

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS  
AT YALE UNIVERSITY

Box 2125, Yale Station  
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 886

Note: Cowles Foundation Discussion Papers are preliminary materials circulated to stimulate discussion and critical comment. Requests for single copies of a Paper will be filled by the Cowles Foundation within the limits of the supply. References in publications to Discussion Papers (other than mere acknowledgment by a writer that he has access to such unpublished material) should be cleared with the author to protect the tentative character of these papers.

A LITTLE MAGIC WITH THE CAUCHY DISTRIBUTION

Peter C. B. Phillips

October 1988

# A LITTLE MAGIC WITH THE CAUCHY DISTRIBUTION

by

P. C. B. Phillips\*

*Cowles Foundation for Research in Economics  
Yale University*

## 0. ABSTRACT

The standard Cauchy distribution is completely characterized by the property that it has no atoms and is distributionally equivalent under the involution  $X \rightarrow -1/X$  i.e.  $X \equiv -1/X$ . Since maximum likelihood is invariant to the choice of normalization rule in structural equation estimation this property establishes that the LIML estimator is standard Cauchy in the leading case of a canonical structural equation. This is a proof by identifying characteristics and is a major improvement over the usual apparatus of change of variable methods and reductions by multiple integration.

The new approach has applications in many other contexts. A second example considered in the paper is the unidentified ARMA with degenerate common factors. Such models commonly arise from overfitting or overdifferentencing. They have eluded conventional asymptotic methods for many years. Yet they are resolved quite simply by the present approach, which yields both an exact finite sample theory and the relevant asymptotics.

**SOME KEY WORDS:** Canonical form; Cauchy property; Identifying characteristics; Involution; Leading case; Maximum likelihood; Unidentified model.

September 1988

---

\*My thanks go to Glenna Ames for her work in keyboarding the manuscript and to the NSF for research support.

## 1. INTRODUCTION

In teaching the theory of the simultaneous equations model I have for some years found it useful to give the distribution theory of various single equation estimators like instrumental variables (IV) and limited information maximum likelihood (LIML) in the totally unidentified case. This is a polar case that is never mentioned in the textbooks yet it is both simple and fruitful to analyze. For example, the asymptotic theory in this case is readily deduced from the exact distributions, which are themselves quite easy to obtain. Both sets of results dramatically illustrate the effects of nonidentifiability. They show how the distributions of the (usually consistent) structural estimators IV and LIML do not collapse as the sample size  $T \rightarrow \infty$  and continue to reflect in the limit the uncertainty about the coefficients that is implicit in their lack of identification. The theory also introduces students to some fascinating little tricks with exact distributions and gives them an early demonstration of the fact that limit distributions are not always normal or chi-squared.

The derivations in the case of the IV estimator are especially simple and are easily followed by students who are familiar with the  $t$  distribution and elementary conditioning arguments. They were given in earlier work (1982, 1983) and have recently been extended to more general partially identified econometric models in (1989).

One purpose of the present note is to introduce a simple mathematical trick that yields the exact distribution of the LIML estimator directly in a couple of lines. This provides a substantial improvement

over the derivation given in my earlier work on LIML (1984, 1986, 1987). As in the case of the IV estimator, the new derivation should be easily understood by students at an early stage in their training.

A second aim of this note is to show how the same trick has fruitful applications to maximum likelihood estimators (MLE's) in other contexts. One that is especially interesting at present is the under-identified ARMA model. This class of model typically arises from overfitting and/or overdifferencing, leading to degenerate common factors in the limit. Such models have eluded conventional asymptotic analysis for many years. Yet they may be treated quite simply by the method presented here.

## 2. A SIMPLE CHARACTERIZATION OF THE CAUCHY DISTRIBUTION

Using " $\equiv$ " to signify equivalence in distribution we write  $X \equiv C(0,1)$  and say that  $X$  is standard Cauchy if it has density  $\text{pdf}(x) = [\pi(1+x^2)]^{-1}$ . If  $Y$  and  $Z$  are independent  $N(0,1)$  it is well known that  $X \equiv Y/Z \equiv Z/Y$  and hence  $X \equiv 1/X$ . The question of whether there is a converse to this result has intrigued probabilists for many years. It has recently been shown by Knight (1976) (see also Hassenforder (1988) and Phillips (1988) that  $X \equiv C(0,1)$  iff

(C)  $X \equiv -1/X$  and has no atoms.

