Econometric Theory, **20**, 2004, 963–987. Printed in the United States of America.
DOI: 10.1017/S0266466604205084

TRANSFORMATIONS FOR MULTIVARIATE STATISTICS

PATRICK MARSH
University of York

This paper derives transformations for multivariate statistics that eliminate asymptotic skewness, extending the results of Niki and Konishi (1986, *Annals of the Institute of Statistical Mathematics* 38, 371–383). Within the context of valid Edgeworth expansions for such statistics we first derive the set of equations that such a transformation must satisfy and second propose a local solution that is sufficient up to the desired order. Application of these results yields two useful corollaries. First, it is possible to eliminate the first correction term in an Edgeworth expansion, thereby accelerating convergence to the leading term normal approximation. Second, bootstrapping the transformed statistic can yield the same rate of convergence of the double, or prepivoted, bootstrap of Beran (1988, *Journal of the American Statistical Association* 83, 687–697), applied to the original statistic, implying a significant computational saving.

The analytic results are illustrated by application to the family of exponential models, in which the transformation is seen to depend only upon the properties of the likelihood. The numerical properties are examined within a class of nonlinear regression models (logit, probit, Poisson, and exponential regressions), where the adequacy of the limiting normal and of the bootstrap (utilizing the *k*-step procedure of Andrews, 2002, *Econometrica* 70, 119–162) as distributional approximations is assessed.

1. INTRODUCTION

Valid asymptotic corrections to limiting distributions, in the form of Edgeworth expansions, are available for a variety of econometric estimators and tests. However, the use of such expansions specifically as inferential tools has proved somewhat limited. Instead, Bartlett correction, the bootstrap, and Laplace or saddlepoint approximations have proved more popular. In fact, all of these techniques rely on the theory of Edgeworth expansions to justify their higher order

This paper is derived from my Ph.D. thesis, "Higher-Order Asymptotics for Econometric Estimators and Tests," for which thanks for patient and helpful supervision go to Grant Hillier. Comments by Karim Abadir, Francesco Bravo, Giovanni Forchini, Soren Johansen, Paul Marriott, Mark Salmon, and Steve Satchell, participants at the conference "Differential Geometric Methods in Econometrics," held at EUI, Florence, and by two anonymous referees proved most helpful. In particular, I thank Peter Phillips for showing interest in the paper, helping with improving the exposition, and providing me with copies of two unpublished research notes. Financial support in the form of a Leverhulme Special Research Fellowship in Economics and Mathematics is gratefully acknowledged. Address correspondence to Patrick Marsh, Department of Economics, University of York, Heslington, York YO10 5DD, United Kingdom; e-mail: pwnm1@york.ac.uk.

 0266-4666/04 $12.00

validity. For example, see the derivation of Bartlett corrections in McCullagh (1987), the synthesis of the bootstrap and Edgeworth expansion in Hall (1992), and the derivations of saddlepoint approximations in Daniels (1954) and Durbin (1980), all of which illustrate this point.

This paper attempts to make use of the Edgeworth expansion as a practical tool. First we detail precisely the conditions under which the expansion is justified, ensuring that the required conditions are verifiable in a straightforward way. Second we consider the statistic for which we want the expansion as a choice in itself, specifically by choosing a nonlinear transformation so as to fix the properties of the resultant expansion. For example, it will be shown that it is possible to improve the rate of convergence of the leading term, from $O(N^{-1/2})$ to $O(N^{-1})$, where N is the asymptotic argument of the series, in most applications the sample size.

The effect of nonlinear transformations upon the efficacy of asymptotic expansions has been considered in the literature. Phillips and Park (1988) examine the effect that different algebraic formulations of nonlinear hypotheses have on the Wald test. Phillips (1979a) and Niki and Konishi (1986) derive Edgeworth expansions for transformations of a univariate statistic, whereas Phillips (1979b), Taniguchi (1991), and Marsh (2001) consider the Edgeworth expansion of Fisher's z-transformation of the serial correlation coefficient.

Here these results are generalized in the following way. It will be assumed that the statistic, or indeed an approximation to it suitable in distribution, permits a valid, that is, with known order of error, Edgeworth expansion. The statistic may be multivariate and the data from which it is derived may be dependent and heterogeneously distributed. Then we identify what is perceived to be a fundamental difficulty in the usage of such expansions. This is that the oscillatory nature of the polynomial corrections to the leading term asymptotic distribution causes nonmonotonicity, particularly in the tails of the distribution, so-called tail difficulty (see Niki and Konishi, 1986; Hall, 1992, App. V). The suggested transformation will, as in Niki and Konishi (1986), thus be that which eliminates the asymptotic skewness coefficient and thereby limits this tail difficulty to the greatest extent.

The main results of this paper are contained in two theorems and two corollaries. The first theorem derives the set of equations (a cube of coupled second-order partial differential equations) that the transformation that removes asymptotic skewness must satisfy. There will be, in general, a multiplicity of solutions to these equations. Specializing the problem slightly (essentially requiring that the moments depend upon some set of parameters of fixed dimension) the second theorem yields a particular solution. The first corollary shows that we may find a transformation such that the leading term has an order of error of $O(N^{-1})$, rather than the usual $O(N^{-1/2})$. Similarly, the second corollary shows that if the bootstrap is used to approximate the distribution of the transformed statistic the order of error committed is $O(N^{-3/2})$. Although these results are simply the usual orders of errors for both the limiting and bootstrap approximations applied to symmetrically distributed statistics, and are therefore somewhat obvious, they highlight a key point of the paper, which is that the Edgeworth expansion may be employed in such a way as to turn a first-order problem into a second-order one.

To be of practical use these results need to offer at least some advantage over the techniques mentioned at the beginning of this introduction. It is certainly true that derivation of Edgeworth expansions can be a formidable exercise, as exemplified by the derivations contained in Sargan (1976), Phillips (1977a, 1977b), Satchell (1984), Sargan and Satchell (1986), and more recently Linton (1997). However, the same is true for both Bartlett corrections and saddlepoint approximations, and indeed the latter are only available under more restrictive conditions.

The bootstrap has become by far the most widely used higher order inferential tool in econometrics. See, for example, Hall (1992), Horowitz and Savin (2000), and Horowitz (2001), among many others. Although it depends upon Edgeworth-type expansions for its validity, appropriate resampling schemes negate the necessity of calculating the expansion itself. Thus the theoretical complexity of the expansion is replaced by the computational complexity of Monte Carlo and resampling. However, unlike approximations derived from an expansion those from the bootstrap are conditional upon the sample, that is, they apply to the data, rather than the model. For any given sample, calculation of bootstrap critical values for a test is trivial given modern equipment and software; however this needs to be repeated for every sample. To overcome the computational burden in nonlinear models Davidson and MacKinnon (1999) and Andrews (2002) propose a more convenient procedure requiring only a finite number of steps in an iterative optimization routine. Combined with double bootstrap, or prepivoting in the language of Beran (1988), and requiring only two steps in the optimization we have potentially a very powerful tool, delivering an order of error of $O(N^{-3/2})$. However, even with a relatively modest sample size, number of bootstrap iterations (at both levels), and Monte Carlo replications in the thousands, a single experiment would require the generation of billions of pseudorandom numbers, even for a univariate statistic. Although perfectly feasible, this does require both time and the availability of sophisticated pseudorandom number generators having very long periods.

Recalling that a one-level bootstrap will, in general, have an order of error of $O(N^{-1})$, in coverage probability, and for the two-level bootstrap it will be $O(N^{-3/2})$ (see Beran, 1988), this puts the results of the two corollaries of this paper in context. That is we may deliver the same order of error but at a significantly lower computational cost, although this must be offset by the cost in deriving the relevant transformation. The computational saving may be quantified. If the last bootstrap level consists of $100 \times B$ iterations, the computation length for the methods of this paper are $(1/B)\%$ of those for the bootstrap yielding the same order of error. In some circumstances, such as calculation of a symmetric confidence interval or application of a symmetric test, the error in coverage and rejection probabilities for the one-level bootstrap are $O(N^{-2})$. In such cases the practical relevance of transformation methods appears limited in comparison, because this order of error cannot always be attained by a leading term approximation.

The methods of this paper are illustrated in two separate sections. The first applies the theorems to the distribution of the sufficient statistic in the exponential family. Doing so reveals that the transformation required depends only upon the log-likelihood and its derivatives. This result adds to those, including, for example, Durbin (1980) and Armstrong and Hillier (1999), in which properties of, for instance, maximum likelihood estimators (MLE), are derived directly from the likelihood, not from the form of the statistic itself. The second illustration concerns the numerical properties of the two corollaries. Over four nonlinear regression models (probit, logit, Poisson, and exponential regressions), the transformation that removes asymptotic skewness and delivers asymptotic normal inference of order $O(N^{-1})$ is found. Then standard normal quantiles are used as approximations to those of the MLEs (as obtained by Monte Carlo simulation) and those of the transformed statistic. Similarly, bootstrap critical values are obtained for both statistics, in the logit and exponential regression cases, and the nominal and true rejection probabilities analyzed.

The plan of the paper is as follows. Section 2 describes the notation used and the assumption required to justify the Edgeworth expansion for the density and distribution of a multivariate statistic, and it motivates consideration of transformations of that statistic. The main results are then contained in Section 3. Section 4 applies the theorems to the cases where the statistic of interest is the sufficient statistic and a linear combination of its elements. Section 5 provides a numerical comparison first with respect to the limiting standard normal and second with respect to the bootstrap within the nonlinear models mentioned previously, Section 6 then concludes. The proofs of the two theorems are contained in the Appendix.

