable.

Our proposed solution is to decompose the test procedure into two stages. First, we make use of Theorem 2.5 and Theorem 3.4 which together imply that a necessary and sufficient condition that there exists a vector $\mathbf{a}$ such that $\mathbf{x}'\mathbf{A}\mathbf{x} \geqq 0$ whenever $\mathbf{a}'\mathbf{x} = 0$ is that the number of non-negative Cholesky values of $\mathbf{A}$, an $n \times n$ Cholesky factorizable real symmetric matrix, must be greater than or equal to $(n-1)$. Thus, if the Hessian matrix is Cholesky factorizable, this implies that there are at least $(n-1)$ non-negative Cholesky values (or alternatively at most one negative Cholesky value) everywhere. This provides a basis for testing the hypothesis of quasi-convexity at the point of approximation. However, because one has no knowledge *a priori* which one of the n Cholesky values will turn out to be negative, if any, this theorem does not lead to a corresponding procedure for constrained estimation. We therefore test the hypothesis that $\mathbf{B}$ has at least $(n-1)$ non-negative Cholesky values simultaneously with the hypothesis that $\alpha_i \geqq 0, \forall i$. Thus

$$\begin{aligned} &H_0{:}\,\alpha_i \geqq 0, && \forall i, \\ &D_{ii} \geqq 0, && \forall i, \text{ except possibly one.} \end{aligned}$$

There are three possible outcomes with respect to the D_{ii}'s. First, we do not reject that $D_{ii} \geqq 0, \forall i$. This implies that not only can we not reject the hypothesis of quasi-convexity, but also that of convexity as well. Second, we do not reject the hypothesis that $D_{ii} \geqq 0$ for only $(n-2)$ i's. This implies that the function cannot possibly be quasi-convex at the point of approximation, by virtue of Theorem 2.5. Third, we do not reject the hypothesis that $D_{ii} \geqq 0$ for only $(n-1)$ i's. This leaves open the possibility that one may find a λ and $\mathbf{D}$ such that $\mathbf{LDL}' = (\mathbf{B} + \lambda \boldsymbol{\alpha}\boldsymbol{\alpha}')$ with $D_{ii} \geqq 0, \forall i$, that is, the function may still be quasi-convex.

Since the computer can only handle finite arithmetic, it is not possible to check through all possible λ's. The next best alternative is therefore to set $\lambda = \bar{\lambda}$, where $\bar{\lambda}$ is the largest positive scalar constant that the computer can recognize, and to test the hypothesis that the resultant D_{ii}'s are all non-negative. This constitutes the second stage of the test procedure.[41]

---

[41] After this paper was essentially completed, Jorgenson and Lau (1975b) have proposed a procedure, also based on the Cholesky factorization, that avoids this inconclusiveness.

5.2. Bonferroni t-Statistics

Suppose that one is interested in testing the null hypothesis

$$H_0: \alpha_i \geqq 0, \qquad i = 1,\dots,n \quad (n \neq 1),$$

against the alternative hypothesis that at least one $\alpha_i < 0$, a natural way to proceed is to use n one-tailed t-statistics. However in order to control for the overall level of significance, the level of significance of each one-tailed test must be scaled down accordingly. If the desired overall level of significance is set at α, then the individual levels of significance α^i must satisfy the following Bonferroni inequality

$$1 - \alpha \geqq 1 - \alpha^1 - \alpha^2 - \cdots - \alpha^n.$$[42]

In the case that the α_i's are distributed independently of each other, we actually have

$$(1 - \alpha) = \prod_{i=1}^{n} (1 - \alpha^i).$$

Let $V(\hat{\alpha})$ be the estimated asymptotic variance–covariance matrix of the estimator of α, so that

$$T_i = \frac{\hat{\alpha}_i}{(V_{ii}(\hat{\alpha}))^{1/2}}, \qquad i = 1,\dots,n,$$

is distributed asymptotically as Student's t with infinite degrees of freedom. Let $t_{\infty}^{\alpha/n}$ be the upper α/n percentile points of the t-distribution with infinite degrees of freedom. Then with probability greater than or equal to $(1 - \alpha)$, simultaneously,

$$0 \leqq \hat{\alpha}_i + t_{\infty}^{\alpha/n}(V_{ii}(\hat{\alpha}))^{1/2}, \qquad i = 1,\dots,n.$$

