

Ronning, Gerd

Working Paper

Linear and nonlinear dirichlet share equations models

Diskussionsbeiträge - Serie II, No. 80

Provided in Cooperation with:

Department of Economics, University of Konstanz

Suggested Citation: Ronning, Gerd (1989) : Linear and nonlinear dirichlet share equations models, Diskussionsbeiträge - Serie II, No. 80, Universität Konstanz, Sonderforschungsbereich 178 - Internationalisierung der Wirtschaft, Konstanz

This Version is available at:

<https://hdl.handle.net/10419/101798>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

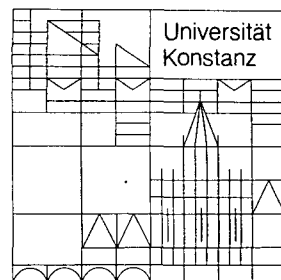
Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Sonderforschungsbereich 178
„Internationalisierung der Wirtschaft“

Diskussionsbeiträge



Juristische
Fakultät

Fakultät für Wirtschafts-
wissenschaften und Statistik

Gerd Ronning

Linear and Nonlinear
Dirichlet Share Equations Models

6. JULI 1989 Weltwirtschaft
Kiel

W 113-80

014 039 341

LINEAR AND NONLINEAR DIRICHLET SHARE EQUATIONS MODELS

Gerd Ronning

Serie II - Nr. 80

Juni 1989

LINEAR AND NONLINEAR DIRICHLET SHARE EQUATIONS MODELS

Gerd Ronning*

Universität Konstanz

P.O. Box 5560, D-7750 Konstanz

Abstract

An adequate stochastic model for shares as dependent variables is provided by the Dirichlet distribution. The paper considers two different parametrizations which lead to linear and nonlinear Dirichlet share equations. Using an inequality for the trigamma function the global concavity of the likelihood function for the nonlinear case is shown. The same inequality is employed in proving positive definiteness of the information matrix for the linear case. Suitability of the Dirichlet specification in econometric demand systems (such as AIDS and Translog) is discussed.

Keywords: Dirichlet distribution, demand systems, trigamma function.

1 Introduction

Descriptive analysis by means of shares or percentages is a frequent statistical task in applied research. However, adequate stochastic specifications have not received

*Research is related to project A2 of SFB 178.

much attention. The state of the art is described very extensively by Aitchison (1982). Econometricians in particular have for a long time analysed budget shares of households and cost shares of firms by means of models which are derived from microeconomic theory and which result in systems of demand equations. See, for example, Deaton and Muellbauer (1980). Almost every author has assumed that shares are normally distributed which does not take into account that shares can vary only within the interval $[0,1]$. The only noteworthy exception is Woodland (1979) who employs the Dirichlet distribution to a set of linear share equations.

In this paper we compare Woodland's linear specification with a nonlinear specification which appears in demand systems such as the *translog* approach. For the latter approach the simple inequality involving the trigamma function ψ_1 ¹

$$\psi_1(x) < \alpha \psi_1(\alpha x) \quad , \quad x > 0, \quad 0 < \alpha < 1 \quad (1-1)$$

(see Ronning 1986) will be used to show that the likelihood function is globally concave. This generalizes a result of Ronning (1989) where numerical aspects of maximum likelihood estimation are discussed, as well. We also use inequality (1-1) to show that for the linear case for which the Hessian matrix depends on the observed shares the information matrix is positive definite. This can be seen as an indication of global identification. See Rothenberg (1971).

The paper is organized as follows: Section 2 considers linear and nonlinear specifications of Dirichlet share equations models. In section 3.1 we prove global concavity for the nonlinear specification by an approach which partly makes use of Dhrymes's (1978) proof for the multinomial logistic model. An alternative (direct) proof is given in appendix A. Section 3.2 considers the linear case which is much more cumbersome to deal with. A rather huge formal effort is necessary in order to prove that the information matrix for this model is positive definite. Therefore most details are treated in three appendices (B - D). Section 4 adds some remarks

¹The trigamma function is the second derivative of the log of the gamma function. See Abramowitz and Stegun (1965, chapter 6).

about reasonableness and feasibility of the two approaches with special emphasis on estimation of demand systems.

2 Dirichlet Share Equations Models

2.1 Econometric Share Equations

For illustrative purpose let us consider the case of budget shares in econometric demand analysis. Let s_{it} be the share of budget (disposable income) in month t which is spent for good i . The shares satisfy $0 \leq s_{it} \leq 1, i = 1, \dots, k$, and $\sum_i s_{it} = 1$ for all t . Then a **nonlinear** (deterministic) share equation as considered in this paper is given by

$$s_{it} = \frac{\alpha_i + \beta_i x_t}{\sum_{j=1}^k \alpha_j + \beta_j x_t} \quad (2-1)$$

and a **linear** (deterministic) share equation is given by

$$s_{it} = \alpha_i + \beta_i x_t \quad (2-2)$$

where α_i and β_i denote unknown parameters and x_t is the value of an explanatory variable for observation t ². For the collection of all k goods we call (2-1) and (2-2), respectively, a *demand system*. (2-1) has the typical form of a demand system derived from *flexible functional forms* (Christensen, Jorgenson and Lau 1975 p.370) and (2-2) depicts the *Almost Ideal Demand System* by Deaton and Muellbauer (1980 p. 313)³. In both cases the explanatory variables are prices of all k goods and the income of household in logarithmic form. Note that it is implicitly assumed that either the x_t or the parameters (or both) are such that the s_{it} of (2-1) and (2-2) satisfy the share restrictions⁴. We also remark that (2-2) is derived from (2-1) by

²A more general specification involving two and more explanatory variables is straightforward. See the discussion in section 3.

³We consider the approximate version which is mostly used in applied work. See Deaton and Muellbauer (1980, p. 316).

⁴This problem could be avoided by using a logistic specification which however so far has not become popular in econometric demand analysis. See Considine and Mount (1984).

the imposition of the following parameter constraints:

$$\sum_{j=1}^k \alpha_j = 1 \quad \text{and} \quad \sum_{j=1}^k \beta_j = 0 \quad . \quad (2-3)$$

2.2 Stochastic Share Equations

We assume that for each t the observed shares for the k categories are generated by a Dirichlet distribution with parameters θ_{it} . If we denote the observed shares by y_{it} then the joint density function is given by

$$f(y_{1t}, y_{2t}, \dots, y_{kt}) = \Gamma \left(\sum_{j=1}^k \theta_{jt} \right) \prod_{i=1}^k \frac{y_{it}^{\theta_{it}-1}}{\Gamma(\theta_{it})} \quad (2-4)$$

with $\theta_{it} > 0$ and $y_{it} \geq 0$ for all i , $\sum_{j=1}^k y_{jt} = 1$ and Γ denotes the gamma function.