Condition (C) gives a simple and convenient characterization of the standard Cauchy distribution. A short and elementary proof of C is given in the author's paper (1988). That paper also gives a characterization theorem for the matrix Cauchy distribution  $C_{n,m}(0,I)$  whose

density is

$$\text{pdf}(X) = \left[ \pi^{nm/2} \Gamma_n(n/2) \right]^{-1} \Gamma_n((n+m)/2) |I + XX'|^{-(n+m)/2}$$

It is shown that  $X \equiv C_{n,m}(0, I)$  iff

$$(i) \quad X \equiv HXK \quad \text{for any } H \in O(n), \quad K \in O(m)$$

(D)

$$(ii) \quad x_{ij} \equiv 1/x_{ij} \quad \text{and has no atoms } (i = 1, \dots, n; j = 1, \dots, m)$$

where in D(i)  $O(k)$  signifies the orthogonal group of order  $k$  and in D(ii)  $x_{ij}$  is the  $(i,j)^{\text{th}}$  element of  $X$ .

Both (C) and (D) will be used in our following applications.

### 3. LEADING CASE THEORY FOR LIML

We consider the linear structural equation

$$(1) \quad y_{1t} + y_{2t} + \dots = u_{1t}$$

where  $y_{1t}$  and  $y_{2t}$  are endogenous variables,  $u_{1t}$  is a disturbance and the undefined elements of the equation may involve further endogenous variables and possibly some fixed exogenous variables also. An alternative normalization of (1) is

$$(2) \quad \alpha y_{1t} + y_{2t} + \dots = u'_{1t}, \quad \alpha = 1/\beta.$$

Let  $\hat{\alpha}$ ,  $\hat{\beta}$  be the LIML estimators of  $\alpha$ ,  $\beta$  in (2) and (1), respectively. By the invariance property of maximum likelihood we know that

$$(3) \quad \hat{\alpha} = 1/\hat{\beta}.$$

To develop a leading case theory we assume that the reduced form coefficient matrix is zero, although the theory can be developed under the weaker requirement that a certain submatrix is zero (see Phillips (1982, 1984) for details). We further assume that the model has been transformed to canonical form, so that  $y_{1t}$ ,  $y_{2t}$  (and any other endogenous variables in equation (1)) are independent, serially independent and normally distributed. Writing  $y'_1 = (y_{11}, \dots, y_{1T})$ ,  $y'_2 = (y_{21}, \dots, y_{2T})$ , ... we then have

$$(4) \quad y_1, y_2, \dots \equiv \text{iid } N(0, I_T) .$$

It follows immediately from (1) and (2) that

$$(5) \quad \hat{a} \equiv \hat{\beta} \equiv -\hat{a}$$

the final equivalence holding because  $y_1 \equiv -y_1$ . Combining (3) and (5) we have

$$(6) \quad \hat{a} \equiv -1/\hat{a} .$$

Moreover, the distribution of  $\hat{a}$  has no atoms since the distribution of the data is continuous in view of the stochastic hypothesis (4). Thus, from condition (C) we deduce directly that

$$(7) \quad \hat{a} \equiv C(0,1) .$$

#### REMARKS

- (a) The great advantage of using Condition (C) to establish the Cauchy property (7) is that it shows clearly how this property is the consequence of two simple features of the problem:

- (i) the invariance property of the MLE as expressed in (3) and representing functional invariance to the equation normalization rule;
- (ii) the leading case and canonical form hypotheses that are embodied in (4) which jointly ensure that (5) holds. This leads to distributional invariance to the normalization rule.

Together (i) and (ii) imply distributional invariance under the involution (6) and thereby ensure the Cauchy property in view of condition (C).