2. PRELIMINARIES AND MOTIVATION

Before fixing the conditions under which the results of this paper apply, some notation is required for expansions for multivariate statistics. Let X_N be a $k \times 1$ random vector with density $f_N(x)$, distribution $F_N(x)$, characteristic function $M_X(\lambda)$, cumulant generating function $K_X(\lambda) = \ln[M_X(\lambda)]$, and vth cumulant defined by

$$\kappa_X^{I_v} = \kappa_X^{i_1,\ldots,i_v} = \left(\sqrt{-1}\right)^{-v} \left. \frac{\partial^v K_X(\lambda)}{\partial \lambda_1^{i_1} \ldots \partial \lambda_k^{i_v}} \right|_{\lambda=0}; \qquad \sum_j i_j = v.$$

For a full account of the notation involved see McCullagh (1987, Ch. 2), in particular for the generalities of index notation and the summation convention. The statistic X_N may be, for example, an estimator or a test derived from some

https://doi.org/10.1017/S0266466604205084 Published online by Cambridge University Press

sample of size N, or indeed an in-distribution approximation to it, of suitable order. In general we will assume that as $N \to \infty$, $X_N \to_d N(0, \kappa_X^{i_1,i_2})$ and that the density $f_N(x)$ may be expanded as

$$f_N(x) = \phi_k(x; \kappa_X^{i_1,i_2}) \left\{ 1 + \sum_{j=3}^{b} c_{j,N}(\kappa) q_j(x) \right\} + o(N^{-(b-2)/2}), \tag{1}$$

for some $b \geq 3$ and where $\phi_k(x; \kappa_X^{i_1,i_2})$ is the k-dimension normal density with covariance $\kappa_X^{i_1,i_2}$, the coefficients $c_{j,N}(\kappa)$ are an asymptotic sequence in N and depend upon the cumulants of X_N, and the $q_j(x)$ are (Hermite tensoral) polynomials of degree $3j$ in the elements of x.

For details of relevant conditions and assumptions under which (1) holds, the reader is referred to Sargan (1976), Phillips (1977a), Bhattacharya and Ghosh (1978), Durbin (1980), Sargan and Satchell (1986), and Hall (1992), among many others. A sufficient—and, more important, verifiable—condition is given when the cumulants of X_N satisfy the following condition.

Assumption 1. Let $\kappa_X^{I_j}$ be the jth cumulant of X_N. Then

$$\kappa_X^{I_j} = N^{-(j-2)/2} \sum_{l=1}^{\infty} \kappa_{X,l}^{I_j} N^{-(l-1)}, \tag{2}$$

where the $O(1)$ cumulant coefficients $\kappa_{X,l}^{I_j}$ are free of N and we also assume $\kappa_{X,1}^{I_1} = 0$.

Assumption 1 guarantees that in (1) the coefficients $c_{j,N}(\kappa)$ are of order $O(N^{-(j-2)/2})$, which in turn determine the rate of convergence of the approximation. However, this paper is concerned with the application of (1) as an approximation to the finite sample density of X_N. Two aspects influence the accuracy of such approximations. Apart from the number of correction terms included, (i.e., $b - 2$), the accuracy will depend upon where, in the sample space for X_N, we evaluate the approximation. As noted by Niki and Konishi (1986) the polynomials $q_j(x)$ tend to be highly oscillatory in particular for large x (i.e., in the tail areas). Moreover, they find that the highest order polynomial occurs, for every b, with the asymptotic skewness term $\kappa_{X,1}^{I_3}$. Therefore, for the univariate case they suggest transforming X_N via a nonlinear function to remove this term. Other investigations of the impact of nonlinear transformations on the accuracy of asymptotic approximations are contained in Hougaard (1982) (for univariate likelihood), Phillips (1979b) (in autoregression), and Phillips and Park (1988) (for Wald tests of nonlinear restrictions).

An interesting slant on the problem is provided by Hall (1992, App. V), who shows that, the product $\phi_k(x; \kappa_X^{i_1,i_2}) c_{j,N}(\kappa) q_j(x)$ involves the dominant term

$$(N^{-1/2} x^3)^j \phi_k(x; \kappa_X^{i_1,i_2}),$$

so that the integral, and hence the approximating distribution, will not converge at all if $x = O(N^{1/6})$. This is relevant because, for example, if we follow some procedure that decreases the size of a hypothesis test for larger sample sizes, then the critical value of the test will grow with N. If we eliminate asymptotic skewness then the polynomials $q_j(x)$ become of degree $2j$ (if j is even) or $2j-1$ (if j is odd), and so the distribution will not converge if now $x = O(N^{1/4})$.

Evidence that the oscillations of the polynomials $q_j(x)$ do cause tail difficulty that is of concern may be found in the numerical analysis of the Edgeworth series for the ordinary least squares (OLS) estimator for the autoregressive parameter in an AR(1), contained in Phillips (1977b). Nonmonotone behavior of the approximating distribution function manifests itself even for moderate values of the parameter and in reasonable sample sizes. Similar results are contained in the analysis of Edgeworth approximations for the information matrix test in Chesher and Spady (1991).

Later in this paper we consider a family of nonlinear regression models involving a single covariate, $\{w_i\}_{i=1}^N$. For each, it will be shown that the asymptotic skewness coefficient of the bias-corrected MLE (for the coefficient on that covariate) takes the form,

$$\kappa_{\hat{h},1}^{I_3} = \frac{k_\alpha r_w}{\sqrt{N}},$$

where $\hat{h} = h(\hat{\beta})$ is the bias-corrected version of the MLE $\hat{\beta}$, k_α is a constant depending upon the particular model characteristics, and r_w is the skewness coefficient of the covariates w_i (precise details are given in Section 5). To order $O(N^{-1})$ the Edgeworth expansion for the distribution of $\hat{h}$ is

$$\Pr(h(\hat{\beta}) \le h) = \Phi(h) - \phi(h)\left\{\frac{k_\alpha r_w(h^2-1)}{6\sqrt{N}}\right\} + O(N^{-1}). \tag{3}$$

The approximation for the exponential regression case for which $k_\alpha = -1$ for a sample size of $N = 25$ and for two different covariate configurations having skewness of approximately $r_w \approx 12.5$ and $r_w \approx 5.7$, respectively, is plotted in Figure 1.

Clearly the approximation is nonmonotonic; in fact for (3) to be monotonic when $r_w \approx 12.5$, we would require over 50,000 observations. If higher order inference based upon Edgeworth expansions is to be made more practical, then the issue of nonmonotonicity, and hence tail difficulty, needs to be overcome.

3. OPTIMAL TRANSFORMATIONS

In this section we first generalize the results of Phillips (1979a) and Niki and Konishi (1986) to the multivariate case. Following the heuristic argument of the previous section, we call a transformation optimal if it removes asymptotic skewness, thereby reducing the degree of the highest order polynomial in (1).

https://doi.org/10.1017/S0266466604205084 Published online by Cambridge University Press

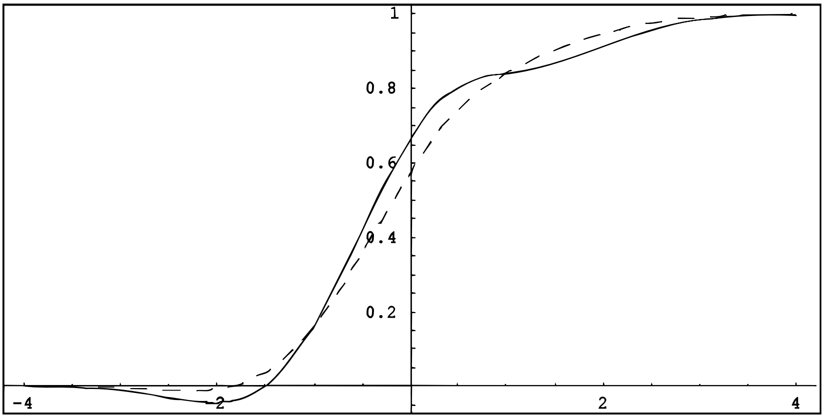

FIGURE 1. Edgeworth approximation (3) for the distribution of $h(\hat{\beta})$, with $N = 25$, for $r_w \approx 12.5$ (solid) and $r_w \approx 5.7$ (dashed) in the exponential case.

Suppose that we have the k-dimension statistic $X = X_N$, with density $f_N(x)$, satisfying Assumption 1, with Edgeworth expansion given by (1) and cumulants $\kappa_X^{I_j}$ and consider an arbitrary nonlinear transformation of the form $g = g(X)$. We will consider transformations satisfying the following assumption.

Assumption 2.

(i) Let the dimension of $g(X)$ be d, where $d \leq k$.
(ii) $g(X)$ is v times differentiable, where $v > b$, with derivatives

$$g_{I_j} = \frac{\partial^j g(X)}{\partial X_1^{i_1} \ldots \partial X_k^{i_k}}; \qquad \sum_{l=1}^{k} i_l = j$$

with the g_{I_v} continuous in a neighbourhood of $\tau = \kappa_X^{I_1} = E[X]$ and all minors of g_{I_1} bounded away from zero.