We now let the parameters θ_{it} depend linearly on the explanatory variable x_t :

$$\theta_{it} = \alpha_i + \beta_i x_t \quad (2-5)$$

where again it is assumed that $\theta_{it} > 0$. Then the expected value of the stochastic share Y_{it} is given by

$$E(Y_{it}) = \frac{\alpha_i + \beta_i x_t}{\sum_{j=1}^k \alpha_j + \beta_j x_t} \quad (2-6)$$

(see, for example, Johnson and Kotz 1972 p.233), that is, the expected value of Y_{it} has the form of the (deterministic) nonlinear share equation in (2-1). Following Woodland (1979) a linear specification can be derived from

$$\theta_{it} = c(\alpha_i + \beta_i x_t) \quad , \quad c > 0 \quad (2-7)$$

together with the parameter restrictions in (2-3) which imply that the additional parameter c satisfies $c = \sum \theta_j$. Therefore we have under (2-3) and (2-7)

$$E(Y_{it}) = \alpha_i + \beta_i x_t \quad . \quad (2-8)$$

Note that the use of (2-7) (instead of (2-5)) implies the flexible scale factor $1/(1+c)$ (instead of $1/2$) in the covariance matrix of the random variables Y_{it} for a certain t ⁵. In (2-8) we implicitly assume that $0 < \alpha_i + \beta_i x_t < 1$ holds.

⁵The covariance matrix of the Dirichlet distribution is given, for example, in Johnson and Kotz (1972 p.233).

3 Global Concavity of Likelihood Function and Global Identification

It will turn out in this section that the analysis of the nonlinear case is much more straightforward. We therefore consider this case first and show that the likelihood function for this model is globally concave since the Hessian matrix is negative definite. This also implies that the information matrix has full rank. For the linear case such an analysis of the likelihood function is not possible since the Hessian matrix depends on the outcomes of the dependent variables. However, we can show that the information matrix for this model is positive definite, too.

Since the (continuous) Dirichlet distribution and the (discrete) multinomial distribution are closely related it is not surprising that the matrix of second-order partial derivatives for the nonlinear share equations model is similar in structure to the multinomial logistic case. We therefore in our proof make use of a result which has been obtained by Dhrymes (1978) in connection with the multinomial logistic model. A second proof which is self-contained is given in appendix A.

In section 3.2 we then show that the information matrix for the linear model is positive definite. In style our proof resembles that of Deprins and Simar (1985) for linear models with gamma disturbances although another result from the trigamma function (our inequality (1-1)) is applied and the dimension of the parameter space makes the analysis more complicated.⁶ Rothenberg (1971) has shown that a positive definite information matrix guarantees global identification of the parameters which is therefore determined for both models.

⁶Deprins and Simar consider only one equation whereas we have to consider $k - 1$ equations.

3.1 The Nonlinear Case: Global Concavity

In this subsection we consider the nonlinear Dirichlet share equations. Starting from density function (2-4) and using (2-5) under random sampling the loglikelihood function is given by

$$L = \sum_{t=1}^n \left[\log(\Gamma(S_t)) + \sum_{j=1}^k [(\alpha_j + \beta_j x_t - 1) \log(y_{jt}) - \log(\Gamma(s_{jt}))] \right] \quad (3-1)$$

where s_{jt} is given by (2-1) and $S_t = \sum_{j=1}^k s_{jt}$. We show in appendix A that the matrix of second-order partial derivatives (Hessian matrix) with respect to the parameter vector $(\alpha_1, \dots, \alpha_k, \beta_1, \dots, \beta_k)$ is given by

$$H = - \sum_{t=1}^n B_t \otimes x_t x_t' \quad (3-2)$$

where \otimes denotes the Kronecker product, and $x_t = (1, x_t)'$. The $(k \times k)$ matrix B_t has the following form:

$$B_t = D_t - q_t \iota \iota' \quad (3-3)$$

where D_t is a diagonal matrix with diagonal elements $d_{it} = \psi_1(\alpha_i + \beta_i x_t)$, ι is a k -dimensional vector of ones and $q_t = \psi_1(\sum \alpha_j + x_t \sum \beta_j)$. Note that the matrix (3-2) does not depend on the observed values y_{it} .

The Hessian matrix (3-2) is similar in structure to the one for the multinomial logistic model. Dhrymes (1978 p. 350) proved global concavity for that model under the assumption that the regressor matrix X has full column rank⁷. He showed that a sufficient condition for negative definiteness of that matrix is provided by showing that B_t is positive definite. Therefore we have to prove

Lemma 1 *Let $\alpha_i + \beta_i x_t > 0$ for $i = 1, \dots, k$. Then B_t in (3-3) is a positive definite matrix.*

⁷Dhrymes considers the case of an arbitrary number of, say, r explanatory variables. In our notation each row of the regressor matrix X would consist of the row vector $x_t = (1, x_{1t}, x_{2t}, \dots, x_{rt})$. This extended vector x_t used in (3-2) would give the Hessian matrix for the case of r explanatory variables instead of just one (beside the constant term).

Proof:⁸ First note that from inequality (1-1) it follows that

$$\psi_1(S_t) < \frac{\theta_{it}}{S_t} \psi_1(\theta_{it}) \quad (3-4)$$

holds for all i . We now consider the j -th principal submatrix of B_t denoted by B_{jt} . For this matrix we obtain $B_{jt} = D_{jt} - q_t E_j$ where D_{jt} contains the first j diagonal elements of D_t and E_j is a $(j \times j)$ matrix of ones. Therefore the j -th principal minor of B_t satisfies

$$\begin{aligned} \det(B_{jt}) &= \left(1 - \sum_{\ell=1}^j \frac{\psi_1(S_t)}{\psi_1(\theta_{\ell t})} \right) \prod_{\ell=1}^j \psi_1(\theta_{\ell t}) \\ &> \left(1 - \frac{1}{S_t} \sum_{\ell=1}^j \theta_{\ell t} \right) \prod_{\ell=1}^j \psi_1(\theta_{\ell t}) \\ &\geq 0 \end{aligned}$$

where the first line follows from, e.g., Graybill (1983, theorem 8.4.3), the second line uses inequality (3-4) and the third line is based on the property $\psi_1(z) > 0$ for all $z > 0$ (Abramowitz and Stegun 1965, (6.4.10)) and the nonnegativity of the term in parentheses. Since the inequality holds for all j the matrix B_t is positive definite. \square

Theorem 1 *Assume that there are at least two linearly independent vectors among the vectors $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ and that $\alpha_i + \beta_i x_t > 0$ holds for $i = 1, \dots, k$ and $t = 1, \dots, k$. Then the Hessian matrix (3-2) is negative definite and therefore the likelihood function for the nonlinear Dirichlet share equations model is globally concave.*

Proof: See Dhrymes (1978 p. 350) where lemma 1 from this paper has to be substituted in order to prove that the matrix B_t is positive definite. Note that for the general case of r explanatory variables the only modification in the theorem would be that at least $r + 1$ linear independent vectors \mathbf{x}_t should exist. \square

Finally we should mention that in appendix A (subsection A.2) a different formulation of the Hessian matrix (3-2) is used to prove theorem 1 directly.

⁸The proof is almost identical to the one used in Ronning (1989) where no explanatory variables were considered, that is all β_i were set equal to zero.