- (b) Note that the above arguments do not involve any of the conventional change of variable or integral reduction arguments that are typical of finite sample theory. The method relies on the physical characteristics of the estimation method, the distributional characteristics of the stochastic hypothesis (4) and the characterization property (C) of the Cauchy distribution. It may therefore be described as a proof by identifying characteristics.
- (c) As discussed in earlier work (1984), the distribution (7) is invariant to the sample size  $T$  and is therefore also the asymptotic distribution of the LIML estimator as  $T \rightarrow \infty$ .
- (d) The assumption of normality (4) is not necessary for the Cauchy property (7) to hold. It is sufficient that the multivariate distribution in (4) be continuous and spherically symmetric. The assumption of serial independence can also be relaxed. Writing  $Y = [y_1, y_2, \dots, y_g]$  as the matrix of observations of the (g) endogenous variables in (1) we need only require that  $Y$  be continuous and matrix spherically symmetric i.e.  $Y \equiv LYM$  for any  $L \in O(T)$  and any  $M \in O(g)$ .

- (e) If fixed exogenous variables  $z_{1t}, \dots, z_{kt}$  occur in equation (1) we may also allow these variables to enter the reduced form. Assumption (4) may then be replaced by the weaker requirement

$$Q_{z_1} y_1, Q_{z_1} y_2, \dots \equiv \text{iid } N(0, Q_{z_1})$$

where the matrix  $Q_{z_1}$  is the orthogonal projection onto the orthogonal complement of the space spanned by  $\{z_{1t}, \dots, z_{kt}; t = \dots, T\}$ .

- (f) If  $y_{2t}$  is an  $n$ -vector we can write equation (1) in the form

$$y_{1t} + \beta' y_{2t} + \dots = u_{1t} .$$

Since  $y_{2t} \equiv H y_{2t}$  for any  $H \in O(n)$  it is clear that

$$(7) \quad \hat{\beta} \equiv H \hat{\beta} k \text{ for any } k \in O(1), H \in O(n) .$$

In view of the argument leading to (6) we also obtain the equivalence

$$(8) \quad \hat{\beta}_i \equiv 1/\hat{\beta}_i$$

for individual elements  $\hat{\beta}_i$  of  $\hat{\beta}$ . Combining (7) and (8) we deduce from condition (D) that

$$\hat{\beta} \equiv C_{n,1}(0, I_n) ,$$

giving the corresponding multivariate Cauchy result.

- (g) When the leading case hypothesis (4) fails the Cauchy property (7) no longer holds. However, even in this case the invariance property of maximum likelihood has a sufficiently powerful effect on  $\hat{a}$  to characterize the tails of the distribution as being of the Cauchy type. This was shown in earlier work by Sargan (1988).

- (h) The above results also hold for the full information maximum likelihood (FIML) estimator. The Cauchy property for the structural FIML estimator was shown in the author's paper (1986). However, the canonical form hypothesis is not innocuous in this case since it is not in general possible to reduce the structural system to canonical form without affecting the exclusion restrictions on individual structural equations.

#### 4. UNIDENTIFIED ARMA MODELS

A simple and typical example is the ARMA(1,1) with a degenerate common factor, viz

$$(9) \quad y_t + ay_{t-1} = u_t + \beta u_{t-1}, \quad \text{with } a = \beta$$

and where  $\{u_t\} \equiv \text{iid}(0, \sigma^2)$ . Models such as (9) may arise in practice from overfitting in ARMA model estimation or even overdifferencing prior to estimation when MA components are fitted. It is then not unusual to find estimates of the AR and MA coefficients in the same vicinity i.e.  $\hat{a} \sim \hat{\beta}$  and the estimated model is nearly degenerate. This is a problem that has received virtually no attention in the statistics literature and yet it is not uncommon in applications. Recent examples where the problem has arisen in econometrics occur in Hansen and Singleton (1988) and in Poterba and Summers (1988).

Before we examine the behavior of the MLE's of  $a$  and  $\beta$  in (9) it is instructive to consider what happens to a simpler typical estimator that is consistent when  $a \neq \beta$ . We take the IV estimator of  $a$  that uses  $y_{t-2}$  in place of  $y_{t-1}$  as an instrument. In this case, we

have

$$(10) \quad a^* = \frac{\sum_2^T y_t y_{t-2}}{\sum_2^T y_{t-1} y_{t-2}} = \frac{T^{-1/2} \sum_2^T u_t u_{t-2}}{T^{-1/2} \sum_2^T y_{t-1} y_{t-2}} \Rightarrow \frac{Y}{Z}$$

$$\equiv \frac{N(0,1)}{N(0,1)} \quad \text{with numerator and denominator independent}$$

$$(11) \quad \equiv C(0,1) .$$

Note that in (10) we use the joint weak convergence of the numerator and denominator, viz

$$(T^{-1/2} \sum_2^T u_t u_{t-2}, T^{-1/2} \sum_2^T u_{t-1} u_{t-2}) \Rightarrow (\sigma^4 Y, \sigma^4 Z) .$$

