1. Introduction

Quarterly or monthly data over ten to fifteen years have become available recently for a large number of economic and firm variables. This makes various dynamic econometric methods, such as distributed lag, auto-correlated errors, and optimal prediction, at least worth trying. At the present time, however, only the asymptotic theory is available for judging the performance of these statistical methods, whereas the length of data available to us is such that the asymptotic theory should neither be rejected on the basis of obvious irrelevance nor be accepted without a careful investigation.

Professor Nagar in a series of papers developed the formulae for the bias to the order of $T^{-1}$ and for the variance to the order of $T^{-2}$, where $T$ is the length of data, for a number of estimators that arise for various models. This kind of approach might be useful for examining the performance of dynamic econometric methods in our particular use. It is rather unfortunate therefore that the method that Nagar used has been subjected to a criticism by Srinivasan [22] denying its validity at least under situations as unrestricted as Nagar has.1)

1) The current state of our knowledge is chaotic indeed. On the one hand Mariano and Sawa [14] have proved that the moments of any order do not exist for the limited information maximum likelihood estimator at least when the equation contains two endogenous variables. On the other hand Kadane [7] has derived an expansion formula for the moments of the maximum likelihood estimators for a general case that includes Mariano-Sawa's. Kadane uses essentially the same method as Nagar. Evidently Nagar's method gives an approximation to something that does not exist, which actually has been known for the Two Stage Least Squares since the important, pioneering work by Basmann [5].

* I have greatly benefited from frequent discussions with Professor Takamitsu Sawa. I am also grateful to Professor Masashi Okamoto and the referee of the paper for giving me important advice on the adaptation of the Lomnicki-Zaremba's theorem, on which I had initially an error. The correspondence with Professors A. L. Nagar, T. N. Srinivasan, and Arnold Zellner has been useful.

---

1973年8月
Prior to the Nagar-Srinivasan controversy there had been a series of closely related developments. The Nagar's method may be looked upon as a kind of the method of statistical differentials, in which the expectation of a function is obtained through the expectations of the terms that appear in the Taylor expansion of the function. The method is used by Klein [9, p. 258] and later by Goldberger, Nagar, and Odeh [6] to derive the covariance matrix of forecasts to the order of $T^{-1}$. Zellner [24] criticized the method, questioning the validity of the theorem conjectured by Klein and allegedly proved by Goldberger, Nagar, and Odeh.

In the statistical literature the method of statistical differentials seems to have a long history (Mann and Wald [12]). It was used in Bartlett [4] and Kendall [8] to obtain the moments of auto-correlation estimates. This induced a theorem by Lomnicki and Zaremba [11]. However, their proof assumes that the function is bounded uniformly with respect to $T$. This assumption has to be weakened to gain the usefulness of their theorem in statistical analysis in general, and in the derivation of moments of econometric estimators in particular. The success in this attempt would restore the Nagar's method as well as the Klein's.

The present paper has triple purposes. (1) We derive the condition under which the $k$-class estimators for the econometric model have moments. (2) Through modifying the Lomnicki-Zaremba's theorem on the method of statistical differentials we derive the approximation formulae for the moments of the $k$-class estimators to the order of $T^{-2}$. (3) We examine various econometric formulae that are allegedly based upon the method of statistical differentials.

2. The Model and the Notations

We shall deal with the same model as Nagar [15 and 16]. I mention here only a few important points. The equation, which is a part of the simultaneous equations system, contains 1 endogenous variable on the left-hand side, $m$ endogenous variables without lags and $l$ exogenous variables on the right-hand side. The system contains $A$ exogenous variables. (The order condition for the identifiability of the equation is $A-l-m \geq 0$.) The disturbances have a normal distribution that is independent and identical over time.

Other notations are also identical to Nagar's [15] unless explained specifically. Those that will appear in the present paper are as follows. $X$ for the exogenous variables, $X_i$ for the included exogenous variables, $Y$ for the endogenous variables on the right-hand side, $Y = X\overline{Y} + \overline{V}$ for the reduced form of $Y$, $y$ for the endogenous variable on the left-hand side, $y = Y_0 + X_0 + u$ for the equation to be estimated, $X = [X_i, X_j]$, $\overline{X} = [\overline{X}_i, \overline{X}_j]$, $\overline{X} = [X\overline{Y}, X_i]$, $Q = (Z'Z)^{-1}$, $M_i = I - X_i(X_i'X_i)^{-1}X_i'$, $M^* = X(X'X)^{-1}X'$, $V_i = [\overline{V}, \overline{O}]$, $C^* = T^{1/2}E(\overline{V}'\overline{V})$.

Needless to say that we need the four assumptions that Nagar makes in [15]. However they are clearly inadequate for our analysis. I add

**Assumption V** $\text{det } T^{-1}Z'Z$, $\text{det } T^{-1}X_0'X_0$, and $\text{det } T^{-1}X_i'X_i$ are non-vanishing.

Only the Two Stage Least Squares method will be considered in the main text.
3. The Existence of Moments of the Two Stage Least Squares Estimates

To avert shame of approximating something which does not exist, I begin with investigating the existence of moments of the Two Stage Least Squares estimators. There has been a large amount of literature on this problem since the pioneering work of Basmann [5]. Dealing with a more restricted model than ours, that is, the case where only two endogenous variables are included in the equation (i.e. \( m=1 \)), Richardson [17], Sawa [19] and Takeuchi [23] gave in terms of the degree of overidentification, \( A-l-m \), what amounts to the necessary and sufficient condition for the existence of moments of the Two Stage Least Squares estimates. Dealing with the same model as ours, Mariano [13] gave the necessary and sufficient condition for the existence of even order moments. In the present paper a sufficient condition will be given for the existence of moments of any order, even or odd, for the model where \( m \) is not restricted to 1.