Assumption 2 requires the following. First, because we are transforming a multivariate statistic, we require the dimension of the transformation to be no larger than the original statistic. Second, we will require that the function $g = g(X)$ may be expanded, around τ, as a power series of the form

$$g^r = \bar{g}^r_{I_0} + \sum_{j=1}^{v} \bar{g}^r_{I_v} Z^{I_v} + O_p(N^{-(v+1)/2}); \qquad r = 1, \ldots, d, \tag{4}$$

where

$$\bar{g}^r_{I_v} = \left. \frac{\partial^v g(X)}{\partial X_1^{i_1} \ldots \partial X_k^{i_k}} \right|_{X=\tau} \quad \text{and} \quad Z = (X - \tau) = O_p(N^{-1/2}).$$

https://doi.org/10.1017/S0266466604205084 Published online by Cambridge University Press

Under Assumption 2, the density and distribution of $g(X)$ permit a valid Edgeworth approximation (see Skovgaard, 1981; Sargan and Satchell, 1986). This assumption is stronger than those used in the latter two papers only in that we desire a transformation that will affect each of the terms in the expansion up to, and including, order b. The density of g may be approximated, for example, to $O(N^{-3/2})$, by

$$f^{(4)}(g;\theta) = \frac{\phi_d(g;\kappa)}{6}\Big\{6 + N^{-1/2}(6\kappa_{g,2}^{I_1}h_{I_1}(x) + \kappa_{g,1}^{I_3}h_{I_3}(x)) + \frac{N^{-1}}{12}(12\kappa_{g,2}^{I_3}h_{I_3}(x) + 3\kappa_{g,1}^{I_4}h_{I_4}(x) + \kappa_{g,1}^{I_3}\kappa_{g,1}^{J_3}h_{I_3,J_3}(x))\Big\}. \quad \textbf{(5)}$$

In (5) $\kappa_{g,j}^{I_v}$ is the jth term in the expansion of the vth cumulant array of the statistic g. Because g is expressible as a power series in X, then the cumulants of g are thus the generalized cumulants of X, as detailed in the Appendix. The transformation we want is such that the asymptotic skewness term vanishes, that is, $\kappa_{g,1}^{I_3} = 0$. This is achieved in the following theorem.

THEOREM 1. *Suppose X satisfies Assumption 1 and $g = g(X)$ satisfies Assumption 2. Then*

$$\kappa_{g,1}^{I_3} = 0$$

if and only if $g(.)$ solves the d-cube of coupled partial differential equations

$$g_{I_1}(\tau)\kappa_{X,1}^{I_3} + 3g_{I_2}(\tau)\kappa_{X,1}^{I_2} = 0. \quad \textbf{(6)}$$

Remarks.

(i) The set of differential equations (6) are invariant with respect to affine transformation; hence issues of standardization (with respect to the mean and variance) do not impact upon solutions to them.

(ii) To eliminate asymptotic skewness $g(.)$ must satisfy (6). However, sometimes the exact cumulants of X will be known, rather than their expansion. Because by definition the exact cumulants and their leading term coincide asymptotically, it may be simpler to solve

$$g_{I_1}(\tau)\kappa_X^{I_3} + 3g_{I_2}(\tau)\kappa_X^{I_2} = 0. \quad \textbf{(7)}$$

(iii) In general, (6) and (7) describe a set of d^3 coupled second-order partial differential equations. Hence, solutions, even when they exist, may be difficult to find (see Hougaard, 1982), whereas if $d < k$ then there will not exist an orthogonal basis for a solution (see McCullagh, 1987, Ch. 5). Overcoming this problem will be the purpose of the next theorem.

https://doi.org/10.1017/S0266466604205084 Published online by Cambridge University Press

(iv) The requirement here is only that asymptotic skewness be eliminated, a much weaker condition than requiring that all excess skewness be removed. Thus, to simplify the problem, we need only look for local (up to an asymptotic order of $O(N^{-1/2})$) solutions. This is still true even if the form (7) is used.

To implement the transformation, that is, to find a feasible solution to (6), we need to specialize the problem. Specifically we impose a parameterization that the density of X is $f_N(x;\theta)$ depending on a d-vector of parameters θ and so X has cumulants given by $\kappa_X^{I_v} = \kappa_X^{I_v}(\theta)$. In this setup, to solve (6) we make a preliminary transformation of the form

$$X \rightarrow h = h(X,\theta),$$

depending both upon the statistic and the parameter. Again we assume that h satisfies Assumption 1 and so permits a valid Edgeworth approximation. The transformation is now $g = g(h)$. Hence a solution to (6) is given by the following theorem.

THEOREM 2. *Let $\bar{g}_i^i$ and $\bar{g}_{ii}^i$ denote the first and second derivatives of the ith element of g with respect to the ith element of h, evaluated at $\kappa_{h,1}^{i_1}$. Then a solution to the set of d ordinary differential equations*

$$\bar{g}_i^i \kappa_{h,1}^{i,i,i} + 3\bar{g}_{i,i}^i = 0; \qquad i = 1,\ldots,d, \tag{8}$$

where $k_{h,1}^{i,i,i}$ is the leading term of the expansion of $\mathrm{cum}(h^i,h^i,h^i)$*, is also a solution to (6). Moreover a solution to (8) exists, unique up to constants of integration.*

Although Theorems 1 and 2 deliver transformations that remove the highest order Hermite polynomial from any bth order approximation, often we are able to obtain a stronger result.

COROLLARY 1. *Suppose that there exists an affine transformation of h, $h \rightarrow \bar{h} = d + Dh$, so that $\tau_{\bar{h}} = E[\bar{h}] = 0 + O(N^{-1})$ and hence $\kappa_{\bar{h},2}^1 = 0$. Defining the transformation, $g = g(\bar{h})$, satisfying (8) and letting*

$$z = (\kappa_{g,1}^{i_1,i_2})^{-1/2}(g(\bar{h}) - g(\tau_{\bar{h}}) + c),$$

$$c = -\kappa_{g,1}^{i_1,i_2}(g_{I_1}(\tau)g_{I_1}(\tau)')^{-1/2}g_{I_2}(\tau),$$

where $\kappa_{g,1}^{i_1,i_2}$ is the asymptotic covariance of g, then

$$f_Z(z) = \phi_k(z)\{1 + O(N^{-1})\}, \tag{9}$$

where $\phi_k(z)$ is the standard normal density.

Proof. Because now also $\kappa^{1}_{\bar{h},2} = 0$ then the only other term of order $N^{-1/2}$ is c, as defined previously, which may be removed because (8) is invariant to affine transformation. ■

We may also consider the transformation in conjunction with an appropriate one-step bootstrap. Suppose that we have a sample $Y_N = (y_1, \ldots, y_N)$ having distribution $P_{N,\theta}$. The statistic of interest is $X = X_N = X(y_1, \ldots, y_N)$, which may be, for example, an estimator for, or test upon, the unknown parameter of interest θ. Let $\hat{\theta}_N$ be an $\sqrt{N}$-consistent estimator of θ. Let Y_N^* be resampled observations having distribution $P_{N,\hat{\theta}_N}$, that is, the empirical distribution of Y_N, and similarly let X^*, h^*, and g^* denote the bootstrapped versions of the statistics considered in Theorems 1 and 2 and Corollary 1. The cumulative distribution functions of X, h, and g will be denoted by $F_N(.,\theta)$, which are to be approximated by $F_N(.,\hat{\theta}_N)$, the empirical distribution of the bootstrapped statistics.

COROLLARY 2. *Suppose that X has an Edgeworth expansion given by (1), $\bar{h}$ satisfies the conditions of Corollary 1, and $g = g(\bar{h})$ solves (8). Then*

(i) $F_N(X,\theta) - F_N(X,\hat{\theta}_N) = O_p(N^{-1})$,
(ii) $F_N(g,\theta) - F_N(g,\hat{\theta}_N) = O_p(N^{-3/2})$.

Proof. Following Beran (1988) the empirical distributions $F_N(.,\hat{\theta}_N)$ permit expansions, uniform in the first argument and local in θ, of the form

$$F_N(X,\hat{\theta}_N) = F_N(X,\theta) + N^{-1/2}\bar{f}_1(X,\theta) + O(N^{-1}),$$

$$F_N(g,\hat{\theta}_N) = F_N(g,\theta) + N^{-1}\bar{f}_2(g,\theta) + O(N^{-3/2}),$$

where $\bar{f}_1(X,\theta)$ and $\bar{f}_2(g,\theta)$ are polynomials derived from the respective Edgeworth expansions for X and g and the rate of convergence is determined by the leading term in that expansion, that is, respectively, from (1) and (9). Consequently, expanding $f_1(X,\theta)$ and $\bar{f}_2(g,\theta)$ around $\hat{\theta}_N$ and noting

$$\bar{f}_1(X,\theta) - \bar{f}_1(X,\hat{\theta}_N) = O_p(N^{-1/2}),$$

$$\bar{f}_2(g,\theta) - \bar{f}_2(g,\hat{\theta}_N) = O_p(N^{-1/2})$$

proves the result. ■

Provided that there is an affine transformation that removes the $O(N^{-1/2})$ term in the expansion of the mean, Corollary 1 contains the stronger result that asymptotic inference is optimized via use of a transformed statistic given by Theorem 2. That is, if we use the limiting normal to obtain probabilities or quantiles, then the transformation minimizes the order of error for these values. Similarly, Corollary 2 optimizes the use of a one-step bootstrap, in the sense that, as is well known, the bootstrap has a faster convergence rate for symmetric rather than asymmetric distributions. In fact, the one-step bootstrap applied to the transformed statistic can have a convergence rate equal to that of the two-step bootstrap applied to the original statistic.

With the general results in the preceding theorems and corollaries, in the next two sections we will analyze the resulting properties of such transformations, first their analytic and then their numerical properties.

4. ANALYTIC PROPERTIES IN THE EXPONENTIAL FAMILY

4.1. General Functions of the Sufficient Statistic

As a general application, consider a sample $Y = \{y_1, \ldots, y_N\}'$, generated from a member of the exponential family, with density

$$f_Y(y;\theta) = \exp\{t'\eta - K(\eta) + r(t)\}, \quad \textbf{(10)}$$

where $\eta = \eta(\theta)$, properties of which are detailed in Barndorff-Nielsen and Cox (1989). Many interesting Gaussian econometric models are exponential models, examples of which are contained in van Garderen (1997).