3.2 The Linear Case: Positive Definiteness of Information Matrix

The loglikelihood function for the linear case is obtained from density function (2-4) with parameters θ_{it} given by (2-7) where α_j and β_j satisfy restrictions (2-3):

$$\begin{aligned}
 L &= n \log(\Gamma(c)) + \sum_{t=1}^n \sum_{j=1}^k ((\theta_{jt} - 1) \log(y_{jt}) - \log(\Gamma(\theta_{jt}))) \\
 &= n \log(\Gamma(c)) + c \sum_{t=1}^n \sum_{j < k} (\alpha_j + \beta_j x_t) \log(y_{jt}/y_{kt}) + c \sum_{t=1}^n \log(y_{kt}) \\
 &\quad - \sum_{t=1}^n \sum_{j=1}^k \log(y_{jt}) - \sum_{t=1}^n \sum_{j < k} \log(\Gamma(c(\alpha_j + \beta_j x_t))) \\
 &\quad - \sum_{t=1}^n \log(\Gamma(c(1 - \sum_{j < k} \alpha_j - x_t \sum_{j < k} \beta_j)))
 \end{aligned} \tag{3-5}$$

From results in appendix B the following form of the information matrix Ω with respect to the parameters $\alpha_1, \beta_1, \dots, \alpha_{k-1}, \beta_{k-1}, c$ is obtained:

$$\Omega = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}$$

where Ω_{11} has $2(k-1)$ rows and columns, respectively, and Ω_{22} is a scalar which we will henceforth denote by ω . We then have for the different submatrices of Ω :⁹

$$\Omega_{11} = c^2 ((X \otimes I_{k-1})'(V + (W_0 \otimes \iota'))(X \otimes I_{k-1})$$

where I_r denotes the $(r \times r)$ identity matrix. Since V and W_0 are positive diagonal matrices, the submatrix Ω_{11} is positive definite if X has full column rank. Furthermore we have

$$\Omega_{12} = c(X \otimes I_{k-1})'(Vu - (W_0 \otimes I_{k-1})(d \otimes \iota)) = \Omega'_{21}$$

Finally the scalar term $\Omega_{22} = \omega$ is given by

$$\omega = -(n \psi_1(c) - u'Vu - d'W_0d) \tag{3-6}$$

⁹The exact definition of the vectors u , d and the matrices X , V , W_0 is given in appendix B. They are not repeated here in order to save space and not to obscure the sketch of the proof.

Since Ω_{11} is positive definite we must only show that the determinant of Ω satisfies:

$$\det(\Omega) > 0 \quad (3-7)$$

For the determinant of Ω we can write:

$$\det(\Omega) = \omega \det\left(\Omega_{11} - \frac{1}{\omega} \Omega_{12} \Omega_{21}\right)$$

and the matrix in brackets can be written as

$$\Omega_{11} - \frac{1}{\omega} \Omega_{12} \Omega_{21} = c^2 (X \otimes I_{k-1})' B (X \otimes I_{k-1}) \quad (3-8)$$

with

$$B = Z - \frac{1}{\omega} h h' \quad (3-9)$$

where $Z = V + (W_0 \otimes \iota \iota')$ and $h = Vu - (W_0 \otimes I_{k-1})(d \otimes \iota)$.

Please note that

$$B^{-1} = Z^{-1} + \frac{1}{\omega - h' Z^{-1} h} h h' \quad (3-10)$$

where a well-known matrix inversion lemma has been applied (see, e.g., Rao 1973 p. 33).

We now use the following two lemmas which are proved in appendices C and D, respectively.

Lemma 2 *The expression ω in (3-6) satisfies the inequality*

$$\omega > 0 \quad (3-11)$$

provided $0 < \alpha_j + \beta_j x_t < 1$ holds for all j and t .

Lemma 3 *The following inequality holds provided $0 < \alpha_j + \beta_j x_t < 1$ is valid for all j and t :*

$$\omega - h' Z^{-1} h > 0 \quad (3-12)$$

We know that Z is the sum of the positive definite matrix V and the positive semidefinite matrix $(W_0 \otimes \iota')$. Therefore Z^{-1} is positive definite. From lemma 2 we see that the term $(1/(\omega - h'Z^{-1}h)hh')$ is positive semidefinite. Therefore B^{-1} in (3-10) is positive definite. Then B and the matrix in (3-8) are positive definite, too, which together with lemma 2 implies (3-7). This completes the proof of

Theorem 2 *Assume that there are at least two linearly independent vectors among the vectors x_1, x_2, \dots, x_n and that $0 < \alpha_i + \beta_i x_i < 1$ holds for $i = 1, \dots, k$ and $t = 1, \dots, k$. Then the information matrix (3-6) is positive definite.*

4 Comments

4.1 The Applicability of the Dirichlet Distribution

Woodland (1979 p. 363) starts his study by considering a system of *nonlinear* deterministic share equations as given by (2-1) and points out: "This formulation of the deterministic shares is particularly convenient in the motivation for the Dirichlet distribution introduced below. Moreover it is a very general formulation." However, both in his simulation study and in his three empirical examples he uses *linear* share equations from type (2-2)¹⁰. Our analysis of the likelihood function for both models reveals that indeed the nonlinear case fits better into the stochastic specification involving the Dirichlet distribution.

The fact, that the stochastic specification of error terms in share equations is more than merely "adding an additive error term", is nicely illustrated for the nonlinear specification: For the deterministic share equations (2-1) parameters α_i and β_i are *not* identified, that is the parameters are only unique up to a scale factor. However our analysis in section 3.1 shows that in the nonlinear Dirichlet

¹⁰This can be motivated by the underlying economic theory which assumes that the denominator in (2-2) equals 1.

share equations model as described in section 2.2 these parameters *are* identified. This is due to the fact that not only the expected values, but also the covariance matrix of the shares for this model are determined by these parameters.

It is intuitively clear that a correct stochastic specification of share equations models must take account of heteroskedasticity¹¹. Moreover the (contemporaneous) correlation between shares should be recognized. As to my opinion the important message from Woodland's (1979) study is that an estimator which takes care of both phenomena will be satisfactory: Firstly, for the linear two-equation model (which due to adding-up reduces to one equation and therefore does not need to consider correlation) it is shown that GLS estimation is only marginally less efficient than ML estimation¹². Estimation of a linear three-equation model is then illustrated by means of three examples with relatively large sample size¹³. In the first two examples the by now common procedure of iterated 3SLS produces almost identical estimates in comparison with ML estimation of the Dirichlet share specification. The third example uses estimation of a truncated normal distribution as the alternative procedure. Here the estimates differ quite a lot from ML estimates of the Dirichlet specification but standard errors for both procedures are large. This last result might be seen as an indication of misspecification and it would be interesting to compare these results with those from a *nonlinear* specification of the "flexible functional form" type as given by (2-1).

A main disadvantage of the Dirichlet distribution is that all correlations between shares are negative. More seriously, these nonzero correlations do not necessarily have any economic meaning: If expenditures with regard to k different goods were *independently* gamma distributed with identical scale parameter and second parameter θ_i , then the budget shares are Dirichlet distributed with parameters θ_i . See Johnson and Kotz (1972 p. 231-233). This reveals two inconsistencies of the

¹¹This is best demonstrated by a graphical representation of (2-1) or (2-2).

¹²See Woodland (1979), in particular table 1.

¹³Example 1: 41 observations, example 2: 61 observations, example 3: 316 observations.