Let \( e \) be the error of the Two Stage Least Squares estimator vector, which has \((l+m)\) components. The boundedness of \( E|e|^r \) is our concern. Let \( \hat{Z}=[X\hat{U}, X_i] \), where \( \hat{U} \) is the least squares estimates of \( U \). Let \( \hat{Z}^{(i)} \) be the matrix that is obtained from \( \hat{Z} \) by replacing its i-th column by \( u \). The i-th element of \( e \) is expressed as \( \text{det} T^{-1}\hat{Z}'\hat{Z}(i) \cdot \text{det} T^{-1}\hat{Z}'\hat{Z} \). Since \( |\text{det} T^{-1}\hat{Z}'\hat{Z}^{(i)} \cdot \text{det} T^{-1}\hat{Z}'\hat{Z} | \leq (\text{det} T^{-1}\hat{Z}'\hat{Z}^{(i)})| (\text{det} T^{-1}\hat{Z}'\hat{Z})^{-1} \), where \( | \cdot | \) denotes the absolute value, it suffices for the first order moments of \( e \) to show that the expected value of \( (\text{det} T^{-1}\hat{Z}'\hat{Z}^{(i)})^{1/2} \cdot \text{det} T^{-1}\hat{Z}'\hat{Z}^{-1/2} \) is bounded.

Put \( p=(\text{det} T^{-1}\hat{Z}'\hat{Z}^{(i)})^{1/2} \) and \( q=(\text{det} T^{-1}\hat{Z}'\hat{Z})^{1/2} \). The random variables involved in \( q \) are \( V \), and those in \( p \) are \( V \) and \( u \). Let \( z_j \) be \( T^{-1/2} \) the j-th column of \( \hat{Z} \), and let \( z_j^{(i)} \) be \( T^{-1/2} \) the j-th column of \( \hat{Z}^{(i)} \). We take \( i=1 \). Then with the exception of \( j=1 \), \( z_j=z_j^{(1)} \). Let \( R \) and \( R^{(i)} \) be the \((m+l) \times (m+l) \) matrix, the \((j, h) \) element of which is respectively \( z_j'z_h/||z_j|| \cdot ||z_h|| \) and \( z_j^{(i)}'z_h^{(i)}/||z_j^{(i)}|| \cdot ||z_h^{(i)}|| \). Then \( p=\Pi_j||z_j^{(i)}| \cdot (\text{det} R^{(i)})^{1/2} \) and \( q=\Pi_j||z_j|| \cdot (\text{det} R)^{1/2} \). We are concerned with \( E(pq^{-1}) \).

The following procedure is divided into two steps. Initially \( E(pq^{-1}) \) is considered. \( q^{-1} \) is completely determined by the fixed values of the conditioned variables in the above expectation since \( ||z_{m+1}||, \ldots, ||z_m|| \) are nonstochastic and fixed. It will be shown that this conditional expectation of \( p \) is bounded except over a linear subspace of \( V \) that is defined independently of the conditioned variable and that has the measure zero in terms of the joint probability of \( V \). Deleting this subspace, which should have no effect upon \( E(pq^{-1}) \), we shall show in the second step that \( E(q^{-1}) \) is bounded.

(i) In view of the fact that \( 0 \leq (\text{det} R^{(i)})^{1/2} \leq 1 \), we have \( E(pq^{-1})=\Pi_j||z_j^{(i)}|| \cdot E(||z_j^{(i)}|| \cdot ||z_j||, \ldots, ||z_m||, (\text{det} R^{(i)})^{1/2}) \); that is, if the left-hand side is unbounded for given values of \( ||z_j||, \ldots, ||z_m|| \), and \( (\text{det} R)^{1/2} \), so is the right-hand side, and hence the conditional expectation of \( ||z_j^{(i)}||^{2} \) is unbounded. This conditional expectation may be taken in two steps. In view of the relationships \( ||z_j^{(i)}||^{2} = T^{-1} u'u \) and \( u=v-\hat{U}' \),
where $v$ is the random disturbance for $y$ in the reduced form, the expectation of $\|z_i\|^2$ conditional upon $\tilde{V}$ is taken first, using the assumption that rows of $\tilde{V}$ are independently and identically distributed with a $m$ dimensional normal distribution, and, then the expectation over the domain of $\tilde{V}$ that is restricted by the fixed values of $\|z_1\|, \ldots, \|z_m\|$ and $(\det R)^i$. If the conditional expectation is unbounded, at least one column of $\tilde{V}$ must lie in the null-space of $X(X'X)^{-1}X'$. Since this null-space is independent of the particular, fixed values of $\|z_1\|, \ldots, \|z_m\|$ and $(\det R)^i$—note that $X$ is not a variable but $V$ is—, we conclude that the conditional expectation of $p$ is unbounded only when $\tilde{V}$ is in the region having the zero probability measure.

(ii) Under the Assumption $V$, $q^2=\det T^{-1}X_i'X_i\cdot\det T^{-1}\tilde{U}_iX_i'M_iX_i\tilde{U}_i$. Consider $\det T^{-1}\tilde{U}_iX_i'M_iX_i\tilde{U}_i$. Under the assumption about $\tilde{V}$, $T^{-1}(\tilde{U}_i-\Pi_2)X_i'M_iX_i(\tilde{U}_i-\Pi_2)$ is distributed in the $m$-dimensional Wishart distribution with $A-\ell$ degrees of freedom and the true covariance matrix $T^{-1}C^*$, provided that $A-\ell-m \geq 0$. The integral of the product of $(\det T^{-1}\tilde{U}_iX_i'M_iX_i\tilde{U}_i)^{-1}$ and the noncentral Wishart density has a finite integral under which the product of $(\det T^{-1}\tilde{U}_iX_i'M_iX_i\tilde{U}_i)^{-1}$ and the noncentral Wishart density has a finite integral is $A-\ell-m \geq 2$. $ee'$ has a bounded expectation provided that $A-\ell-m \geq 2$.

The above reasoning may be extended to the $r$-th order moments. A sufficient condition for the existence of the $r$-th order moments of the Two Stage Least Squares estimates is that the degree of overidentification is no less than $r$.