Inference about θ will be conducted through functions of the minimal sufficient statistic, t. In this section we will derive the transformation of t, as described in the previous section. In particular, it will be seen that this transformation is defined only by the properties of the sample density itself. First we need to check when Assumption 1 holds. The cumulant generating function of t is

$$\begin{aligned} K_t(\lambda) &= \ln \int_{\mathbb{R}^k} \exp\{t'\eta + t'\lambda - K(\eta) + r(t)\}\, dt \\ &= K(\lambda + \eta) - K(\eta), \end{aligned}$$

and hence the cumulants are

$$\kappa_t^{I_v} = \left. \frac{\partial^v K(\lambda + \eta)}{\partial \lambda_1^{i_1} \ldots \partial \lambda_k^{i_k}} \right|_{\lambda = 0} = K^{(v)}(\eta). \quad \textbf{(11)}$$

To proceed, note that if $s = At + a$, where A and a may depend upon N but not η, then (i) s is a canonical statistic and (ii) the sample density may be written

$$f_Y(y;\theta) = \exp\{s'\gamma - K(\gamma) + r^*(t)\}.$$

Thus the exponential form of the density is preserved by affine transformations, and moreover we may assume without loss of generality that $E[t] = 0$ by choosing the constant a appropriately. In particular we can define a canonical statistic $S = N^{1/2}t$, and if in addition the cumulants of S satisfy

$$\kappa_S^{I_v} = K^{(v)}(\gamma) = O(N), \quad \textbf{(12)}$$

https://doi.org/10.1017/S0266466604205084 Published online by Cambridge University Press

then Assumption 1 holds because for S, $\gamma = N^{-1/2}\eta$, then (12) implies

$$\kappa_t^{I_v} = K^{(v)}(\eta) = O(N^{-(2v-1)/2}), \tag{13}$$

noting that higher order terms in the cumulant expansion vanish. As a consequence, the density of t admits a formal Edgeworth expansion.

An analogous result for a nonlinear transformation, $g = g(t)$, may be established, given only that (12) holds, as in the following theorem, proved in Marsh (2001).

THEOREM 3. *Given a transformation $t \to g(t)$, satisfying Assumption 2, the density of g permits a valid Edgeworth expansion provided only that there exists an S satisfying (12).*

Because both t and g permit Edgeworth approximations for their densities, we may apply the criterion of an "optimal" transformation. If we consider the case where $\eta = \theta$, implying that $f_Y(y;\theta)$ is full exponential (i.e., $k = d$), then applying Theorem 2, $g(.)$ must satisfy, using (11),

$$g_{I_1}(\tau)K^{(3)}(\theta) + 3g_{I_2}(\tau)K^{(2)}(\theta) = 0, \tag{14}$$

where $\tau = E(t) = K'(\theta)$. Notice that (14) is specified purely in terms of the properties of the sample density. That is, if we define the log-likelihood for the sample by $l(\theta) = \theta' t - K(\theta) + h(y)$, then $K^{(v)}(\theta) = -l^{(v)}(\theta)$ for $v \geq 2$, and $\tau = t - l'(\theta)$. A solution to (14) is found by applying Theorem 2. Define a symmetric $P = p_j^i$ such that

$$P^{-1}P^{-1} = -l^{(2)}(\theta)$$

and let $z = Pt$, so the cumulants of z are

$$\kappa_z^{i_1,i_2} = p_{j_1}^{i_1} p_{j_2}^{i_2} K^{j_1,j_2}(\theta); \qquad \kappa_z^{i_1,i_2,i_3} = p_{j_1}^{i_1} p_{j_2}^{i_2} p_{j_3}^{i_3} K^{j_1,j_2,j_3}(\theta);$$

then the optimal transformation must satisfy

$$g_i^i(\tau_z)\kappa_{z,1}^{i,i,i} + 3g_{ii}^i(\tau_z) = 0, \qquad i = 1,\ldots,d, \tag{15}$$

where $\tau_z = P\tau$.

This explicit dependence on the likelihood is a further result in the spirit of Durbin (1980), Hougaard (1982), and Armstrong and Hillier (1999). In those papers the object is to derive exact or higher order asymptotic densities of estimators and tests, purely in terms of functionals of the likelihood and its derivatives. Here, the optimal transformation is defined entirely in terms of a differential equation in the derivatives of the log-likelihood.

4.2. Tests of a Particular Form

Often, even for multivariable hypotheses, we require a univariate statistic. In this section we analyze the properties of the transformation acting upon a weighted average of the components of the sufficient statistic. In a slightly altered setup, consider the full exponential model

$$f_Y(y;\gamma) = \exp\{s'\gamma - K(\gamma)\},$$

where $s = \{s_1,\ldots,s_k\}'$ and $\gamma = \{\gamma_1,\ldots,\gamma_k\}$, and the set of statistics defined by

$$h = \sum_{i=1}^{k} a_i s_i, \tag{16}$$

the set of weighted averages of the canonical statistic. Statistics such as (16) encompass point optimal and locally most powerful tests of $H_0 : \gamma = \gamma_0$ against $H_1 : \gamma = \gamma_1$ (see van Garderen, 1998) and also statistics with asymptotically optimal properties (see Elliott, Rothenberg, and Stock, 1996). As a consequence, we look for a transformation of h, $g = g(h)$, satisfying the conditions of Theorem 2. Defining

$$t = As; \qquad A = \operatorname{diag}\{a_i\},$$

then the density may be written

$$f_Y(y;\gamma) = f_Y(y;\theta) = \exp\{t'\theta - K(\theta) + r(t)\},$$

and h then becomes $h = \sum_{i=1}^{k} t_i$. Now from (11), the cumulants of t are

$$\kappa_t^j = \frac{\partial^j K(\theta)}{\partial\theta^{j_1}\ldots\partial\theta^{j_k}}, \qquad \sum_{j_i} = j,$$

and hence the cumulants of h are

$$\kappa_h^j = \sum_{j_1+\cdots+j_k=j} \frac{\partial^j K(\theta)}{\partial\theta_1^{j_1}\ldots\partial\theta_k^{j_k}},$$

and in particular

$$E(h) = \tau_h = \sum_{i=1}^{k} \frac{\partial K(\theta)}{\partial\theta_i}, \tag{17}$$

so that

$$\kappa_h^j = \sum_{j_1+\cdots+j_k=j-1} \frac{\partial^j \tau_h}{\partial\theta_1^{j_1}\ldots\partial\theta_k^{j_k}}.$$

https://doi.org/10.1017/S0266466604205084 Published online by Cambridge University Press

On applying Theorem 1, the optimal transformation of h is given by

$$g'(\tau_h)\left(\sum_{j_1+\cdots+j_k=2}\frac{\partial^j\tau_h}{\partial\theta_1^{j_1}\ldots\partial\theta_k^{j_k}}\right)+3g''(\tau_h)\left(\sum_{j_1+\cdots+j_k=1}\frac{\partial^j\tau_h}{\partial\theta_1^{j_1}\ldots\partial\theta_k^{j_k}}\right)=0. \tag{18}$$

Solutions to (18) will obviously depend upon the particular form of the likelihood; however, rearranging gives

$$\frac{d\ln[g'(\tau_h)]}{d\tau_h}=-\frac{1}{3}\left(\frac{\displaystyle\sum_{j_1+\cdots+j_k=2}\frac{\partial^j\tau_h}{\partial\theta_1^{j_1}\ldots\partial\theta_k^{j_k}}}{\displaystyle\sum_{j_1+\cdots+j_k=1}\frac{\partial^j\tau_h}{\partial\theta_1^{j_1}\ldots\partial\theta_k^{j_k}}}\right),$$

so that, noting (17), the shape of the transformation, once again, is explicitly determined by the shape of the likelihood. The general solution to (18) is

$$g(\tau_h)=C_1+\exp\left\{C_2+\int_{\mathbb{R}}\frac{1}{3}\left(\sum_{j_1+\cdots+j_k=2}\frac{\partial^j\tau_h}{\partial\theta_1^{j_1}\ldots\partial\theta_k^{j_k}}\right)\times\left(\sum_{j_1+\cdots+j_k=1}\frac{\partial^j\tau_h}{\partial\theta_1^{j_1}\ldots\partial\theta_k^{j_k}}\right)^{-1}d\tau_h\right\},$$

for constants C_1 and C_2.

5. NUMERICAL ANALYSIS

In this section we apply the results of Section 3 to some simple likelihood based analysis of nonlinear regression problems. In particular, we consider (a) exponential and (b) Poisson regression models and also two binary choice regression models (c) logit and (d) probit. In general, small sample results for these models are relatively scarce. Some exceptions are the work of McCullagh (1987, Ch. 7), who calculates Bartlett corrections for the first two models, Armstrong and Hillier (1999), who derive an expression for the exact distribution of the MLE for exponential regression, and Horowitz (1994) and Horowitz and Savin (2000) who apply higher order bootstrap methods in binary probit.