Dirichlet distribution approach: (i) Expenditures for some goods may very well be correlated with each other according to substitutional or complementary relations of the corresponding goods. (ii) The specification of gamma distributed expenditures does not take into account the budget constraint as a (finite) upper bound of the random variable. Needless to say that these points should not be taken as arguments in favour of the purely heuristic specification of *normally* distributed shares.

4.2 Competing Approaches

Econometric analysis of share equations such as (2-1) or (2-2) still has to consider another problem which is almost never discussed: There is no guarantee that the estimated shares lie within the interval $[0, 1]$. Estimates of α_i and β_i which imply estimated shares outside the unit interval may occur even under the Dirichlet specification. Woodland (1979, section 6) demonstrates that this event has very small probability if the economic model is correctly specified. An attractive alternative which does not have this defect is proposed by Considine and Mount (1984)¹⁴. They start from the *stochastic* share equations:

$$Y_{it} = \frac{\exp(\alpha_i + \beta_i x_t + \epsilon_{it})}{\sum_{j=1}^k \exp(\alpha_j + \beta_j x_t + \epsilon_{jt})} \quad (4-1)$$

where the ϵ 's are normally distributed random variables. Clearly, for this model estimated shares will always lie within the unit interval¹⁵. Moreover the ϵ 's could follow any distribution. From an econometric point of view a disadvantage is that demand-theoretic restrictions vary over the sample points. This could however be as well seen as an advantage in checking whether the restrictions (such as concavity) are satisfied for all points¹⁶.

¹⁴I am grateful to Timothy J. Considine for providing me with some subsequent work not yet published.

¹⁵Note that Aitchison (1982) has already proposed such a share model for the analysis of budget data. However he did not discuss its economic aspects.

¹⁶Considine raises this argument in an unpublished paper.

The various difficulties arising in the econometric analysis of share equations also provoke discussion whether econometricians should stick to share equations at all. From a statistical point of view it would be much easier to model the *quantities* demanded. However, such an approach does not necessarily lead to a simpler solution if economic considerations are recognized. For example, Chipman and Tian (1987) derive a generalized form of the “linear expenditure system” from a *stochastic* utility function. Since utility for this approach is assumed to be positive, the random term has to follow a distribution which is bounded from below. Chipman and Tian use the three-parameter lognormal distribution. The specification of a stochastic utility function can also be used in connection with share equations: Lee and Pitt (1986) illustrate this for the case of the translog demand system. In their model random components may follow any distribution. Therefore it is similar in its structure to the logistic approach of Considine and Mount (1984) although the primary concern of this model is the phenomenon of “zero consumption” for certain goods.

4.3 Concluding Remarks

Econometric analysis of share equations needs special treatment of stochastic specifications which are frequently not taken into account in an appropriate manner. The Dirichlet distribution offers a formally acceptable model for shares and its numerical treatment seems to be no longer of major concern¹⁷. This paper shows that the nonlinear specification (2-1) is “more natural” than the linear approach. However the rather restrictive covariance structure limits its applicability in both cases. It seems that so far a satisfactory specification has not yet been found.

¹⁷See the discussion in Ronning (1989).

A Proof of (3-2) and Alternative Proof of Theorem 1

A.1 Proof of (3-2)

We consider first and second derivatives of the loglikelihood function (3-1) with respect to the parameters α_i and β_i , $i = 1, \dots, k$. For the first-order partial derivatives we obtain:

$$\begin{aligned}\frac{\partial L}{\partial \alpha_i} &= \sum_t (\psi(S_t) - \psi(\theta_{it}) + \log(y_{it})) \\ \frac{\partial L}{\partial \beta_i} &= \sum_t (\psi(S_t) - \psi(\theta_{it}) + \log(y_{it})) x_t\end{aligned}\tag{A-1}$$

where ψ denotes the digamma function¹⁸. The second-order partial derivatives are given by

$$\begin{aligned}\frac{\partial^2 L}{(\partial \alpha_i)^2} &= \sum_t (\psi_1(S_t) - \psi_1(\theta_{it})) \\ \frac{\partial^2 L}{\partial \alpha_i \partial \alpha_j} &= \sum_t \psi_1(S_t) \quad , \quad i \neq j \\ \frac{\partial^2 L}{(\partial \beta_i)^2} &= \sum_t (\psi_1(S_t) - \psi_1(\theta_{it})) x_t^2 \\ \frac{\partial^2 L}{\partial \beta_i \partial \beta_j} &= \sum_t \psi_1(S_t) x_t^2 \quad , \quad i \neq j \\ \frac{\partial^2 L}{\partial \alpha_i \partial \beta_i} &= \sum_t (\psi_1(S_t) - \psi_1(\theta_{it})) x_t \\ \frac{\partial^2 L}{\partial \alpha_i \partial \beta_j} &= \sum_t \psi_1(S_t) x_t \quad , \quad i \neq j\end{aligned}\tag{A-2}$$

¹⁸The digamma function $\psi(x)$ is the first derivative of $\log(\Gamma(x))$.

We now define the following expressions:

$$p_{it} = \psi_1(\alpha_i + \beta_i x_t)$$

$$q_t = \psi_1(S_t)$$

$$\mathbf{x}_t = \begin{pmatrix} 1 \\ x_t \end{pmatrix}$$

$$X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$$

$$\boldsymbol{\iota} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \quad (k-1 \text{ elements})$$

$$P_i = \begin{pmatrix} p_{i1} & & & \\ & p_{i2} & & \\ & & \ddots & \\ & & & p_{in} \end{pmatrix}$$

$$Q_0 = \begin{pmatrix} q_1 & & & \\ & q_2 & & \\ & & \ddots & \\ & & & q_n \end{pmatrix}$$

$$P = \begin{pmatrix} P_1 & & & \\ & P_2 & & \\ & & \ddots & \\ & & & P_k \end{pmatrix}$$

$$Q = \begin{pmatrix} Q_0 & Q_0 & \dots & Q_0 \\ Q_0 & Q_0 & \dots & Q_0 \\ \vdots & \vdots & \ddots & \vdots \\ Q_0 & Q_0 & Q_0 & Q_0 \end{pmatrix} = Q_0 \otimes \iota'$$

$$D_t = \begin{pmatrix} p_{1t} & & & \\ & p_{2t} & & \\ & & \ddots & \\ & & & p_{kt} \end{pmatrix}$$

Then the matrix H of second-order partial derivatives (Hessian matrix) can be written as follows:

$$H = - \sum_{t=1}^n B_t \otimes x x' \quad (\text{A} - 3)$$

where the $(k \times k)$ matrix B_t has the following form:

$$B_t = D_t - q_t \iota \iota' \quad (\text{A} - 4)$$

The proof of theorem 1 in the text is based on these two formulae.

A.2 Alternative Proof of Theorem 1

An alternative formulation of the Hessian matrix H is given by

$$H = - (X \otimes I_k)' (P - Q) (X \otimes I_k) \quad (\text{A} - 5)$$

If we assume that the matrix X has full column rank¹⁹, positive definiteness of $P - Q$ is sufficient for the negative definiteness of H . We will consider $(P - Q)^{-1}$ and show that this matrix is positive definite implying that $P - Q$ has the same property.