4. Expected Value to the Order of $T^{-m}$

To turn to the approximation to the moments of the Two Stage Least Squares estimates we begin with a general explanation. Let $\{\xi(T)\}$ be a sequence of random variables. When $E(\xi(T))$ is expanded in the power series of $T^{-1}$ to be written as $\sum_{k=0}^{\infty} a_k T^{-k}$, the expected value of $\xi(T)$ to the order of $T^{-m}$ is defined as $\sum_{k=0}^{m} a_k T^{-k}$. (This is the definition which Srinivasan [22] also uses.) In the present paper a reference will be made to the expected value in the order of $T^{-m}$ to be defined as $\alpha_m T^{-m}$.

In comparison with this approach Nagar's method involves the notion of a random variable being in the order of $T^{-m/2}$ in probability where $m$ is a positive integer.

---

2) See Anderson and Girshick [1] and Anderson [2], especially [2, p. 419] together with the correction in Anderson [3]. The expression in Theorem 3 there is the density function of $a_{ij}$ under the condition that $N-1 \geq p_1$, i.e., that the degrees of freedom exceed the number of dimensions. The integral of this function over the domain of $a_{ij}$ such that $((a_{ij}))$ is positive definite must be unity. The condition that $N-1 \geq p_1$ is $A-\ell-m \geq 1$ for the integral mentioned in the text.

3) Start with the Hölder inequality for $r$ variables.
Along this line Sargan and Mikhail [18] showed for a single element of the Two Stage Least Squares estimate that Nagar's moment formulae can be deduced from a version of statistical differentials of estimation errors, where one can determine the order in probability of the remainder and thereby remedy the defects of Nagar's approach.

On the other hand the expected value (or moments) rather than a stochastic variable is what we expand in the present study on the basis of the Lomnicki-Zaremba version of the statistical differentials. When we are concerned with low-order moments of estimates rather than their sampling distributions, our procedure is simpler than Sargan and Mikhail's in regard to both the concepts and calculations.4)

5. Adapting the Lomnicki-Zaremba Version of the Method of Statistical Differentials

We adapt the theorem due to Lomnicki and Zaremba [11, p. 151] in order to enhance its usefulness for statistical investigations.

**Theorem 1 (A)**

(a) Let \( z(T) = (z_1(T), \ldots, z_q(T)) \) be a vector of nonstochastic variables, and suppose that \( \lim_{T \to \infty} z(T) \) exists.

(b) Let \( x(T) = (x_1(T), \ldots, x_p(T)) \) be a vector of stochastic variables that satisfy

1) \( E(x(T)) = 0 \), for each \( T \)

2) \( \lim_{T \to \infty} TE(x_i(T)x_j(T)) \) exists for \( i, j = 1, \ldots, p \)

and, letting \( \bar{r} \) be the maximum of positive integers \( r \) such that \( E(x_i(T)^{2r}) = O(T^{-r}) \),

3) \( i = 1, \ldots, p, \) (put \( \bar{r} = \infty \) if \( r \) is unbounded),

4) \( 2\bar{r} \geq 4 \).

(c) Let \( H(x, z) \) be a function such that its partial derivatives with respect to \( x \) up to the second order are continuous at \((0, z)\), and, moreover, such that for some \( u > 1 \)

7A) i) \( E(|H(x(T), z(T))|^{u}) \) exists for every \( T \),

ii) \( E(|H(x(T), z(T))|^{u}) = O(T^{1-u}) \).

Then we have

8) \( \lim_{T \to \infty} T[E(H(x(T), z(T))) - H(0, z(T))]/2 \)

9) \( \lim_{T \to \infty} T \sum_{i,j=1}^{p} \partial^2 H/\partial x_i \partial x_j (0, z(T)) E(x_i(T)x_j(T)) \),

where \( \partial^2 H/\partial x_i \partial x_j \) is the partial derivative evaluated at \( x = 0, z = z(T) \).

**Theorem 1 (B)**

(a) Let \( \{z(T)\} \) be as assumed in Theorem 1 (A).

(b) Let \( \{x(T)\} \) be a sequence of stochastic vectors that satisfy (1) and (2) in Theorem

4) Sargan and Mikhail are primarily concerned with the distribution. I would also like to join Sewell [21] and Srinivasan [22] in emphasizing that there is no guarantee, in general, for the limits between the finite sample moments (our approach) and the moments of the limiting distribution (Sargan and Mikhail's approach).

5) \( a(T) = o(T^p) \) if \( \lim_{T \to \infty} |a(T)| T^{-p} = 0 \), and \( a(T) = O(T^p) \) if \( \lim_{T \to \infty} |a(T)| T^{-p} < \infty \).
1 (A) and
\[ \lim_{T \to \infty} T^{-1} \lim_{T \to \infty} T E(x_i(T) x_j(T)) \] for \( i, j = 1, \ldots, p \)
(3B) \[ \lim_{T \to \infty} T^2 E(x_i(T) x_j(T)^2 x_k(T)) \] exists for \( i, j, k = 1, \ldots, p \)
(4B) \[ \lim_{T \to \infty} T^2 E(x_i(T)^2 x_j(T)^2 x_k(T)) \] exists for \( i, j, k = 1, \ldots, p \)
(5B) \[ \lim_{T \to \infty} T^2 E(x_i(T)^2 x_j(T)^2 x_k(T)^2 x_h(T)) \] exists for \( i, j, k, h = 1, \ldots, p \)
(6B) \[ 2F \geq 8. \]

(c) Suppose that the partial derivatives of \( H(x, z) \) with respect to \( x \) up to the fourth order are continuous at \((0, z)\) and also that for some \( u > 1 \)
(7B)
\begin{itemize}
  \item[(i)] the same as (i) in (7A)
  \item[(ii)] \[ E\{ H(x(T), z(T))^u \} = O (T^{u-1-2v}). \]
\end{itemize}

Then we have
\[ \lim_{T \to \infty} T^2 \left[ E\{ H(x(T), z(T)) - H(0, z(T)) \} \right] \]
\[ - \frac{1}{2} \lim_{T \to \infty} T \sum_{i, j = 1}^p \frac{\partial^2 H}{\partial x_i \partial x_j} (0, z(T)) E(x_i(T) x_j(T)) \]
\[ = \frac{1}{2} \lim_{T \to \infty} T \sum_{i, j = 1}^p \frac{\partial^3 H}{\partial x_i \partial x_j} (0, z(T)) E(x_i(T) x_j(T) x_k(T)) \]
\[ + \frac{1}{3!} \lim_{T \to \infty} T \sum_{i, j, k = 1}^p \frac{\partial^4 H}{\partial x_i \partial x_j \partial x_k} (0, z(T)) E(x_i(T) x_j(T) x_k(T) x_h(T)). \]

Remark When divided by \( T \), (8) gives \( E\{ H(x(T), z(T)) - H(0, z(T)) \} \) in the order of \( T^{-1} \), and, when divided by \( T^2 \), (9) gives \( E\{ H(x(T), z(T)) - H(0, z(T)) \} \) in the order of \( T^{-2} \).