Before proceeding, we note some crucial properties of likelihood based on independent observations, which will vastly simplify application of the transformation for all these models. Let y_i, $i=1,\ldots,N$, be independent observations with density $f_i(\beta)$ depending upon some $d\times 1$ parameter, β, and with sample log-likelihood

$$\ln(\beta)=\sum_{i=1}^{N}\ln[f_i(\beta)]$$

and denote the MLE for β by $\hat{\beta}$. Define the generalized mean information measures (see McCullagh, 1987, Ch. 7) by

https://doi.org/10.1017/S0266466604205084 Published online by Cambridge University Press

$$\mathcal{I}_{I_a,\ldots,I_b} = \mathcal{I}^{I_a,\ldots,I_b} = \frac{1}{N} E\left[\left(\frac{\partial \ln(\beta)}{\partial\beta_1\ldots\partial\beta_d}\right)^a \cdots \left(\frac{\partial^v \ln(\beta)}{\partial\beta_1^{j_1}\ldots\partial\beta_d^{j_v}}\right)^b\right],$$

where $\sum_l j_l = v$ and, for example, $-\mathcal{I}_{i_1,i_2}$ is Fisher information. As a consequence, the first three cumulants of $\hat{\beta}$ are asymptotically

$$\begin{aligned}
\kappa_{\hat{\beta}}^{I_1}(\beta) &= \beta + N^{-1}(\mathcal{I}^{i_1,i_2}\mathcal{I}^{j_1,j_2}(\mathcal{I}_{i_2,j_1,j_2} - \mathcal{I}_{i_2,j_1 j_2})) + O(N^{-2}),\\
\kappa_{\hat{\beta}}^{I_2}(\beta) &= N^{-1}\mathcal{I}_{i_1,i_2} + O(N^{-2}),\\
\kappa_{\hat{\beta}}^{I_3}(\beta) &= N^{-2}\mathcal{I}^{i_1,i_2}\mathcal{I}^{j_1,j_2}\mathcal{I}^{k_1,k_2}(\mathcal{I}_{i_2,j_2,k_2} - \mathcal{I}_{i_2 j_2 k_2}).
\end{aligned} \tag{19}$$

For all the models listed previously we will consider approximating the distribution of the standardized, bias-corrected MLE, given by

$$h = h(\hat{\beta}) = \sqrt{\mathcal{I}_{i_1,i_2}}\sqrt{N}(\hat{\beta} - \beta_0 - N^{-1}(\mathcal{I}^{i_1,i_2}\mathcal{I}^{j_1,j_2}(\mathcal{I}_{i_2,j_1,j_2} - \mathcal{I}_{i_2,j_1 j_2}))), \tag{20}$$

with asymptotic cumulants

$$\begin{aligned}
\kappa_{h,1}^{I_1} &= O(N^{-3/2}),\\
\kappa_{h,1}^{I_2} &= I_d + O(N^{-1}),\\
\kappa_{h,1}^{I_3} &= N^{-1/2}\sqrt{\mathcal{I}^{i_1,i_2}}\mathcal{I}^{j_1,j_2}(\mathcal{I}_{i_2,j_2,k_2} - \mathcal{I}_{i_2 j_2 k_2}) + O(N^{-3/2}).
\end{aligned} \tag{21}$$

Crucially, by standardizing the MLE as in (20), the following points should be noted. From (21) Assumption 1 holds for $h(\hat{\beta})$ (with $b = 3$) and moreover for any function satisfying Assumption 2. The elements of $h(\hat{\beta})$, $(h_1,\ldots,h_d)$, are asymptotically independent, and therefore the conditions for Theorem 2 are met; that is, we may transform the h_i individually. Finally, by considering the bias-corrected MLE the conditions of Corollary 1 are met, and the transformation will be optimal. As a consequence, and also to keep the algebra as brief as possible, we will only consider the case where the mean function for all of the regressions depends only upon $\lambda_i = \exp\{\alpha + \beta w_i\}$, for covariates w_i satisfying $\sum_i w_i = 0$. We will assume α is known and that we are interested in a homogeneity hypothesis, that is, $H_0{:}\,\beta = 0$. Assuming α is unknown and replacing it with an $\sqrt{N}$-consistent estimator would not affect the order of error of the approximation.

For each case, the density of the ith observation is

$$\begin{aligned}
&\text{(a)}\quad f_i(\beta) = \lambda_i^{-1}\exp\{-y_i/\lambda_i\},\\
&\text{(b)}\quad f_i(\beta) = \exp\{-\lambda_i\}\lambda_i^{y_i}/y_i!,\\
&\text{(c)}\quad f_i(\beta) = y_i^{p_{1i}}(1-y_i)^{1-p_{1i}}; \qquad p_{1i} = \frac{\lambda_i}{1+\lambda_i},\\
&\text{(d)}\quad f_i(\beta) = y_i^{\Phi_i}(1-y_i)^{1-\Phi_i}; \qquad \Phi_i = \int_{-\infty}^{\ln[\lambda_i]} \phi(z)\,dz,
\end{aligned} \tag{22}$$

where $\phi(z)$ is the standard normal density.

https://doi.org/10.1017/S0266466604205084 Published online by Cambridge University Press

The simplifications we have placed upon the regressions give, for each case and under $H_0 : \beta = 0$,

$$\kappa_{h,1}^{I_3} = N^{-1/2} k_\alpha \frac{\sum_{i=1}^{N} w_i^3}{\left(\sum_{i=1}^{N} w_i^2\right)^{3/2}}, \tag{23}$$

where k_α is a constant depending upon the value of α and the particular characteristics of each likelihood. Indeed, after some algebra, we find

(a) exponential: $k_\alpha = -1$,
(b) Poisson: $k_\alpha = -2e^{-\alpha/2}$,
(c) logit: $k_\alpha = 2\dfrac{\sqrt{e^\alpha}}{1+e^\alpha}(e^a - 1)$,
(d) probit: $k_\alpha = (\phi_\alpha \Phi_\alpha (1-\Phi_\alpha))^{-1/2}(\alpha\Phi_\alpha(1-\Phi_\alpha) + \phi_\alpha)$,

where $\Phi_\alpha = \int_{-\infty}^{\alpha} \phi(z)\,dz$ and $\phi_\alpha = e^{-\alpha^2/2}\big/\sqrt{2\pi}$. So, given a value α, the success of low-order Edgeworth approximations for the density of the standardized MLE $h(\hat{\beta})$, even for these very simple cases, depends crucially upon the value of

$$r_w = \frac{\sum_{i=1}^{N} w_i^3}{\left(\sum_{i=1}^{N} w_i^2\right)^{3/2}}, \tag{24}$$

the sample skewness coefficient of the covariate $w = (w_1, \ldots, w_N)'$.

Because in all cases the conditions of Theorem 2 are met, the transformation we seek satisfies

$$g''(h) + 3k_\alpha r_w g'(h) = 0, \tag{25}$$

so that upon standardization,

$$r(\hat{\beta}) = \sqrt{N}\left(\frac{g(h(\hat{\beta})) - g(h(0))}{g'(h(0))}\right) \to_d N(0,1), \tag{26}$$

and solving (25) we find

$$g(h(\hat{\beta})) = \frac{\exp\{-3k_\alpha r_w h(\hat{\beta})\}}{3k_\alpha r_w}.$$

To assess the numerical performance, for each model we fix $\alpha = 1$ and $N = 25$ and consider two sets of (centered) covariates. Set (i) was generated from an exponential random variable with mean 2.718 and set (ii) from a uniform [0,5], giving values of (24); $r_w^{(i)} = 12.538$ and $r_w^{(ii)} = 5.693$, respectively. Based on likelihoods from (22) and setting $\beta = 0$, $h(\hat{\beta})$ and $r(\hat{\beta})$ were simulated with 20,000 replications, using the internal numerical optimization routine in Mathematica.

The quantiles for the empirical distribution of $h(\hat{\beta})$ and $r(\hat{\beta})$, for each model configuration, are given in Table 1, along with those for the standard normal, the limiting distribution for either. Although the numerical accuracy of the transformation is by no means perfect, there is an obvious improvement over that of the bias-corrected MLE. This is particularly so for the binary choice models and when the covariate has significant skewness. Thus for these cases, the theoretical improvement implied by Corollary 1 manifests itself in the improved ability of the standard normal to approximate the distribution of the transformed statistic (26).

The second set of experiments concerns the application of the bootstrap in these nonlinear models, specifically the logit and exponential regressions. Of interest will be the true cumulative probabilities of bootstrap critical values for

TABLE 1. Quantiles of the empirical distribution of the bias-corrected MLE $h(\hat{\beta})$ and the transformed statistic $r(\hat{\beta})$

	$r_w^{(i)} \approx 12.5$		$r_w^{(ii)} \approx 5.7$		$r_w^{(i)} \approx 12.5$		$r_w^{(ii)} \approx 5.7$	
$N(0,1)$	$h(\hat{\beta})$	$r(\hat{\beta})$	$h(\hat{\beta})$	$r(\hat{\beta})$	$h(\hat{\beta})$	$r(\hat{\beta})$	$h(\hat{\beta})$	$r(\hat{\beta})$
	(a) Exponential Regression				(b) Poisson Regression			
−1.282	−1.143	−1.292	−1.308	−1.306	−1.197	−1.225	−1.200	−1.225
−0.842	−0.771	−0.796	−0.909	−0.844	−0.753	−0.850	−0.753	−0.850
−0.524	−0.481	−0.503	−0.582	−0.535	−0.480	−0.579	−0.481	−0.558
−0.253	−0.249	−0.260	−0.284	−0.262	−0.261	−0.332	−0.259	−0.303
0	−0.030	−0.032	−0.013	−0.011	−0.036	−0.040	−0.029	−0.043
0.253	0.149	0.178	0.216	0.254	0.151	0.201	0.151	0.208
0.524	0.343	0.419	0.452	0.527	0.341	0.486	0.342	0.489
0.842	0.567	0.722	0.735	0.864	0.563	0.851	0.566	0.848
1.282	0.864	1.167	0.110	1.305	0.838	1.361	0.846	1.322
	(c) Binary Logit				(d) Binary Probit			
−1.282	−0.588	−1.070	−0.645	−1.136	−0.467	−0.976	−0.942	−1.092
−0.842	−0.371	−0.671	−0.422	−0.753	−0.323	−0.695	−0.619	−0.790
−0.524	−0.247	−0.442	−0.254	−0.464	−0.161	−0.457	−0.375	−0.467
−0.253	−0.113	−0.214	−0.115	−0.229	0.037	−0.181	−0.153	−0.219
0	0.052	0.034	0.021	0.004	0.192	0.057	0.081	0.030
0.253	0.274	0.282	0.161	0.241	0.339	0.295	0.320	0.266
0.524	0.489	0.551	0.325	0.492	0.523	0.536	0.595	0.549
0.842	0.784	0.861	0.546	0.832	0.908	0.941	0.942	0.884
1.282	1.339	1.295	0.894	1.342	1.616	1.474	1.493	1.353

https://doi.org/10.1017/S0266466604205084 Published online by Cambridge University Press

both the standardized MLE $h(\hat{\beta})$ and the transformed statistic $r(\hat{\beta})$. The procedure used is the k-step bootstrap described by Andrews (2002).