We define the $(nk \times k)$ matrix F by $F = W_0^{1/2} \otimes \iota$. Then we can write

$$P - Q = P - FF' \quad (\text{A} - 6)$$

and by a well-known matrix inversion lemma (see, for example, Rao 1973, p. 33.)

$$(P - Q)^{-1} = P^{-1} + P^{-1}F \left[I - F'P^{-1}F \right]^{-1} F'P^{-1} \quad (\text{A} - 7)$$

Since $I - F'P^{-1}F = I - Q_0 \sum_{j=1}^k P_j^{-1}$ is diagonal, it would be positive definite if all diagonal elements were positive. For the t -th diagonal element we have (suppressing the index t)

$$1 - q \sum_{j=1}^k \frac{1}{p_j} = \frac{\prod_{j=1}^k p_j - q \prod_{j \neq 1} p_j - q \prod_{j \neq 2} p_j \cdots - q \prod_{j \neq k} p_j}{\prod_{j=1}^k p_j} \quad (\text{A} - 8)$$

We now use inequality (3-4) of section 3.1 which in the notation of this appendix can be written as follows:

$$q < \frac{\theta_i}{S} p_i \quad \text{for each } i \quad (\text{A} - 9)$$

Applying this inequality to the right hand side numerator of (A-8) and noticing that $S = \sum_j \theta_j$ we find that all diagonal elements of the matrix $I - F'P^{-1}F$ are positive which shows that $(P - Q)^{-1}$ is positive definite. It then follows that H in (A-5) is negative definite. \square

¹⁹We consider the case of only one explanatory variable in the text, that is our matrix X has only two columns. However, the proof applies to an arbitrary number of columns.

B Derivation of the Hessian Matrix for the Linear Case

We consider first and second derivatives of the loglikelihood function (3-6) with respect to the parameters α_i , β_i , $i = 1, \dots, k-1$, and c . For the first-order partial derivatives we obtain²⁰:

$$\begin{aligned}\frac{\partial L}{\partial \alpha_i} &= c \sum_t \left(\log \frac{y_{it}}{y_{kt}} - \psi(c(\alpha_i + \beta_i x_t)) + \psi(c(1 - \sum_{j < k} \alpha_j - x_t \sum_{j < k} \beta_j)) \right) \\ \frac{\partial L}{\partial \beta_i} &= c \sum_t \left(\log \frac{y_{it}}{y_{kt}} - \psi(c(\alpha_i + \beta_i x_t)) + \psi(c(1 - \sum_{j < k} \alpha_j - x_t \sum_{j < k} \beta_j)) \right) x_t \\ \frac{\partial L}{\partial c} &= n\psi(c) + \sum_t \sum_{j < k} \left((\alpha_j + \beta_j x_t) \log \frac{y_{jt}}{y_{kt}} - (\alpha_j + \beta_j x_t) \psi(c(\alpha_j + \beta_j x_t)) \right) \\ &\quad + \sum_t \left(\log(y_{kt}) - (1 - \sum_{j < k} \alpha_j - x_t \sum_{j < k} \beta_j) \psi(c(1 - \sum_{j < k} \alpha_j - x_t \sum_{j < k} \beta_j)) \right)\end{aligned}$$

The second-order partial derivatives with respect to α_i and β_i , $i = 1, \dots, k-1$, are given by

$$\begin{aligned}\frac{\partial^2 L}{(\partial \alpha_i)^2} &= -c^2 \sum_t \left(\psi_1(c(\alpha_i + \beta_i x_t)) + \psi_1(c(1 - \sum_{\ell < k} \alpha_\ell - x_t \sum_{\ell < k} \beta_\ell)) \right) \\ \frac{\partial^2 L}{\partial \alpha_i \partial \alpha_j} &= -c^2 \sum_t \psi_1(c(1 - \sum_{\ell < k} \alpha_\ell - x_t \sum_{\ell < k} \beta_\ell)) \quad , \quad i \neq j \\ \frac{\partial^2 L}{(\partial \beta_i)^2} &= -c^2 \sum_t \left(\psi_1(c(\alpha_i + \beta_i x_t)) + \psi_1(c(1 - \sum_{\ell < k} \alpha_\ell - x_t \sum_{\ell < k} \beta_\ell)) \right) x_t^2 \\ \frac{\partial^2 L}{\partial \beta_i \partial \beta_j} &= -c^2 \sum_t \psi_1(c(1 - \sum_{\ell < k} \alpha_\ell - x_t \sum_{\ell < k} \beta_\ell)) x_t^2 \quad , \quad i \neq j \\ \frac{\partial^2 L}{\partial \alpha_i \partial \beta_i} &= -c^2 \sum_t \left(\psi_1(c(\alpha_i + \beta_i x_t)) + \psi_1(c(1 - \sum_{\ell < k} \alpha_\ell - x_t \sum_{\ell < k} \beta_\ell)) \right) x_t \\ \frac{\partial^2 L}{\partial \alpha_i \partial \beta_j} &= -c^2 \sum_t \psi_1(c(1 - \sum_{\ell < k} \alpha_\ell - x_t \sum_{\ell < k} \beta_\ell)) x_t \quad , \quad i \neq j\end{aligned}$$

²⁰ $\psi(x)$ denotes the digamma function. See Appendix A.

Note that the three cross derivatives labelled $i \neq j$ do not depend on i or j . Moreover we have

$$\begin{aligned} \frac{\partial^2 L}{\partial \alpha_i \partial c} &= \sum_t \left(\log \frac{y_{it}}{y_{kt}} - \psi(c(\alpha_i + \beta_i x_t)) + \psi(c(1 - \sum_{j < k} \alpha_j - x_t \sum_{j < k} \beta_j)) \right) \\ &\quad - c \sum_t (\alpha_i + \beta_i x_t) \psi_1(c(\alpha_i + \beta_i x_t)) \\ &\quad + c \sum_t \left((1 - \sum_{j < k} \alpha_j - x_t \sum_{j < k} \beta_j) \psi_1(c(1 - \sum_{j < k} \alpha_j - x_t \sum_{j < k} \beta_j)) \right) \end{aligned} \quad (\text{B-1})$$

$$\begin{aligned} \frac{\partial^2 L}{\partial \beta_i \partial c} &= \sum_t \left(\log \frac{y_{it}}{y_{kt}} - \psi(c(\alpha_i + \beta_i x_t)) - \psi(c(1 - \sum_{j < k} \alpha_j - x_t \sum_{j < k} \beta_j)) \right) \cdot x_t \\ &\quad - c \sum_t (\alpha_i + \beta_i x_t) \psi_1(c(\alpha_i + \beta_i x_t)) x_t \\ &\quad + \sum_t \left((1 - \sum_{j < k} \alpha_j - x_t \sum_{j < k} \beta_j) \psi_1(c(1 - \sum_{j < k} \alpha_j - x_t \sum_{j < k} \beta_j)) \right) x_t \end{aligned} \quad (\text{B-2})$$