Proofs of Theorems 1 (A) and 1 (B) The proofs of Theorems 1 (A) and 1 (B) are obtained through an adaptation of the Lomnicki-Zaremba’s proof [11, pp. 151–154]. In the following presentation the portions that are identical to the Lomnicki-Zaremba’s are condensed to the minimum explanation. The proof of Theorem 1 (B) is so analogous to that of Theorem 1 (A) that only the latter will be shown.

Let \( \varepsilon > 0 \) be an arbitrary number, and put
\[ \eta \equiv 2\varepsilon \left[ 1 + \left( \sum_{i=1}^p c_i \right)^2 \right] \]
where \( c_i \equiv \lim_{T \to \infty} T E(x_i(T) x_i(T)) \). There exists positive constants \( \delta x(i), i = 1, \ldots, p \), and \( \delta z(j), j = 1, \ldots, q \), such that, whenever \( |x_i(T)| < \delta x(i) \) and \( |z_j(T) - z_j| < \delta z(j) \), all of the second order partial derivatives of \( H(\cdot) \) at \((x(T), z(T))\) differ from their values at \((0, z)\) by less than \( \eta \). Since \( |z(T) - z_j| < \delta z(j) \) for sufficiently large \( T \), and since we shall consider always \( \lim_{T \to \infty} T \) or \( \lim_{T \to \infty} T^2 \), we may concentrate upon \( |x_i(T)| < \delta x(i) \). Put \( \min \delta x(i) = \delta \).

From the Markov inequality and the definition of \( \delta \) it follows that for sufficiently

---

(6) The Lomnicki-Zaremba’s original theorem has the existence of \( \lim_{T \to \infty} T E(x(T)) \) in place of (1), and some alterations in (2) and (5B) due to this difference. Further it does not have \( z(T) \), and, does not state (9) explicitly. Most important of all, the function \( H \) is assumed to be bounded in the ordinary sense in the Lomnicki-Zaremba’s theorem, which considerably reduces the usefulness of the theorem since we rarely deal with the bounded function in statistics.
large $T$, which word will be omitted henceforth,

\[(11) \; \Pr\{ |x_i^{(T)}| > \delta \} \leq \delta^{-1} O(T^{-r}). \]

Let $\mathcal{E}$ be the event that $|x_i^{(T)}| \leq \delta$ for every $i = 1, \ldots, p$, and let $\overline{\mathcal{E}}$ be its complement. From (11) it follows that

\[(12) \; \Pr(\overline{\mathcal{E}}) \leq p \delta^{-1} O(T^{-r}). \]

We investigate the right-hand side of

\[(13) \; T E\{H(x^{(T)}, z^{(T)})\} = \int_{\mathcal{E}} T H(x^{(T)}, z^{(T)})dP(x^{(T)}) + \int_{\overline{\mathcal{E}}} T H(x^{(T)}, z^{(T)})dP(x^{(T)}). \]

Begin with the second term. For any $u > 1$

\[(14) \; \lim_{T \to \infty} T \int_{\mathcal{E}} H^* dP \leq \lim_{T \to \infty} T \int_{\mathcal{E}} |H|^* dP \leq \lim_{T \to \infty} \left[ \int_{\mathcal{E}} |H|^* dP \right]^{1/u} \cdot T \left[ \int_{\mathcal{E}} dP \right]^{-1/\alpha} \]

\[\leq \lim_{T \to \infty} \left[ \int_{\mathcal{E}} |H|^* dP \right]^{1/\alpha} \cdot \left( T \left[ \int_{\mathcal{E}} dP \right]^{-1/\alpha} \right) \]

due to the Hölder inequality. The term inside $\{ \}$ in the last expression of (14) is $O(T^{1-\gamma(1-1/\alpha)}).$ Therefore if $\left[ \int_{\mathcal{E}} |H|^* dP \right]^{1/\alpha} = o(T^{(1-1/\alpha)-1/\alpha})$, that is, $\int_{\mathcal{E}} |H|^* dP = o(T^{(1-1/\alpha)-1/\alpha})$ which is ii) of (7A), then the last term in the sequence of inequalities in (14) is zero. Thus

\[(15) \; \lim_{T \to \infty} T \int_{\mathcal{E}} H^* dP = 0. \]

The first term on the right-hand side of (13) is treated just like in Lomnicki-Zaremba's proof.

\[(16) \; T \int_{\mathcal{E}} H(x^{(T)}, z^{(T)})dP = TH(0, z^{(T)}) + T \sum_{i=1}^{p} \frac{\partial H}{\partial x_i}(0, z^{(T)})E(x_i^{(T)})
\]

\[+ \frac{1}{2} \sum_{i,j} \left\{ \frac{\partial^2 H}{\partial x_i \partial x_j}(0, z^{(T)})E(x_i^{(T)}x_j^{(T)}) - H(0, z^{(T)}) \right\} dP
\]

\[- \sum_{i} \frac{\partial H}{\partial x_i}(0, z^{(T)})T \int_{\mathcal{E}} x_i^{(T)} dP
\]

\[- \frac{1}{2} \sum_{i,j} \frac{\partial H}{\partial x_i \partial x_j}(0, z^{(T)})T \int_{\mathcal{E}} x_i^{(T)}x_j^{(T)} dP
\]