Let $Y_N = (y_1,\ldots,y_N)$ be the data having likelihood $f(Y_N;\beta)$ and yielding MLE $\hat{\beta}$. A bootstrap sample Y_N^* is generated from the likelihood $f(Y_N^*;\hat{\beta})$ and a k-step bootstrap estimator based on a Newton–Raphson iterative routine is defined by

$$\beta_{(k)}^* = \beta_{(k-1)}^* - \left(\frac{\partial^2 \ln f(Y_N^*;\beta)}{\partial\beta\partial\beta'}\right)^{-1}_{\beta=\beta_{(k-1)}^*} \left(\frac{\partial \ln f(Y_N^*;\beta)}{\partial\beta}\right)_{\beta=\beta_{(k-1)}^*}, \tag{27}$$

where $\beta_{(0)}^* = \hat{\beta}$. For the parametric bootstrap to be employed here it is sufficient to consider $k = 1,2$ for the one- and two-step bootstrap to have the same order of error in coverage probability as a bootstrap procedure based on full estimation of the parameter (see Andrews, 2002).

Define the k-step standardized MLE and transformed statistic by $h_{(k)} = h(\beta_{(k)}^*)$ and $r_{(k)} = r(\beta_{(k)}^*)$ and let $F_{h,N} = F_N(h_{(k)};\hat{\beta})$ and $F_{r,N} = F_N(r_{(2)};\hat{\beta})$ be their empirical bootstrap distributions based on B bootstrap iterations. Notice that for the transformed statistic at least two steps are needed to ensure that $r_{(2)}$ and $r(\hat{\beta})$ have coincident asymptotic skewness. We will also consider the two-step bootstrap for $h_{(2)}$. Following Beran (1988) let $F_{2,N} = F_{2,N}(F_N(h_{(2)};\beta_{(2)}^*);\hat{\beta})$ denote the empirical bootstrap distribution of the prepivoted standardized MLE. Finally, denote the bootstrap critical values, of nominal size ρ, by, respectively, $c_{h,N}(\rho)$, $c_{r,N}(\rho)$, and $c_{1,N}(\rho)$ as the $[\rho B]$th element in the bootstrap empirical distributions $F_{h,N}, F_{r,N}$, and $F_{1,N}$. These critical values have error in probabilities of order $O(N^{-1})$, $O(N^{-3/2})$, and $O(N^{-3/2})$ (see Beran, 1988; and Corollary 2 in Section 3 of this paper).

Details of these experiments are as follows. For both the logit and exponential regression, data were generated with $\beta = 0$ and $\alpha = 1$, for two sample sizes, $N = 25$ and $N = 50$; the set of regressors (i) was used (in the case $N = 50$, the same set was sampled twice). The full bias-corrected MLE was evaluated, say, $\hat{\beta}_m$, for $m = 1,\ldots,2{,}500$ Monte Carlo replications. For each value of $\hat{\beta}_m$ a bootstrap sample $Y_N^{b_1}$ was generated and using (27) the k-step bootstrap estimates calculated, and from that, $h_{(1)}^{b_1}$, $h_{(2)}^{b_1}$, and $r_{(2)}^{b_1}$, for $b_1 = 1,\ldots,400$ bootstrap iterations. Similarly the distribution of $h_{(2)}^{b_1}$ itself was bootstrapped (prepivoted) giving values $F_{h,N}^{b_2}$ for $b_2 = 1,\ldots,200$ double bootstrap replications. The bootstrap critical values, of nominal size ρ are then found as the $[\rho B]$th entry in the sorted values for $h_{(1)}^{b_1}$, $r_{(2)}^{b_1}$, and $F_{h,N}^{b_2}$, giving $c_{h,N}(\rho)$, $c_{r,N}(\rho)$, and $c_{1,N}(\rho)$, respectively. The true cumulative probability of these critical values is then calculated from the empirical distributions of the $h(\hat{\beta}_m)$, $r(\hat{\beta}_m)$, and the $h_{(2)}^{b_1}$ for the double bootstrap. The results are presented in Table 2.

The results confirm the relevance of Corollary 2. That is, the one-step bootstrap applied to the transformed statistic does seem to have improved finite sample performance compared with the one-step bootstrap applied to the stan-

https://doi.org/10.1017/S0266466604205084 Published online by Cambridge University Press

TABLE 2. Monte Carlo cumulative probabilities for bootstrap critical values for the bias-corrected MLE $h(\hat{\beta})$ and the transformed statistic $r(\hat{\beta})$

	$N = 25$			$N = 50$		
ρ	$c_{h,N}(\rho)$	$c_{r,N}(\rho)$	$c_{1,N}(\rho)$	$c_{h,N}(\rho)$	$c_{r,N}(\rho)$	$c_{1,N}(\rho)$
0.05	0.208	0.166	0.207	0.089	0.090	0.089
0.10	0.267	0.237	0.267	0.204	0.200	0.203
0.90	0.709	0.728	0.747	0.679	0.804	0.799
0.95	0.775	0.783	0.819	0.765	0.898	0.901
0.05	0.234	0.195	0.177	0.183	0.101	0.081
0.10	0.242	0.245	0.238	0.233	0.162	0.136
0.90	0.696	0.717	0.709	0.708	0.846	0.815
0.95	0.747	0.798	0.759	0.801	0.926	0.858

dardized estimator. However, the asymptotic improvement offered by the transformation applies, in this case, only for one-sided coverage probabilities. For symmetrical confidence intervals the error orders are identical.

The performance of the double, or prepivoted, bootstrap is not significantly better than that of the transformed statistic, whereas the computation time for the latter is 200 times that of the one-step bootstrap. The experiments were run on a 2 MHz PC. A single Monte Carlo replication of 400 bootstrap iterations for a sample size of 25 required approximately 0.8 seconds, and a period of 160 seconds was required for a double bootstrap with 200 iterations at the second level. Consequently, with 2,500 Monte Carlo replications the single bootstrap required just over an hour of computation time whereas the double bootstrap required approximately 1 week. Times for the experiments involving a sample size of 50 were approximately a third longer.

6. CONCLUSIONS

This paper has generalized the results of Phillips (1979a) and Niki and Konishi (1986) on transformations in univariate Edgeworth series to the multivariate case. To solve the system of differential equations that the transformation must satisfy, (6), more structure on the inferential problem was imposed, yielding an asymptotically local solution. Although we lose some generality in the process, the resultant theorem, Theorem 2, seems a powerful tool for higher order inference in a multivariate setting. This is particularly the case if the conditions required for Corollaries 1 and 2 hold. Specifically, we can achieve accelerated asymptotic, that is, $O(N^{-1})$ normal, inference and thus utilize the appropriate tabulated critical values and confidence intervals. Moreover, in combination with the bootstrap, inference of order $O(N^{-3/2})$ is available with a single bootstrap,

https://doi.org/10.1017/S0266466604205084 Published online by Cambridge University Press

offering a considerable computational saving over the equivalent, in order of error, double bootstrap.

A full analysis of the properties of the transformation is obtained in the special case of the exponential family. For general functions of the sufficient statistics, the transformation is defined by a set of second-order partial differential equations in the derivatives of the log-likelihood. A similar result is obtained for transformations of the class of statistics defined as a weighted average of the components of the sufficient statistic. Perhaps more important than the analytic are the numerical properties. For the class of simple nonlinear regression models considered here, the transformation is relatively simple to apply and has reasonable numerical properties.

Although the theoretical properties of such transformations are precisely detailed in this paper, one potential weakness is the practical difficulty of obtaining the transformation in cases more complex than considered here. The transformed statistic is found by solving the set of equations given in (8); how demanding this might be will vary from case to case. However, a possible way forward might be to utilize methods analogous to those of Andrews (2002): specifically, to find numerical solutions to these equations that are equivalent to the analytic ones, up to some appropriate order.

REFERENCES

Andrews, D.W.K. (2002) Higher-order improvements of a computationally attractive *k*-step bootstrap for extremum estimators. *Econometrica* 70, 119–162.

Armstrong, M. & G.H. Hillier (1999) The density of the maximum likelihood estimator. *Econometrica* 67, 1459–1470.

Barndorff-Nielsen, O.E. & D. Cox (1989) *Asymptotic Techniques for Use in Statistics*. Chapman and Hall.

Beran, R. (1988) Prepivoting test statistics: A bootstrap view of asymptotic refinements. *Journal of the American Statistical Association* 83, 687–697.