These two expressions have the following two special characteristics in common: the first term depends on the outcome of the random variables y_{it} and is proportional to the gradients $\frac{\partial L}{\partial \alpha_i}$ and $\frac{\partial L}{\partial \beta_i}$, respectively. Since it can be verified²¹ that

$$E \left(\log \frac{y_{it}}{y_{kt}} \right) = \psi(\theta_{jt}) - \psi(\theta_{kt})$$

holds, the three first terms vanish if the expected value of the two cross derivatives in (B-2) and (B-3) are considered. Finally we have

$$\begin{aligned} \frac{\partial^2 L}{(\partial c)^2} &= n \psi_1(c) - \sum_t \sum_{j < k} \left((\alpha_j + \beta_j x_t)^2 \psi_1(c(\alpha_j + \beta_j x_t)) \right) \\ &\quad - \sum_t \left((1 - \sum_{j < k} \alpha_j - x_t \sum_{j < k} \beta_j)^2 \psi_1(c(1 - \sum_{j < k} \alpha_j - x_t \sum_{j < k} \beta_j)) \right) \end{aligned}$$

Since two elements of the Hessian matrix depend on the y_{it} , the loglikelihood function cannot be globally concave. We therefore consider the information matrix, that is the negative of the expected value of the Hessian matrix. We want to show that

²¹For the special case of $k = 2$ see Johuson and Kotz (1970 p. 52)

this matrix is positive definite. A convenient formulation of this matrix makes use of the following definitions:

$$\begin{aligned}
A &= \sum_{j < k} \alpha_j \\
B &= \sum_{j < k} \beta_j \\
u_{ti} &= \alpha_i + \beta_i x_t & i = 1, \dots, k-1 \\
v_{ti} &= \psi_1(c(\alpha_i + \beta_i x_t)) \\
d_t &= 1 - \sum_{j < k} u_{tj} \\
w_t &= \psi_1(c(1 - A - Bx_t)) = \psi_1(cd_t)
\end{aligned}$$

$$\begin{matrix} e \\ (n \times 1) \end{matrix} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \quad \begin{matrix} u_i \\ (n \times 1) \end{matrix} = \begin{pmatrix} u_{1i} \\ u_{2i} \\ \vdots \\ u_{ni} \end{pmatrix}, \quad \begin{matrix} x \\ (n \times 1) \end{matrix} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

$$\begin{matrix} u \\ ((k-1)n \times 1) \end{matrix} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_{k-1} \end{pmatrix} \quad \begin{matrix} \iota \\ ((k-1) \times 1) \end{matrix} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

$$d = e - \sum_{j < k} u_j, \quad \epsilon = e \otimes \iota$$

$$X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$$

$$\begin{aligned}
V_i &= \begin{pmatrix} v_{1i} & & & \\ & v_{2i} & & \\ & & \ddots & \\ & & & v_{ni} \end{pmatrix} & V &= \begin{pmatrix} V_1 & & & \\ & V_2 & & \\ & & \ddots & \\ & & & V_n \end{pmatrix} \\
W_0 &= \begin{pmatrix} w_1 & & & \\ & w_2 & & \\ & & \ddots & \\ & & & w_n \end{pmatrix} & , W &= W_0 \otimes \mathcal{U}'
\end{aligned}$$

We denote the information matrix by Ω , that is $\Omega = -E(H)$ where H is the Hessian matrix. We shall illustrate the structure of the information matrix for the case $k = 3$, that is we have parameters $\alpha_1, \beta_1, \alpha_2, \beta_2, c$ which have to be estimated.

	α_1	β_1	α_2	β_2	c
α_1	$c^2 \sum_t v_{t1}$ $+ c^2 \sum_t w_t$	$c^2 \sum_t x_t v_{t1}$ $+ c^2 \sum_t x_t w_t$	$c^2 \sum_t w_t$	$c^2 \sum_t x_t w_t$	$c \sum_t u_{t1} v_{t1}$ $- c \sum_t d_t w_t$
β_1	$c^2 \sum_t x_t v_{t1}$ $+ c^2 \sum_t x_t w_t$	$c^2 \sum_t x_t^2 v_{t1}$ $+ c^2 \sum_t x_t^2 w_t$	$c^2 \sum_t x_t w_t$	$\sum_t x_t^2 w_t$	$c \sum_t x_t u_{t1} v_{t1}$ $- c \sum_t x_t d_t w_t$
α_2	$c^2 \sum_t w_t$	$c^2 \sum_t x_t w_t$	$c^2 \sum_t v_{t2}$ $+ c^2 \sum_t w_t$	$c^2 \sum_t x_t v_{t2}$ $+ c^2 \sum_t x_t w_t$	$c \sum_t u_{t2} v_{t2}$ $- c \sum_t d_t w_t$
β_2	$c^2 \sum_t x_t w_t$	$c^2 \sum_t x_t^2 w_t$	$c^2 \sum_t x_t v_{t2}$ $+ c^2 \sum_t x_t w_t$	$c^2 \sum_t x_t^2 v_{t2}$ $+ c^2 \sum_t x_t^2 w_t$	$c \sum_t x_t u_{t2} v_{t2}$ $- c \sum_t x_t d_t w_t$
c	$c \sum_t u_{t1} v_{t1}$ $- c \sum_t d_t w_t$	$c \sum_t x_t u_{t1} v_{t1}$ $- c \sum_t x_t d_t w_t$	$c \sum_t u_{t2} v_{t2}$ $- c \sum_t d_t w_t$	$c \sum_t x_t u_{t2} v_{t2}$ $- c \sum_t x_t d_t w_t$	$-n\psi_1(c)$ $+ \sum_t \sum_{j < k} u_{tj}^2 v_{tj}$ $+ \sum_t d_t^2 w_t$

This matrix can be written in a more compact form as follows:

	(α_1, β_1)	(α_2, β_2)	c
$\begin{pmatrix} \alpha_1 \\ \beta_1 \end{pmatrix}$	$c^2 X'(V_1 + W_0)X$	$c^2 X'W_0X$	$cX'V_1u_1$ $-cX'Wd$
$\begin{pmatrix} \alpha_2 \\ \beta_2 \end{pmatrix}$	$c^2 X'W_0X$	$c^2 X'(V_2 + W_0)X$	$cX'V_2u_2$ $-cX'W_0d$
c	cu'_1V_1X $-cd'W_0X$	cu'_2V_2X $-cd'W_0X$	$-n\psi_1(c)$ $+\sum_{j < k} u'_j V_j u_j + d'W_0d$

We obtain the following form of the information matrix Ω with respect to the parameters $\alpha_1, \beta_1, \dots, \alpha_{k-1}, \beta_{k-1}, c$:

$$\Omega = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}$$

where Ω_{11} has $2(k-1)$ rows and columns, respectively, and Ω_{22} is a scalar which we will henceforth denote by ω . We then have for the different submatrices of Ω :

$$\Omega_{11} = c^2 ((X \otimes I_{k-1})'(V + (W_0 \otimes \iota'))(X \otimes I_{k-1}))$$

where I_r denotes the $(r \times r)$ identity matrix. Since V and W_0 are positive diagonal matrices, the submatrix Ω_{11} is positive definite if X has full column rank. Furthermore we have

$$\Omega_{12} = c(X \otimes I_{k-1})'(Vu - (W_0 \otimes I_{k-1})(d \otimes \iota)) = \Omega'_{21}$$