\[+ \frac{1}{2} T \sum_{i,j} \left\{ \frac{\partial^2 H}{\partial x_i \partial x_j}(0, z^{(T)}) - \frac{\partial^2 H}{\partial x_i \partial x_j}(0, z^{(T)}) \right\} x_i^{(T)}x_j^{(T)} dP. \]

where $0 < \theta < 1$. The second term on the right-hand side of (16) vanishes due to (1). Moreover

\[(17) \; \lim_{T \to \infty} T \int_{\mathcal{E}} dP = 0, \]

\[(18) \; \lim_{T \to \infty} T \int_{\mathcal{E}} x_i^{(T)} dP = 0, \]

\[(19) \; \lim_{T \to \infty} T \int_{\mathcal{E}} x_i^{(T)}x_j^{(T)} dP = 0, \]

\[(20) \; \lim_{T \to \infty} T \int_{\mathcal{E}} x_i^{(T)}x_j^{(T)} dP = \eta, \]

where (6A) and (10) have been used. From (15), (17), (18), (19), and (20) we have (8) since $\varepsilon$ is arbitrary.

6. Remarks on the Econometric Applications

I feel that not only the assumptions (2) and (4B) hold but also $\tilde{r} = \infty$ in many econometric problems. If $\tilde{r} = \infty$, then ii) of each of (7A) and (7B) is turned into a seem-
ingly trivial statement that the left-hand side of each of these equations is \( O(T^s) \) for some \( s \) that can be positive and as large as we wish. The assumption still excludes, for example, the case where the left-hand side is \( O(\exp(T)) \).

Concerning \( u \) in \((7A)\) and \((7B)\) the best strategy is to try \( 1+\varepsilon \) where \( \varepsilon > 0 \) is as small as we wish. This is important in the case where, the larger the power \( u \), the more restricted is the existence of the expectation on the left-hand side.

7. The Nagar's Moment Formulae for the Two Stage Least Squares Estimates

We continue to use the model and the notations in section 2. The error of the Two Stage Least Squares estimates is

\[
(21) \quad e = B^{-1}a,
\]

where

\[
(22) \quad B = T^{-1} [Q^{-1} + Z'V + V'Z' + V'M^*V],
\]

\[
a = T^{-1} [Z'V + V'M^*V].
\]

We now attempt to derive the Nagar's moment formulae of \( e \) to the order of \( T^{-1} \).

To apply Theorem 1 (A) and 1 (B) we use the notations

\[
(23) \quad B_e = T^{-1} \{Q^{-1} + EV'M^*V\},
\]

\[
a_e = T^{-1} \{EV'M^*a\},
\]

\[
\tilde{B} = T^{-1} \{Z'V + V'Z' + V'M^*V - E(V'M^*V)\},
\]

\[
\tilde{a} = T^{-1} \{Z'V + V'M^*a - E(V'M^*a)\}.
\]

Then

\[
B = B_e + \tilde{B}, \quad a = a_e + \tilde{a},
\]

\[
E(\tilde{B}) = 0, \quad E(\tilde{a}) = 0,
\]

and \( B \) and \( a \) are nonstochastic.

The elements of \( \tilde{B} \) and \( \tilde{a} \) correspond to the elements of \( x^{(T)} \), and the elements of \( B_e \) and \( a_e \) to the elements of \( \text{V}^{(T)} \) in the theorems. The function \( H(*) \) is \((21)\) for the first order moments, and \( ee' \) for the second order moments.

We must confirm that the assumptions in Theorem 1 (B) hold. We clearly need

**Assumption VI** \[ \lim_{T \to \infty} T^{-1}Z'Z \text{ exists, and, it is nonsingular.} \]

This is sufficient to assure the existence of the limits of \( B \), \( a \), since \( \text{tr}M^* = \Lambda \) that is independent of \( T \).

Both \( e \) and \( ee' \) have the partial derivatives with respect to \( \tilde{B} \) and \( \tilde{a} \) up to the fourth order. The derivatives are continuous at \( \tilde{B} = 0, \tilde{a} = 0 \) by virtue of Assumption VI.

The conditions (2) through (5B) hold for the elements of \( \tilde{B} \) and \( \tilde{a} \). The following theorem is useful to establish this.

**Theorem 2** Suppose that \( x_1, \ldots, x_n \) are distributed in a \( m \)-dimensional normal distribution with the zero mean vector, and left \( \xi_1, \ldots, \xi_p \) be any \( p \) of these \( x \)'s, allowing a repetition such as \((x_1, x_1, x_2)\). If \( p \) is odd, \( E(\tilde{U}^{(p)}\xi_1) = 0 \), and if \( p \) is even, \( E(\tilde{U}^{(p)}\xi_1) = \Sigma II E(\xi_1, \xi_1) \) where \( \Sigma \) is taken over all possible ways (altogether \((p-1)(p-3)\)
...3.1) of dividing $p$ integers $(1, \ldots, p)$ into $\frac{1}{2}p$ distinct sets of 2 distinct integers and $\Pi$ is taken over such $\frac{1}{2}p$ sets.\textsuperscript{7)

For the case of $p=4$ $E(\xi_1, \xi_2, \xi_3, \xi_4)=E(\xi_1, \xi_3)E(\xi_2, \xi_4)+E(\xi_1, \xi_4)E(\xi_2, \xi_3)+E(\xi_1, \xi_2)E(\xi_3, \xi_4)$. Also crucial is the fact that the rank of $M^*$ is independent of $T$. $\tilde{r}$ is indeed $\infty$.

Turning to (7B) we assume

**Assumption VII**

(i) For the first order moment $A-l-m \geq 1$, and (ii) for the second order moment $A-l-m \geq 2$.