We note that the distribution of Student's t with infinite degrees of freedom is precisely the uninormal distribution. Thus one can equivalently define the intervals above in terms of a uninormal distribution. For each component interval above, the level of significance is set at α/n. Equal significance levels can be abandoned, and unequal allocation substituted. Any combination of significance levels summing to α will produce the same bound α for the probability error rate. The reader is referred to Miller (1966) for further details.

Thus, one can apply the Bonferroni t statistics to construct simultaneous rejection regions for the monotonicity hypothesis, which for both the quadratic function (at $\mathbf{x} = [0]$) and the transcendental logarithmic function (at $\mathbf{x} = [1]$) amounts to $\alpha_i \geqq 0, \forall i$; and for the convexity and quasi-convexity hypotheses which for both the quadratic and transcendental logarithmic function amount to respectively $D_{ii} \geqq 0, \forall i$, and $D_{ii} \geqq 0, \forall i$, except possibly one, for the Cholesky factorization of appropriate matrices. For the generalized linear function, one can test for global monotonicity and convexity simultaneously,

---

[42] Superscripts are used to distinguish the levels of significance from the parameters.

$$\alpha_i \leqq 0, \qquad B_{ij} \leqq 0, \qquad i \neq j, \quad \forall i,j.$$

Finally, it should be noted that in the case of interval constraints, such as $\bar{\alpha}_i \geqq \alpha_i \geqq \underline{\alpha}_i, \forall i$, a similar Bonferroni t statistics procedure resulting in simultaneous intervals of the type

$$\bar{\alpha}_i + t_{\infty}^{\alpha/2n}(V_{ii}(\hat{\alpha}))^{1/2} \geqq \hat{\alpha}_i \geqq \underline{\alpha}_i - t_{\infty}^{\alpha/2n}(V_{ii}(\hat{\alpha}))^{1/2}, \quad \forall i,$$

will apply.

5.3. *Distribution of Extremes*

A second alternative, which is the original proposal made by Lau (1974), makes use of the theory of ordered statistics. Again consider the null hypothesis

$$H_0 : \alpha_i \geqq 0, \qquad i = 1, \ldots, n.$$

This null hypothesis may be transformed to the following one:

$$H_0 : \min_i \{\alpha_1, \alpha_2, \ldots, \alpha_n\} \geqq 0.$$

Now given a known joint distribution of α, one can presumably derive the distribution of the minimum element of α, making use of the techniques of the theory of ordered statistics.[43] Knowing the distribution of $\min_i\{\alpha_i\}$, one can then immediately construct a rejection region for the hypothesis that $\min_i\{\alpha_i\} \geqq 0$ for any given level of significance. The advantage of this test procedure is that it is exact, unlike the Bonferroni t-statistics procedure, which only gives a bound on the overall level of significance and it is in principle quite feasible. The disadvantage of this procedure, is, of course, that elaborate computations are required, even in the case that $V(\hat{\alpha})$ is diagonal.[44] In general, a multivariate normal integral must be evaluated by numerical methods as appropriate tables do not yet exist for $n \geqq 3$. To make matters worse, $V(\hat{\alpha})$ is in general unknown and only an estimate is available. Thus in computing the distribution of $\min_i\{\hat{\alpha}\}$ one will have to integrate over the distribution of $V(\hat{\alpha})$ as well. This all seems to be an extremely high price to pay just for obtaining an exact level of significance. However, this may be the only possible procedure if the Bonferroni t-statistics procedure does not give clearcut results; for example, if zero is on the boundary of the rejection region.