Finally the scalar term $\Omega_{22} = \omega$ is given by

$$\omega = -(n\psi_1(c) - u'Vu - d'W_0d)$$

C Proof of Lemma 2

We consider the expression

$$\begin{aligned}
 -\omega &= n\psi_1(c) - u'Vu - d'W_0d \\
 &= \sum_{t=1}^n \left[\psi_1(c) - \sum_{j=1}^{k-1} u_{tj}^2 v_{tj} - d_t^2 w_t \right] \\
 &= \sum_{t=1}^n \left[\sum_{j=1}^{k-1} u_{tj} \psi_1(c) + d_t \psi_1(c) - \sum_{j=1}^{k-1} u_{tj}^2 v_{tj} - d_t^2 w_t \right] \\
 &= \sum_{t=1}^n \sum_{j=1}^{k-1} u_{tj} (\psi_1(c) - u_{tj} v_{tj}) + \sum_{t=1}^n d_t (\psi_1(c) - d_t w_t)
 \end{aligned}$$

Please note that u_{tj} and d_t are shares, that is, $0 < u_{tj} < 1$ and $0 < d_t < 1$ holds for each j and t . Since $v_{tj} = \psi_1(cu_{tj})$ and $w_t = \psi_1(cd_t)$ (see appendix B), we obtain from the basic inequality (1-1) of section 1.

$$\psi_1(c) < u_{tj} v_{tj} \quad \text{for all } t \text{ and } j$$

and

$$\psi_1(c) < d_t w_t \quad \text{for all } t$$

This implies that $\omega > 0$. \square

D Proof of Lemma 3

We consider the expression

$$\begin{aligned}\omega - h'Z^{-1}h &= -n\psi_1(c) + u'Vu + d'W_0d \\ &\quad - [Vu - (W_0 \otimes I)(d \otimes \iota)]'(V + W)^{-1}[Vu - (W_0 \otimes I)(d \otimes \iota)]\end{aligned}\tag{D-1}$$

Since

$$d = e - (I \otimes \iota)'u$$

and

$$\begin{aligned}(d \otimes \iota) &= (e \otimes \iota) - ((I \otimes \iota)'u \otimes \iota) \\ &= \epsilon - (I \otimes \iota)'u\end{aligned}$$

(see Appendix B) we can write for (D-1):

$$\begin{aligned}\omega - h'Z^{-1}h &= -n\psi_1(c) + u'Vu + e'W_0e \\ &\quad + ((I \otimes \iota)'u)'W_0((I \otimes \iota)'u) - 2((I \otimes \iota)'u)'W_0e \\ &\quad - u'V(V + W)^{-1}Vu - \epsilon'(W_0 \otimes I)(V + W)^{-1}(W_0 \otimes I)\epsilon \\ &\quad - u'(I \otimes \iota)'(W_0 \otimes I)(V + W)^{-1}(W_0 \otimes I)(I \otimes \iota)'u \\ &\quad + 2\epsilon'(W_0 \otimes I)(V + W)^{-1}(W_0 \otimes I)(I \otimes \iota)'u \\ &\quad + 2u'V(V + W)^{-1}(W_0 \otimes I)\epsilon \\ &\quad - 2u'V(V + W)^{-1}(W_0 \otimes I)(I \otimes \iota)'u \\ &= u'Vu + u'(I \otimes \iota)'W_0(I \otimes \iota)'u - u'V(V + W)^{-1}Vu \\ &\quad - u'W(V + W)^{-1}Wu - 2u'V(V + W)^{-1}Wu \\ &\quad - n\psi_1(c) + e'W_0e - \epsilon'(W_0 \otimes I)(V + W)^{-1}(W_0 \otimes I)\epsilon \\ &\quad - 2u'(I \otimes \iota)'W_0e + 2u'W(V + W)^{-1}(W_0 \otimes I)\epsilon \\ &\quad + 2u'V(V + W)^{-1}(W_0 \otimes I)\epsilon\end{aligned}\tag{D-2}$$

Since

$$\begin{aligned}u'Vu &= u'V(V + W)^{-1}(V + W)Vu \\ &= u'V((V + W)^{-1}Vu + u'V(V + W)^{-1}Wu \\ u'Wu &= u'W(V + W)^{-1}Wu + u'W(V + W)^{-1}Vu\end{aligned}$$

and

$$\begin{aligned}(I \otimes \iota)W_0(I \otimes \iota)' &= (W_0 \otimes \iota \iota') = W \\ (I \otimes \iota)W_0 e &= W_0 e \otimes \iota = (W_0 \otimes I)\epsilon\end{aligned}$$

we obtain from (D-2)

$$\begin{aligned}\omega - h'Z^{-1}h &= u'V(V+W)^{-1}Wu + u'V(V+W)^{-1}Wu \\ &\quad - 2u'V(V+W)^{-1}Wu \\ &\quad - n\psi(c) + e'W_0e - \epsilon'(W_0 \otimes I)(V+W)^{-1}(W_0 \otimes I)\epsilon \\ &\quad - 2u'(I \otimes \iota)W_0e + 2u'(I \otimes \iota)W_0\epsilon \\ &= -n\psi_1(c) + e'W_0e - \epsilon'(W_0 \otimes I)(V+W)^{-1}(W_0 \otimes I)\epsilon\end{aligned}\tag{D-3}$$

Since

$$\begin{aligned}&\epsilon'(W_0 \otimes I)(V+W)^{-1}(W_0 \otimes I)\epsilon - \epsilon'W_0e \\ &= \frac{1}{(k-1)^2} \left(\epsilon'W(V+W)^{-1}W\epsilon - \epsilon'W\epsilon \right) \\ &= \frac{1}{(k-1)^2} \left(\epsilon'W(V+W)^{-1}W\epsilon - \epsilon'(V+W)(V+W)^{-1}W\epsilon \right) \\ &= -\frac{1}{(k-1)^2} \epsilon'V(V+W)^{-1}W\epsilon \\ &= -\frac{1}{k-1} \epsilon'V(V+W)^{-1}(W_0 \otimes I)\epsilon\end{aligned}$$

and

$$n\psi_1(c) = e'\Psi e$$

with $\Psi = \psi_1(c)I_n$ we finally obtain from (D-3)

$$\begin{aligned}\omega - h'Z^{-1}h &= -e'\Psi e + \frac{1}{k-1} \epsilon'V(V+W)^{-1}(W_0 \otimes I)\epsilon \\ &= -e'\Psi e + \frac{1}{(k-1)^2} \epsilon'V(V+W)^{-1}W\epsilon\end{aligned}\tag{D-4}$$

Application of the matrix inversion lemma (see, e.g., Rao 1973 p. 33) gives

$$\begin{aligned}(V+W)^{-1} &= \left[V + (W_0^{1/2} \otimes \iota)(W_0^{1/2} \otimes \iota)' \right]^{-1} \\ &= V^{-1} - V^{-1}(W_0^{1/2} \otimes \iota) \left[I + (W_0^{1/2} \otimes \iota)'V^{-1}(W_0^{1/2} \otimes \iota) \right]^{-1} (W_0^{1/2} \otimes \iota)'V^{-1} \\ &= V^{-1} - V^{-1}(W_0^{1/2} \otimes \iota) \left[I + W_0 \sum_{j=1}^{k-1} V_j^{-1} \right]^{-1} (W_0^{1/2} \otimes \iota)'V^{-1} \\ &= V^{-1} - V^{-1}(W_0^{1/2} \otimes \iota)C^{-1}(W_0^{1/2} \otimes \iota)'V^{-1}\end{aligned}$$

with

$$C = I + W_0 \sum_{j=1}^{k-1} V_j^{-1}$$

Note that

$$\begin{aligned} (W_0^{1/2} \otimes \iota)' V^{-1} (W_0^{1/2} \otimes \iota) &= W_0 \sum_{j=1}^{k-1} V_j^{-1} = C - I \\ (W_0^{1/2} \otimes \iota) C^{-1} (W_0^{1/2} \otimes \iota)' &= (W_0 C^{-1}) \otimes \iota \iota' \\ W_0 C^{-1} &= \left[W_0^{-1} + \sum_{j=1}^{k-1} V_j^{-1} \right]^{-1} \end{aligned}$$

where we have exploited the diagonal structure of W_0 and $V_i, i = 1, \dots, k-1$.