This assumption is sufficient for (i) of (7B) because the reasoning in (ii) of section 3 can be strengthened when nonintegers are allowed to enter. There, a weaker sufficient condition\textsuperscript{8)} for the existence of finite integral of the product of $(\det T^{-1} \hat{H}'X_2'M_1X_2\hat{H})^{-\frac{1}{2}}$ and the noncentral Wishart density is $\frac{1}{2}(A-l-m-1)\geq 0$, which we turned into $A-l-m \geq 1$ in section 3 because $A-l-m$ can only be an integer. As for (7B) the power $u$ need not be an integer. A sufficient condition for the existence of a finite integral of the product of $(\det T^{-1} \hat{H}'X_2'M_1X_2\hat{H})^{-u/2}$ and the noncentral Wishart density is $\frac{1}{2}(A-l-m-1)-\frac{u}{2} > 0$, i.e., $A-l-m > u-1$. Further a similar condition for $(\det T^{-1} \hat{H}'X_2'M_1X_2\hat{H})^{-u}$ is $A-l-m > 2u-1$. For $u$ such that $\frac{3}{2} > u > 1$, i) of (7B) holds in the cases of both the first and the second order moments.

For (ii) of (7B) we introduce

**Assumption VIII**

$\lim_{T \rightarrow \infty} T^{-1} \hat{H}'X_2'M_1X_2\hat{H}$ exists.

Since $\tilde{r} = \infty$, even an extremely inefficient bound to the left-hand side of (7B) would

\textsuperscript{7)} For a proof, differentiate successively the characteristic function of $(x_1, \ldots, x_m)$.

\textsuperscript{8)} In connection with the expression in Theorem 3 of Anderson [2] that was brought up in the previous footnote 1), it can be shown that the integral of the function $W$ over the entire domain of $A^{def}((a_{ij}))$ such that $A$ is positive definite is finite as far as $\frac{1}{2}(N-p-2) > -1$. This statement is nothing but a well known property of the Gamma function if $W$ is the density function of the central $x^2$ distribution. The proof of the statement for the case of the noncentral Wishart distribution is as follow.

Without losing generality one may assume that $E$ is the identity matrix. Then $-\frac{1}{2}\sum_{i,j} a_{ij}$ in Anderson's expression becomes $-\frac{1}{2}\sum_{i=1}^{N} a_{ii}$, where $a_{ii} \geq 0$ is the eigenvalue of $A$. Further $|A| = \prod_{i=1}^{N} a_{ii}$.

Turning to the complicated integral with respect to $z_1, \ldots, z_t$, $\prod_{i}^{t} z_i$ are the eigenvalues of $TA$, all of which are indeed real and non negative. (Here $T$ is the Anderson's, that is, the means sigma matrix.) The largest of them is equal to the largest eigenvalue of $T^t AT^t$. The latter cannot exceed the product of the largest eigenvalues of $T$ and of $A$ by virtue of a theorem on the norm of matrices defined as the square root of the largest eigenvalue of $X^t X$ for a matrix $X$. In view of $\sum_{i=1}^{t} w_i$ we see that $\exp \left(\sum_{i=1}^{t} w_i \right)$ is bounded by $\exp Ka_{1},$ where $a_{1}$ is the largest of the $a_{i}$'s, and $K$ is $t$ times the square root of the largest eigenvalue of $T$. Except for some unimportant constants the function $W$ is bounded by the product of $a_{1}^{-\frac{N-p-2}{2}} \cdot \exp \left(\frac{1}{2}(N-p-2)\right)$ and $\prod_{i=1}^{t} \text{det} a_{i}^{-\frac{N-p-2}{2}} \cdot \exp \left(-\frac{1}{2}K^2\right)$. Transform $a_{1}$ to $\beta = (a_{1}^{-\frac{1}{2}} - K^2)$ so that the first factor times the Jacobian is (the polynomial of $\beta^t$ in the order $N-p-2$) times $\exp \left(-\frac{1}{2}\beta^2\right) \exp \left(-\frac{1}{2}K^2\right)$. The condition for the convergence of the Gamma function establishes the desired result.
be adequate. In fact it is easy to show that the left-hand side is $O(T^{m/2(n+1)-1})$ as far as $\frac{3}{2} > u > 1$. This completes the assurance that the assumptions in Theorem 1 (A) and 1 (B) hold in our problem.

Notice that in order to get the expected value of $H(x^{(T)}, z^{(T)})$ through the theorems in section 5 the terms in the order of $T^{-1}$ or $T^{-2}$ of $H(0, z^{(T)})$ should be added to (8) divided by $T$ or (9) divided by $T^2$. In the case of the first order moment of the Two Stage Least Squares estimates $B_{-1}a$, is $H(0, z^{(T)})$, and, it has a non zero term in the order of $T^{-1}$. In the case of the second order moments $B_{-1}a, a, B_{-1}c$ is $H(0, z^{(T)})$, and, it has a non zero term in the order of $T^{-2}$.

I have found that our method yields precisely what Nagar [15] gives as the moment formulae in the order of $T^{-1}$ and $T^{-2}$ for the first and the second order moments. I have also examined the bias, to the order of $T^{-1}$, of the estimate of $\sigma^2$ that is given in Nagar [16], only to find that it is identical to the result derived from our theorem.

A crucial step in the Nagar's procedure is the expansion of $(Q^{-1}+\mathcal{J})^{-1}$ which is discussed in Nagar [15, p. 582, footnote 8]. This part of the Nagar's procedure may be interpreted as a version of our theorem through relating this Nagar's expansion to the partial derivatives of the functions whose expectations are to be evaluated.

**Corollary** Suppose that $C^{(T)}$ and $D^{(T)}$ are $p \times p$ symmetric matrices with elements $C_{ij}^{(T)}$ and $D_{ij}^{(T)}$ and that $a^{(T)}$ is a $p \times 1$ vector with elements $a_i^{(T)}$. We assume

\begin{align}
(24) & \quad C^{(T)} is nonstochastic and positive definite; \lim_{T \to \infty} C^{(T)} exists, and it is positive definite as well,

(25) & \quad D^{(T)} is stochastic, and \lim_{T \to \infty} TE(D^{(T)}) exists,
\end{align}

9) Again we consider $(\det T^{-1/2}X'X^*M_1X^*X_2^*)^{-u/2}$ times the noncentral Wishart density function for $T^{-1/2}X'X^*M_1X^*X_2^*$. (Here we use $T$ for the length of data.) In the previous footnote 8) the power of $|A|$ was not restricted to $\frac{1}{2}$ times an integer. Using the previous footnote as the basis we investigate to see how the integral of the bound to $W$ (not the bound itself) is affected as $T$ varies. The degrees of freedom and the dimension are unaffected by $T$. Since the true covariance matrix has the factor $T^{-1}$, we assume it is $T^{-1}$ times the identity. The largest eigenvalue of the means sigma matrix goes to a limit as $T \to \infty$ if Assumption VIII holds. We then see that the dependence of $K$ upon $T$ is such that $\lim_{T \to \infty} T^{-1}K$ exists. Then $\exp\left(-\frac{K^2}{2}\right) \to 0$. The integrals (i.e., the Gamma functions) with respect to $\beta, \alpha_2, \ldots, \alpha_p$ have the factor $T^{-1(N-p-2)}$ and hence also go to zero.