Bhattacharya, R.N. & J.K. Ghosh (1978) On the validity of the formal Edgeworth expansion. *Annals of Statistics* 6, 434–451.

Chesher, A. & R. Spady (1991) Asymptotic expansion of the information matrix test statistic. *Econometrica* 59, 787–815.

Daniels, H.E. (1954) Saddlepoint approximations in statistics. *Annals of Mathematical Statistics* 25, 631–650.

Davidson, R. & J.G. MacKinnon (1999) Bootstrap testing in nonlinear models. *International Economic Review* 40, 487–508.

Durbin, J. (1980) Approximations for the densities of sufficient estimates. *Biometrika* 77, 311–333.

Elliott, G., T.J. Rothenberg, & J.H. Stock (1996) Efficient tests for an autoregressive unit root. *Econometrica* 64, 813–836.

Hall, P. (1992) *The Bootstrap and Edgeworth Expansion*. Springer Series in Statistics. Springer-Verlag.

Horowitz, J.L. (1994) Bootstrap-based critical values for the information matrix test. *Journal of Econometrics* 61, 395–411.

Horowitz, J.L. (2001) The bootstrap. In J.J. Heckman & E. Leamer (eds.), *Handbook of Econometrics*, vol. 5, chap. 52, pp. 3159–3228. Elsevier Science.

Horowitz, J.L. & N.E. Savin (2000) Empirically relevant critical values for hypothesis tests: A bootstrap approach. *Journal of Econometrics* 95, 375–389.

Hougaard, P. (1982) Parameterizations of nonlinear models. *Journal of the Royal Statistical Society, Series B* 44, 244–252.
Linton, O. (1997) An asymptotic expansion in the GARCH(1,1) model. *Econometric Theory* 13, 558–581.
Marsh, P.W.N. (2001) Edgeworth expansions in Gaussian autoregression. *Statistics and Probability Letters* 54, 233–241.
McCullagh, P. (1984) Local sufficiency. *Biometrika* 71, 233–244.
McCullagh, P. (1987) *Tensor Methods in Statistics*. Chapman and Hall.
Niki, N. & S. Konishi (1986) Effects of transformations in higher order asymptotic expansions. *Annals of the Institute of Statistical Mathematics* 38, 371–383.
Phillips, P.C.B. (1977a) Approximations to some finite sample distributions associated with a first-order stochastic difference equation. *Econometrica* 45, 463–485.
Phillips, P.C.B. (1977b) A general theorem in the theory of asymptotic expansions as approximations to the finite sample distribution of econometric estimators. *Econometrica* 45, 1517–1534.
Phillips, P.C.B. (1979a) Expansions for Transformations of Statistics. Research note, University of Birmingham, U.K.
Phillips, P.C.B. (1979b) Normalising transformation for $\hat{\alpha}$. Research note, University of Birmingham, U.K.
Phillips, P.C.B. & J. Park (1988) On the formulation of Wald tests of nonlinear restrictions. *Econometrica* 45, 463–485.
Sargan, J.D. (1976) Econometric estimators and the Edgeworth expansion. *Econometrica* 44, 421–448.
Sargan, J.D. & S. Satchell (1986) A theorem of validity for Edgeworth expansions. *Econometrica* 54, 189–213.
Satchell, S. (1984) Approximation to the finite sample distribution for non-stable first-order stochastic difference equation. *Econometrica*, 52, 1271–1289.
Skovgaard, I.M. (1981) Transformation of an Edgeworth expansion by a sequence of smooth functions. *Scandinavian Journal of Statistics* 8, 207–217.
Taniguchi, M. (1991) *Higher Order Asymptotic Theory for Time Series Analysis*. Lecture Notes in Statistics. Springer-Berlin.
van Garderen, K.-J. (1997) Curved exponential models in statistics. *Econometric Theory* 15, 771–790.
van Garderen, K.-J. (1998) An alternative comparison of classical tests: Assessing the effects of curvature. In M. Salmon & P. Marriott (eds.), *Applications of Differential Geometry to Econometrics*, pp. 230–280. Cambridge University Press.

APPENDIX

We will require the cumulants of products of the elements of X, the generalized cumulants, as detailed in McCullagh (1987, Ch. 3). For our purposes the first four will suffice and, for reference, are

$$\kappa^{i_1,i_2 i_3} = \kappa^{i_1,i_2,i_3} + \kappa^{i_1}\kappa^{i_2,i_3} + \kappa^{i_3}\kappa^{i_1,i_2},$$

$$\kappa^{i_1 i_2,i_3 i_4} = \kappa^{i_1,i_2,i_3,i_4} + \kappa^{i_2}\kappa^{i_1,i_3,i_4}[3] + \kappa^{i_1,i_2}\kappa^{i_3,i_4}[3] + \kappa^{i_1,i_2}\kappa^{i_3}\kappa^{i_4}[3],$$

$$\kappa^{i_1 i_2,i_3 i_4} = \kappa^{i_1,i_2,i_3,i_4} + \kappa^{i_1}\kappa^{i_2,i_3,i_4}[2] + \kappa^{i_3}\kappa^{i_1,i_2,i_4}[2] + \kappa^{i_1,i_3}\kappa^{i_2,i_4}[2],$$

$$+ \kappa^{i_1}\kappa^{i_3}\kappa^{i_2,i_4}[4],$$

$$\kappa^{i_1,i_2,i_3 i_4} = \kappa^{i_1,i_2,i_3,i_4} + \kappa^{i_3}\kappa^{i_1,i_2,i_4}[2] + \kappa^{i_1,i_3}\kappa^{i_2,i_4}[2], \tag{A.1}$$

https://doi.org/10.1017/S0266466604205084 Published online by Cambridge University Press

where $[i]$ implies summation over similar objects, for example, $\kappa^{i_1,i_3}\kappa^{i_2,i_4}[2] = \kappa^{i_1,i_3}\kappa^{i_2,i_4} + \kappa^{i_1,i_4}\kappa^{i_2,i_3}$.

Proof of Theorem 1. The transformed statistic is, upon expansion,

$$g = g(\tau) + (X-\tau)' g_{i_1,i_2}(\tau) + \frac{1}{2}(X_N-\tau)' g_{i_1,i_2}(X_N-\tau) + \cdots,$$

and because $E(X_N-\tau) = O_p(N^{-1/2})$, then it is sufficient to consider only up to the quadratic term. We write

$$\tilde{g}^r = \bar{g}^r + \bar{g}^r_{i_1} Z^{i_1} + \bar{g}^r_{i_1,i_2} X^{i_1} X^{i_2}; \qquad r = 1,\dots,d,$$

where $Z = X-\tau$, $\bar{g}_{I_v} = (\partial^v g(X)/\partial X_1^{i_1}\dots\partial X_k^{i_k})|_{X=\tau}$ and note that $\tilde{g}^r$ and g^r have identical asymptotic skewness terms, which follows from before and the expansion of cumulants in terms of generalized cumulants. Hence, for notational convenience we drop the tildes. Following McCullagh (1987, Ch. 3),

$$g^r = (G_0^r + G_1^r + G_2^r) Z^r,$$

where $(G_0^r + G_1^r + G_2^r)$ is an operator acting on X, with, for example, $G_2^r Z$ containing all quadratic and bilinear terms in the elements of X. Denote this operator by D^r, for a generic element of g; then the cumulant generating function of g is simply

$$K_g(\lambda) = \exp\{\lambda_r D^r\}\kappa_x; \tag{A.2}$$

that is, $\exp\{\lambda_r D^r\}$ is an operator acting upon the cumulants of X. Expansion of this cumulant operator gives

$$K_g(\lambda) = \{1 + \lambda_{r_1} D^{r_1} + \lambda_{r_1}\lambda_{r_2} D^{r_1} D^{r_2} + \lambda_{r_1}\lambda_{r_2}\lambda_{r_3} D^{r_1} D^{r_2} D^{r_3} + \cdots\}\kappa_x,$$

with

$$D^{r_1}\kappa_x = \bar{g}^{r_1} + \bar{g}^{r_1}_{i_1}\kappa_X^{i_1} + \frac{1}{2}\bar{g}^{r_1}_{i_1,i_2}\kappa_X^{i_1,i_2}.$$

The compound operators produce terms, for example, for the cubic,

$$G_1^{r_1} G_1^{r_2} G_1^{r_3} = \bar{g}^{r_1}_{i_1}\bar{g}^{r_2}_{i_2}\bar{g}^{r_3}_{i_3}\kappa^{I_3},$$

and so on. Consequently, the cumulant generating function may be written as

$$\begin{aligned} K_g(\lambda) = {} & \lambda_r\left[\bar{g}^{r_1} + \bar{g}^{r_1}_{i_1}\kappa_X^{i_1} + \frac{1}{2}\bar{g}^{r_1}_{i_1,i_2}\kappa_X^{i_1,i_2}\right] \\ & + \lambda_{r_1}\lambda_{r_2}\left[\bar{g}^{r_1}_{i_1}\bar{g}^{r_2}_{i_2}\kappa_X^{i_1,i_2} + \bar{g}^{r_1}_{i_1}\bar{g}^{r_2}_{i_2,i_3}\kappa_X^{i_1,i_2i_3}[2] + \bar{g}^{r_1}_{i_1,i_2}\bar{g}^{r_2}_{i_3,i_4}\kappa_X^{i_1i_2,i_3i_4}\right] \\ & + \lambda_{r_1}\lambda_{r_2}\lambda_{r_3}\left[\bar{g}^{r_1}_{i_1}\bar{g}^{r_2}_{i_2}\bar{g}^{r3}_{i_3}\kappa_X^{i_1,i_2,i_3} + \bar{g}^{r_1}_{i_1}\bar{g}^{r_2}_{i_2}\bar{g}^{r_3}_{i_3,i_4}\kappa_X^{i_1,i_2,i_3i_4}[3]\right. \\ & \left. + \bar{g}^{r_1}_{i_1}\bar{g}^{r_2}_{i_2,i_3}\bar{g}^{r_3}_{i_4,i_5}\kappa_X^{i_1,i_2i_3,i_4i_5}[3] + \bar{g}^{r_1}_{i_1,i_2}\bar{g}^{r_2}_{i_3,i_4}\bar{g}^{r3}_{i_5,i_6}\kappa_X^{i_1i_2i_3,i_4i_5i_6}\right] \\ & + [\text{Higher order terms}]. \end{aligned} \tag{A.3}$$

https://doi.org/10.1017/S0266466604205084 Published online by Cambridge University Press