---

[43] For an excellent introductory exposition to the theory of ordered statistics, see Kendall and Stuart (1969, Vol. 1, Ch. 14, pp. 325–346).

For the null hypothesis

$$H_0{:}\alpha_i \geqq 0, \quad \forall i, \quad \text{except possibly one.}$$

The appropriate transformation is the following:

$$H_0{:}\text{second } \underset{i}{\text{smallest}}\{\alpha_1,\alpha_2,\ldots,\alpha_n\} \geqq 0.$$

Again, this problem may in principle be solved by making use of the theory of ordered statistics. Actual computation is of course a different matter.

5.4. *Likelihood Ratio Tests*

It is well known that a null hypothesis of the type $\alpha_i = 0, i = 1,\ldots,n$, may be tested by the likelihood ratio procedure. Essentially, $-2 \ln \lambda$, where λ is the ratio of the maximized likelihood function under the null hypothesis to the unconstrained maximized likelihood function, is asymptotically distributed as a χ^2 variable with n degrees of freedom. However, the same is not true when the null hypothesis consists of inequalities rather than equality constraints. Here there is no reduction in the dimensionality of the feasible parameter space under the null hypothesis. Asymptotically, the value of the inequality constrained maximized likelihood function will converge to the value of the unconstrained maximum likelihood function and $-2 \ln \lambda$ is identically zero.

Despite this shortcoming, the use of the likelihood ratio procedure in this context does have a certain amount of intuitive appeal. If the unconstrained estimates satisfy the constraints, the likelihood ratio will be identically one. Otherwise, it will be less than one. A rejection region may be constructed on the basis of λ. Consider for the sake of simplicity a one-dimensional example. We want to test the hypothesis of $\alpha \geqq 0$. Let $\hat{\alpha}$ be the unconstrained estimator of α. The likelihood ratio is identically one if $\hat{\alpha} \geqq 0$. The likelihood ratio is given by

---

[44] Note that if $V(\hat{\alpha})$ is indeed diagonal, it implies that the components of $\hat{\alpha}$ are distributed independently of one another, and the Bonferroni t-statistics procedure is exact.

$$\lambda = \frac{L(\alpha = 0)}{L(\hat{\alpha})}, \qquad \hat{\alpha} < 0,$$

which will be less than one and decreases monotonically as $\hat{\alpha}$ decreases. (Of course, in practice, we use the one-tailed t-test for this case.)

However, in order to make use of the likelihood ratio in this manner it is necessary to compute either its exact distribution or at least a finite sample approximation to the exact distribution. Unfortunately, the finite sample distribution will most likely depend on the values of both the dependent and independent variables, and will have to be computed on a case by case basis. Further research on the finite sample distribution theory of inequality-constrained estimators is needed before the likelihood ratio procedure can be fruitfully employed.

6. Conclusion

In this paper we have outlined an implementable procedure for testing the hypotheses of monotonicity, convexity and quasi-convexity of estimated functions, and to obtain parametric estimates of functions under the constraints of these hypotheses. Since these hypotheses are of fundamental importance in economic analysis, it is essential to test their validity under as unrestrictive a maintained hypothesis as possible. Functions which provide second-order numerical approximation to arbitrary functions are therefore well suited for this purpose. The availability of a simple procedure for constrained estimation to ensure monotonicity, convexity, or quasi-convexity also enhances immensely the usefulness of the new functional forms in practical economic applications. The proposed procedure may be applied to any functional form with the second-order numerical approximation property, including the class of general linear profit functions introduced by McFadden (Chapter II.2) and others yet to be invented.

The procedures considered here may be easily extended to solve other types of problems: linear inequality constraints, interval constraints, estimation of variance–covariance matrices (which must be positive semidefinite) and transition probability matrices (which must be nonnegative and have column sums equal to ones), and estimation of saddle functions (convex-concave or concave-convex functions).