We now have to investigate the constants which were not considered in the previous footnote. $|\sigma_{ij}|^{1/2}$ in Anderson's notation is $O(T^{m/2(n+1)-1})$ in our problem. $\exp\left(-\frac{1}{2}\sum x_i^2\right)$ in Anderson's notation goes to the unity as $T \to \infty$, because the dependence of $\varepsilon^2$ upon $T$ is such that $\lim_{T \to \infty} T\varepsilon^2$ exists.

10) Our investigation also shows that the moments converge to those which are given as the asymptotic formulae in the literature, that is, zero for the first order moment and $\sigma^2 Q$ for the second order moments, in the case of Two Stage Least Squares estimates. Sewell [21, p. 42] makes some remark that is contradictory to my result. I think he has overlooked Basmann's assumption [5, p. 624, equation (2.12)].

11) To be precise, Nagar's $Q$ in the moment formulae should be read as $T^{-1} \lim_{T \to \infty} (T^{-1}Z'Z)^{-1}$.

12) Professor Takamitsu Sawa suggested to me consideration of this topic.
(26) $a(T)$ is stochastic, and $\lim_{T \to \infty} TE(a(T))$ exists,

(27) $D_i(T) - E(D_i(T))$, $i, j = 1, \ldots, p$, and $a_i(T) - E(a_i(T))$, $i = 1, \ldots, p$, satisfy the conditions (2) through (6B) for $x(T)$ in Theorem 1 (A) and 1 (B),

(28) when $B(T) \triangleq C(T) + D(T)$, both $B(T)^{-1}a(T)$ and $B(T)^{-1}a(T)a(T)'B(T)^{-1}$ satisfy the condition (7B) for $H$ in Theorem 1 (B).

Then, dropping the superscript $(T)$ from $B(T), C(T), D(T),$ and $a(T)$ to simplify the notations, we have

(29) $\lim_{T \to \infty} TE(B^{-1}a) = \lim_{T \to \infty} TE((I-C^{-1}D)C^{-1}a)$,

(30) $\lim_{T \to \infty} TE(B^{-1}aa'B^{-1}) = \lim_{T \to \infty} TE((I-C^{-1}D)C^{-1}aa'C^{-1}(I-DC^{-1}))$,

moreover when the limit in (30) is written as $G$

(31) $\lim_{T \to \infty} T^2\{E(B^{-1}aa'B^{-1}) - T^{-1}G\} = \lim_{T \to \infty} T^2[E((I-C^{-1}D)C^{-1}aa'C^{-1}(I-DC^{-1}) + DC^{-1}DC^{-1}) - T^{-1}G]$

8. Kadane's $\sigma$-expansion

Just like Nagar's formulae, Kadane's $\sigma$-expansion [7] is also subjected to Srivivasan's criticism. It can also be saved from the criticism by our theorem in the case of the Two Stage Least Squares estimate, if the degree of overidentification is as specified in Assumption VII.

Using $i = 1, 2, \ldots$, rewrite (22) and (23) as

where each element of $u$ is $N(0, 1)$. Hold $T$ fixed. Let $\sigma(i)^{-1} i^{-1}$, and treat $i$ as if it were $T$ in our theorems. $B(i)$ and $a(i)$ correspond to $x(T)$, and $B(i)_c$ and $a(i)_c$ to $z(T)$. The assumptions (1) through (6B) hold. The part i) of (7B) has been already established. The part ii) also holds. Theorems 1 (A) and 1 (B) give $\lim_{i \to \infty} E(B(i)^{-1}a(i))$ and $\lim_{i \to \infty} E(B(i)^{-1}a(i)a(i)'B(i)^{-1})$, which, in turn, yields the small $\sigma$-expansion even when the $\sigma$ such that $\sigma^{-1} < \sigma < \sigma^{-1}$ is allowed for.

Looking at the partial derivatives involved in the expectations on the right-hand sides of (29), (30) and (31) prior to applying $\lim_{i \to \infty}$, one sees that $\sigma(i)^{2i}$ and $T$ appear in a pair, $\sigma(i)^{2i}/T$, and, that the remaining, not involving $\sigma(i)$, converges to limits as $T \to \infty$. No wonder why Kadane's formulae are so similar to Nagar's.

9. The Goldberger-Nagar-Odeh Formula for the Forecast Variances

Though the method of statistical differentials lies behind the Goldberger-Nagar-Odeh formula for the forecast Variances [6]—see especially their Lemma in p. 558—, it is very difficult to justify it by our theorems. Let $y_i'T + x_i'B = u_i'$, their
equation (2.1), be the system of structural equations. For practically any specification of the probability structure of \( u' \) and for practically any reasonable estimate \( \hat{\Gamma}, \hat{B} \), it is very unlikely that \( \hat{B}\hat{\Gamma}^{-1} \) has moments of any orders whatsoever. To be more precise, if \( \text{det} \hat{\Gamma} \) is distributed in a continuous probability density function that takes a positive value at \( \text{det} \hat{\Gamma} = 0 \), then \( E[|\text{det} \hat{\Gamma}|^{-1}] \) does not exist. In addition, if the expectation of the product of \( \hat{B} \) and the adjoint matrix of \( \hat{\Gamma} \), conditional upon \( \text{det} \hat{\Gamma} \), is continuous with respect to \( \text{det} \hat{\Gamma} \) with at least one element being non zero at \( \text{det} \hat{\Gamma} = 0 \), then the absolute moments of \( \hat{B}\hat{\Gamma}^{-1} \) do not exist. The condition (7A) of our Theorem 1 (A) does not hold. In fact the difficulty is not in our theorem, but in the fact that moments are used to summarize the distribution of \( \hat{B}\hat{\Gamma}^{-1} \) in spite of that \( \hat{\Gamma} \) is not definite in sign.