Now, for the skewness term, enclosed in the third set of brackets, consider the objects in the third term, which contribute to the asymptotic skewness of g, namely,

$$\bar{g}_{i_1}^{r_1}\bar{g}_{i_2}^{r_2}\bar{g}_{i_3}^{r3}\kappa_X^{i_1,i_2,i_3}+\bar{g}_{i_1}^{r_1}\bar{g}_{i_2}^{r_2}\bar{g}_{i_3,i_4}^{r_3}\kappa_X^{i_1,i_2,i_3 i_4}[3].$$

From (A.1) we have that

$$\kappa_X^{i_1,i_2,i_3 i_4}=\kappa_X^{i_1,i_2,i_3,i_4}+\kappa_X^{i_1}\kappa_X^{i_2,i_3,i_4}[2]+\kappa_X^{i_1,i_2}\kappa_X^{i_3,i_4}[2],$$

so, noting the cumulant expansions given in Assumption 1, particularly

$$\kappa_{x,1}^{i_1}=0,\quad\text{and}\quad\kappa_X^{i_1,i_2,i_3,i_4}=O(N^{-1}),$$

denoting asymptotic equivalence by $\sim$, that is, $a_n\sim b_n\Rightarrow a_n/b_n=O(1)$, we have

$$\kappa_X^{i_1,i_2,i_3 i_4}\sim\kappa_X^{i_1,i_2}\kappa_X^{i_3,i_4}[2]\sim\kappa_{x,1}^{i_1,i_2}\kappa_{x,1}^{i_3,i_4}[2].$$

Further, again because $\kappa_{x,1}^{i_1}=0$,

$$\kappa_X^{i_1,i_2}=\kappa_X^{i_3,i_4}=\Sigma=\text{var}[X]\sim\kappa_{x,1}^{i_1,i_2},$$

which implies that

$$\kappa_X^{i_1,i_2,i_3 i_4}\sim 2(\kappa_{x,1}^{i_1,i_2})^2, \tag{A.4}$$

because $E[X^{i_1}X^{i_2}]=E[X^{i_2}X^{i_1}]$.

We note that (A.4) holds for all permutations over i_1,i_2,i_3,i_4, giving

$$\kappa_X^{i_1,i_2,i_3 i_4}[3]=6(\kappa_{x,1}^{i_1,i_2})^2.$$

Thus to reduce asymptotic skewness to zero, to our desired order, and because $\kappa_X^{i_1,i_2,i_3}\sim\kappa_{x,1}^{i_1,i_2,i_3}$ we need to solve

$$\bar{g}_{i_1}^{r_1}\bar{g}_{i_2}^{r_2}\bar{g}_{i_3}^{r3}\kappa_{x,1}^{i_1,i_2,i_3}+6\bar{g}_{i_1}^{r_1}\bar{g}_{i_2}^{r_2}\bar{g}_{i_3,i_4}^{r_3}(\kappa_{x,1}^{i_1,i_2})^2=0.$$

Premultiplying twice by the inverse of the first derivative matrix (which exists by Assumption 2(ii)) gives

$$\bar{g}_{i_3}^{r_3}\kappa_X^{i_1,i_2,i_3}+6\bar{g}_{i_3,i_4}^{r_3}(\kappa_X^{i_1,i_2})^2=0, \tag{A.5}$$

and noting the definition of $\bar{g}_{i_3,i_4}^{r_3}$, for $r_3=1,\ldots,d$, (A.5) proves the result. ■

Proof of Theorem 2. Suppose initially $k=d$. Now X has asymptotic cumulants $\kappa_{x,1}^{i_1}$, $\kappa_{x,2}^{i_1,i_2}$, and so on, and in particular consider

$$\kappa_{x,1}^{i_1,i_2}=\Omega(\theta),$$

which is a function of the m parameters and is by definition is positive definite. Under the null $H_0:\theta=\theta_0$; $\kappa_X^{i_1,i_2}=\Sigma(\theta_0)$ with θ_0 a fixed point in Ω_θ. Hence there exists a positive definite matrix $P_0^{-1}=P(\theta_0)^{-1}$ such that

$$P_0^{-1}P_0^{-1}=\Sigma(\theta_0).$$

https://doi.org/10.1017/S0266466604205084 Published online by Cambridge University Press

We define

$z_0 = P_0 x,$

so z has cumulants $\kappa_{z_0}^{i_1}$, I_k, $\kappa_{z_0}^{i_1,i_2,i_3}$, and so on. For any $\theta \in \Omega'_\theta = \{\theta : |\theta - \theta_0| < \delta/\sqrt{N}\}$ we say z is locally canonical for families $f_Y(y;\theta_0 + \delta/\sqrt{N})$. Moreover, for $\theta \in \Omega'_\theta$ and $z_\theta = P(\theta)x$ we have

$\Pr\{|z_\theta - z_{\theta_0}| > \delta\} = O(N^{-1/2}),$

and denoting the cumulants of z_θ appropriately gives

$$\kappa_{z_0}^{I} = \kappa_{z_\theta}^{I}\{1 + O(N^{-1/2})\} \tag{A.6}$$

for all $|I| = j$.

That is, the asymptotic cumulants of z_θ and z_{θ_0} agree. Because, asymptotically, the elements of z_0 are independent we choose the ith element of g to be a function of the ith element of z_0 alone. Thus we have the derivatives of $g(.)$ as

$$\bar{g}_{i_1} = \begin{pmatrix} \bar{g}_1^1 & . & 0 \\ \vdots & . & \vdots \\ . & . & \bar{g}_1^d \end{pmatrix} \quad \text{and} \quad \bar{g}_{i_1,i_2} = \begin{pmatrix} \bar{g}_{11}^1 & . & 0 \\ \vdots & . & \vdots \\ . & . & 0 \end{pmatrix} : \ldots : \begin{pmatrix} 0 & . & 0 \\ \vdots & . & \vdots \\ . & . & \bar{g}_{dd}^1 \end{pmatrix}. \tag{A.7}$$

Substituting (A.7) into system (6) gives

$$\bar{g}_i^i \kappa_{z_0,1}^{i,i,i} + 3\bar{g}_{ii}^i = 0, \qquad i = 1, \ldots, d, \tag{A.8}$$

defined for the range of $\kappa_{z_0}^{i_1}$ in $\mathbb{R}^d$, Ω_κ, say, implicitly defined by the set Ω'_θ, such that (A.6) holds. To prove the existence of a solution we note first that each equation in system (A.8) is uncoupled from every other, and second that each is simply a linear, homogeneous second-order differential equation. Finally, a sufficient condition for a solution is that $\kappa_{z_0,1}^{i,i,i}$ is a continuous function of $\kappa_{z_0,1}^{i_1}$for $\kappa_{z_0,1}^{i_1} \in \Omega_\kappa$, and sufficient for this is that $P(\theta)$ is continuous in Ω'_θ. Consider

$P\Sigma(\theta)P = I_d.$

Denoting derivatives with respect to θ by P_θ and Σ_θ, we have

$P_\theta \Sigma P^{-1} + P^{-1}\Sigma P_\theta = \Sigma_\theta,$

and Σ_θ exists in Ω'_θ, because $b \geq 3$ in Assumption 1; then so does P_θ. Hence P is continuous in a neighborhood of θ_0.

Returning to the $k > d$ case, we make a "pre-preliminary" transformation as

$$x \to \begin{pmatrix} \hat{\theta} \\ c \end{pmatrix}, \tag{A.9}$$

https://doi.org/10.1017/S0266466604205084 Published online by Cambridge University Press

where $\hat{\theta}$ is the MLE for θ, $c = Bx$, and B is a $(k-d) \times k$ matrix and is chosen, using the implicit function theorem, such that

$$B \frac{\partial \hat{\theta}}{\partial x_1 \ldots \partial x_k}\bigg|_{x=\tau} = 0 \quad \text{and} \quad BB' = I_{k-d}. \tag{A.10}$$

Now $\hat{\theta}$ is locally sufficient in an $O(N^{-1/2})$ neighborhood of θ_0 (see McCullagh, 1984) in that the density of X may be factored

$$f_X(x;\theta) = G(\hat{\theta};\theta)\{1 + O(N^{-1/2})\},$$

and by (A.10) $\hat{\theta}$ and c are independent in this neighborhood. Hence, marginalizing locally, noting from (A.10) that the Jacobian of the transformation (A.9) does not depend on B, and then transforming from $\hat{\theta}$ to $z = P\hat{\theta}$, where $PP' = \text{Var}(\hat{\theta})$ retains the local properties required for the solution in the $k = d$ case, the analysis follows similarly. ■

https://doi.org/10.1017/S0266466604205084 Published online by Cambridge University Press