Therefore we can write

$$\begin{aligned} V(V+W)^{-1}W &= W - (W_0^{1/2} \otimes \iota) C^{-1} (W_0^{1/2} \otimes \iota)' V^{-1} (W_0^{1/2} \otimes \iota) (W_0^{1/2} \otimes \iota)' \\ &= W - (W_0^{1/2} \otimes \iota) C^{-1} W_0 \left(\sum_{j=1}^{k-1} V_j^{-1} \right) (W_0^{1/2} \otimes \iota)' \\ &= (W_0^{1/2} \otimes \iota) \left[I - C^{-1} (C - I) \right] (W_0^{1/2} \otimes \iota)' \\ &= (W_0^{1/2} \otimes \iota) C^{-1} (W_0^{1/2} \otimes \iota)' \\ &= (W_0 C^{-1}) \otimes \iota \iota' \end{aligned}$$

Therefore our expression under consideration reduces further from (D-4) to

$$\begin{aligned} \omega - h' Z^{-1} h &= -e' \Psi e + \frac{1}{(k-1)^2} e' \left((W_0 C^{-1}) \otimes \iota \iota' \right) e \quad (\text{D-5}) \\ &= -e' \Psi e + \frac{1}{k-1} e' \left((W_0 C^{-1}) \otimes I \right) e \\ &= -e' \left(\Psi - \left[W_0^{-1} + \sum_{j=1}^{k-1} V_j^{-1} \right]^{-1} \right) e \end{aligned}$$

We see from equation (D-5) that our expression would be positive if all diagonal elements of the matrix

$$\Psi - \left[W_0^{-1} + \sum_{j=1}^{k-1} V_j^{-1} \right]^{-1}$$

would be negative. But this can be shown as follows:²² If we write ψ instead of $\psi_1(c)$ and denote the t -th diagonal element of W_0 by w and the t -th diagonal element of

²²A similar procedure was used in section A.2 of appendix A for the proof of theorem 1.

V_j by v_j , then we have for each diagonal element of the matrix

$$\Psi = \left[W_0^{-1} + \sum_{j=1}^{k-1} V_j^{-1} \right]^{-1}:$$

$$\begin{aligned} \psi &= \left(\frac{1}{w} + \frac{1}{v_1} + \frac{1}{v_2} + \dots + \frac{1}{v_{k-1}} \right)^{-1} \\ &= \psi - \left(\frac{\prod_{j=1}^{k-1} v_j + w \prod_{j \neq 1} v_j + w \prod_{j \neq 2} v_j + \dots + w \prod_{j \neq k-1} v_j}{w \prod_{j=1}^{k-1} v_j} \right)^{-1} \\ &= \psi - \frac{w \prod_j v_j}{\prod_{j=1}^{k-1} v_j + w \prod_{j \neq 1} v_j + w \prod_{j \neq 2} v_j + \dots + w \prod_{j \neq k-1} v_j} \\ &= \frac{\psi \prod_j v_j + \psi w \prod_{j \neq 1} v_j + \psi w \prod_{j \neq 2} v_j + \dots + \psi w \prod_{j \neq k-1} v_j - w \prod_j v_j}{\prod_j v_j + w \prod_{j \neq 1} v_j + w \prod_{j \neq 2} v_j + \dots + w \prod_{j \neq k-1} v_j} \\ &< \frac{(d + u_1 + u_2 + \dots + u_{k-1} - 1) w \prod_j v_j}{\prod_j v_j + w \prod_{j \neq 1} v_j + w \prod_{j \neq 2} v_j + \dots + w \prod_{j \neq k-1} v_j} \\ &= 0 \end{aligned}$$

since $u_1 + u_2 + \dots + u_{k-1} + d = 1$ for each t (see appendix B). Again we have applied inequality (1-1) which in the notation of this appendix is given by $\psi_1(c) < d_t w_t$ and $\psi_1(c) < u_{tj} v_{tj}$ for all t and j . This completes the proof that $\omega - h'Z^{-1}h > 0$. \square

References

- Abramowitz, M., and I.A. Stegun 1965. *Handbook of Mathematical Functions* (eighth printing). New York: Dover.
- Aitchison, J. 1982. The Analysis of compositional data. *Journal of the Royal Statistical Society, Ser. B* 44, 139-177 (with discussion).
- Chipman, J.S., and G. Tian 1989. Generalized Maximum Likelihood Estimation of the Linear Expenditure System with Lognormal Distribution. *University of Minnesota*. mimeo.
- Christensen, L.R., D.W. Jorgenson and L.J. Lau 1975. Transcendental logarithmic utility functions. *American Economic Review* 65, 367-383.
- Considine, T.J., and T.D. Mount 1984. The use of linear logit models for dynamic input demand systems. *Review of Economics and Statistics* 66, 434-443.
- Deaton, A., and J. Muellbauer 1980. An almost ideal demand system. *American Economic Review* 70, 312-326.
- Deprins, D., and L. Simar 1985. A note on the asymptotic relative efficiency of m.l.e. in a linear model with gamma disturbances. *Journal of Econometrics* 27, 383-386.
- Dhrymes, P.J. 1978. *Introductory Econometrics*. New York: Springer-Verlag.
- Graybill, F.A. 1983. *Matrices with Applications in Statistics*. Second Edition. Belmont: Wadsworth.
- Johnson, N.L., and S. Kotz 1970. *Continuous Univariate Distributions-2*. Boston: Houghton Mifflin.
- Johnson, N.L., and S. Kotz 1972. *Distributions in Statistics: Continuous Multivariate Distributions*. New York: Wiley.

- Lee, L.F., and Pitt, M.M. 1986. Microeconomic Demand Systems With Binding Nonnegativity Constraints: The Dual Approach. *Econometrica* 54, 1237-1242.
- Rao, C.R. 1973. *Linear statistical inference and its Applications*. New York: Wiley.
- Ronning, G. 1986. On the curvature of the trigamma function. *Journal of Computational and Applied Mathematics* 15, 397-399.
- Ronning, G. 1989. Maximum likelihood estimation of Dirichlet distributions. *Journal of Computational Statistics and Simulation* 32 (to appear).
- Rothenberg, T.J. 1971. Identification in parametric models. *Econometrica* 39, 577-591.
- Woodland, A.D. 1979. Stochastic specification and the estimation of share equations. *Journal of Econometrics* 10, 361-383.