This follows from Billingsley's (1961) central limit theorem for stationary and ergodic martingale differences. The limit variates  $Y$  and  $Z$  are independent since they are normal and since the sample components  $u_t u_{t-2}$  are uncorrelated with  $u_{t-1} u_{t-2}$ .

Unlike (10) there is no explicit formula for the MLE's of  $a$  and  $\beta$  in (9). Indeed, the likelihood itself must be built up from recursive numerical methods. However, this is not a serious obstacle to analysis. We shall instead employ the identifying characteristics approach of the previous section.

First, observe that (9) may be written in the alternatively normalized form

$$(12) \quad \gamma y_t + y_{t-1} = \delta u'_t + u'_{t-2}$$

where

$$\gamma = 1/\alpha, \quad \delta = 1/\beta, \quad \gamma = \delta$$

$$\{u_t\} \equiv \text{iid}(0, \sigma^2\beta^2/\alpha^2) \equiv \text{iid}(0, \sigma^2).$$

Maximum likelihood estimation of (9) and (12) under Gaussian assumptions leads to the pairs  $(\hat{\alpha}, \hat{\beta})$ ,  $(\hat{\gamma}, \hat{\delta})$  with

$$(13) \quad \hat{\gamma} = 1/\hat{\alpha}, \quad \hat{\delta} = 1/\hat{\beta}$$

in view of the invariance properties of the MLE.

We now observe that since  $y_t = u_t \equiv \text{iid } N(0, \sigma^2)$  we may write (12) in the alternative form

$$(14) \quad z_t + \gamma z_{t-1} = \eta_t + \delta \eta_{t-1}$$

where

$$(15) \quad \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_T \end{bmatrix} = \begin{bmatrix} y_T \\ y_{T-2} \\ \vdots \\ y_1 \end{bmatrix} \equiv \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix} \equiv N(0, \sigma^2 I_T)$$

and  $\{\eta_t\} \equiv \text{iid } N(0, \sigma^2)$ . In effect, (14) is simply the time reversed version of the ARMA(1,1) given by (9). However, in view of (15), the statistical properties of the MLE's  $(\hat{\gamma}, \hat{\delta})$  from (14) are identical to those of the corresponding pair  $(\hat{\alpha}, \hat{\beta})$  from (9). We therefore deduce that

$$(16) \quad \hat{\alpha} \equiv \hat{\gamma} \equiv -\hat{\gamma}, \quad \hat{\beta} \equiv \hat{\delta} \equiv -\hat{\delta}$$

the final equivalence holding in each case because of the symmetry of the multivariate distribution (15).

Using (13) and (16) we find

$$\hat{a} \equiv -1/\hat{a} , \quad \hat{\beta} \equiv -1/\hat{\beta}$$

and, thus, by Condition (C)

$$(17) \quad \hat{a} \equiv C(0,1) , \quad \hat{\beta} \equiv C(0,1) .$$

#### REMARKS

- (a) Although  $a$  and  $\beta$  are unidentified in (9) it is clear that  $a-\beta = 0$  and this linear combination of the coefficients is identified. The model is therefore an example of a partially identified model. Such models have been examined in detail in other recent work by the author (1989). In the present case of the partially identified ARMA model (9) it can be shown that  $\hat{a}-\hat{\beta} \xrightarrow{p} 0$  and that  $\sqrt{T}(\hat{a}-\hat{\beta})$  has a nondegenerate limiting distribution. These aspects of the ARMA model (9) are being examined in another paper.
- (b) Note that although the model (9) is unidentified as it stands the use of model selection procedures can help to identify the true mechanism of  $y_t$  as ARMA(0,0). Unfortunately, some of the model selection procedures like AIC tend in any case to lead to overfitting and all of the ones in popular use tend to have poor sampling properties in nearly unidentified cases (Hannan and Rissanen (1982) and Hannan and McDougall (1988) provide some simulation evidence of this). We must therefore expect some erratic performance in such criteria when the ARMA model is degenerate and of a lower order than expected.
- (c) The above derivation of the distribution of the estimates  $(\hat{a}, \hat{\beta})$  assumes that no constraints of stationarity or invertibility are placed on the Gaussian likelihood. This is the usual case. However, when such constraints are placed on the domain of definition

of the coefficients the final result (17) needs to be modified. For example, if the Tunnicliffe-Wilson (1969) flip rule (in which  $\hat{\beta}$  is replaced by  $1/\hat{\beta}$  if  $|\hat{\beta}| > 1$ ) is used, then the modified estimate of  $\beta$  is simply a truncated Cauchy variate on the interval  $(-1,1)$ .

- (d) Note that the result (17) does not depend on the normality assumption. It is sufficient that  $y_t = u_t \equiv \text{iid}(0, \sigma^2)$  and be continuously distributed. This assumption induces the continuous spherical symmetry of  $y' = (y_1, \dots, y_T)$  and that is all that is required for the characterization argument to go through.
- (e) As for the SEM model, the exact theory (17) holds for all  $T$  and therefore delivers the asymptotic theory also.
- (d) The above theory leading to (17) covers direct estimation by maximum likelihood. Some methods of efficient estimation involve recursive algorithms. Durbin's (1960) method and the Hannan-Rissanen (1982) procedure, for example, involve the preliminary fitting of a long autoregression to the data. These methods lead to estimates whose properties can be very different from those of (17). They will be discussed in a later paper.

## REFERENCES

- Billingsley, P. (1961). "The Lindeberg Levy theorem for martingales," *Transactions of the American Mathematical Society*, 83, 250-268.
- Durbin, J. (1960). "The fitting of time series models," *International Statistical Review*, 28, 233-244.
- Hannan, E. J. and M. Deistter (1988). *The Statistical Theory of Linear Systems*. Wiley: New York.
- Hannan, E. J. and A. J. McDougall (1988). "Regression procedures for ARMA estimation," *Journal of the American Statistical Association*, 83, 490-498.
- Hannan, E. J. and R. Rissanen (1982). "Recursive estimation of mixed autoregressive moving average order," *Biometrika*, 69, 81-94.
- Hansen, L. P. and K. Singleton (1988). "Computing semi-parametric efficiency bounds for linear time series models," mimeographed, July 1988.
- Hassenforder, C. (1988). "An extension of Knight's theorem on Cauchy distribution," *Journal of Theoretical Probability*, 1, 205-209.
- Knight, F. B. (1976). "A characterization of the Cauchy type," *Proceedings of the American Mathematics Society*, 55, 130-135.
- Phillips, P. C. B. (1982). "Exact small sample theory in the simultaneous equations model," Ch. 8, pp. 449-516 in M. D. Intriligator and Z. Griliches (eds.), *Handbook of Econometrics*. North Holland: Amsterdam.
- \_\_\_\_\_ (1984). "The exact distribution of LIML: I," *International Economic Review*, 25, 249-261.
- \_\_\_\_\_ (1986). "The exact distribution of FIML in the leading case," *International Economic Review*, 27, 239-243.
- \_\_\_\_\_ (1987). "The distribution of LIML in the leading case," *Econometric Theory*, 3, 469.
- \_\_\_\_\_ (1988). "A new proof of Knight's theorem on the Cauchy distribution," Cowles Foundation, Yale University (mimeographed).
- \_\_\_\_\_ (1989). "Partially identified econometric models," *Econometric Theory* (forthcoming).
- Poterba, J. M. and L. H. Summers (1998). "Mean reversion in stock prices: Evidence and implications," MIT Working Paper No. 457.

Sargan, J. (1988). "The finite sample distribution of FIML estimators,"  
Ch. 3 in J. D. Sargan, *Contributions to Econometrics*, Vol. 2.  
Cambridge University: Cambridge.

Tunncliffe-Wilson, G. (1969). "Factorisation of the covariance  
generating function of a pure moving average process," *SIAM Journal  
on Numerical Analysis*, 6, 1-7.