I think that the Goldberger-Nagar-Odeh formula should be justified through the limiting distribution of a nonlinear function of random variables that are distributed asymptotically in a normal distribution.

10. Conclusion

(1) A method has been developed to evaluate to the order of \( T^{-2} \) the moments of estimators of a wide category. The method is hoped to provide a useful approximation in the cases where the exact moments are hard to obtain and the size of available data is inadequate to rely upon the moments of the limiting distribution even if it is valid for sufficiently large \( T \). (2) The application of our method has confirmed the validity of the Nagar’s and Kadane’s moments formulae for the Two Stage Least Squares estimates under some conditions, among others, on the degrees of overidentification. (3) A half of the Basmann’s conjecture on the moments of the Two Stage Least Squares estimates has been proved for the model that is more general than those hitherto dealt with in the literature. (2) and (3) above will be extended to the \( k \)-class estimators in the Appendix.) (4) The Goldberger-Nagar-Odeh formula of the forecast variances cannot be justified by our theorem.

On the other hand I have left out the following important problems. (1) The usefulness of our method should be established in the framework of dynamic econometric models in which lagged endogenous variables are involved. (2) Nothing along the present study has been done on the limited information maximum likelihood estimators. (3) The unproved part of the Basmann’s conjecture is that when \( r \) is the degree of overidentification, the moments of the odd orders higher than the \( r \)-th do not exist in the case of the Two Stage Least Squares estimates.

(Osaka University)

Appendix: The Moments of the \( k \)-Class Estimators for \( k<1 \)

The treatment in section 3 can be extended to the \( k \)-class estimators for a non-stochastic \( k \). However only the case, \( k<1 \), will be considered. On the case, \( k>1 \), Sawa [20] proved that the moments of any order do not exist for the equation that contains two endogenous variables.

— 12 —
Define \( \alpha(\geq 0) \) through \( k=1-\alpha^2 \). Put \( \hat{Z} \overset{\text{def}}{=} T^{-i}(\hat{Y}+\alpha\hat{V}, X) \), where \( \hat{Y}=X\hat{U} \) and \( \hat{V}=Y-\hat{Y} \). Perform the least squares calculation with \( y \) (the left-hand side endogenous variable) as the dependent variable and \( X \) as the independent variable. Let it be \( y=\hat{y}+\epsilon, \hat{y}=X\hat{\epsilon} \). Define \( \hat{Z}_{\alpha}^{(i)} \) as the matrix that is obtained from \( \hat{Z}_{\alpha} \) through replacing its \( i \)-th column by \( T^{-1/2}(Y+\alpha\epsilon, X) \). The \( k \)-class estimator for \( k \leq 1 \) may be denoted as \( \det \hat{Z}_{\alpha}^* \hat{Z}_{\alpha}^{(i)}(\det \hat{Z}_{\alpha}^{(i)}\hat{Z}_{\alpha}^{(i)})^{-1} \).

The reasoning in section 3 may now be applied. For the first order moment we deal with \( \det \hat{Z}_{\alpha}^* \hat{Z}_{\alpha}^{(i)} \). Since \( \hat{Z}_{\alpha}^* \hat{Z}_{\alpha}^{(i)} = \det \tilde{T}^{-1}X'X_1 \cdot \det \tilde{T}^{-1}X_2', X_2, X_2, X_2, + \alpha \tilde{V}' \tilde{V} \), let us put \( P=\tilde{T}^{-1}X_2', X_2, X_2, \tilde{U}_2 \) and \( Q=\tilde{T}^{-1}X_1' \tilde{V} \), and consider \( \det [P+\alpha^2 Q]^{-1} \) only. The distribution of \( P \) has been stated in section 3. Let \( \Sigma \) be its means sigma matrix, \( \tilde{T}^{-1}X_2'X_2X_2X_2 \). \( P \) is distributed in the \( m \)-dimensional central Wishart distribution with \( T-A \) degrees of freedom and the true covariance matrix, \( \tilde{T}^{-1}C^* \). Moreover \( P \) and \( Q \) are independently distributed.

Henceforth assume \( \alpha \geq 0 \), which means \( k<1 \). The transformation of the variables,\(^\dagger\) \( A_0=P+\alpha^2 Q \), \( B_0=P+\alpha^2 Q \) has the Jacobian, \( \alpha^{m(m+1)}(\det B_0)^{\alpha^{m(m+1)}} \) because the transformation from \( (P, Q) \) to \( (P, B_0) \) has the Jacobian \( \alpha^{m(m+1)} \) and the transformation from \( (P, B_0) \) to \( (A_0, B_0) \) has the Jacobian, \( (\det B_0)^{\alpha^{m(m+1)}} \).

The essential part of the integral for the expectation of \( \det [P+\alpha^2 Q]^{-1} \) is

\[
(A1) \quad \left[ (\det B_0)^{\alpha^{m(m+1)}}(\det I-A_0)^{\alpha^{m(m+1)}}(\det B_0)^{\alpha^{m(m+1)}} \right] \cdot \exp \left[ -\frac{1}{2}(2-\alpha^2)T \operatorname{tr} C^{+1}B_0 \right] \cdot (\text{some integral involving the roots of} \quad \left| \Sigma -AT^{-1}C^*B_0^{-1}A_0^{-1}B_0^{-1}C^* \right|^2 = 0) \cdot \exp \left[ \frac{1}{2}(1-\alpha^2)T \operatorname{tr} C^{+1}B_0 A_0 B_0 \right] dA_0 dB_0,
\]

where the integral with respect to \( A_0 \) is in the domain such that \( I-A_0 \) is positive definite.

It can be shown that this integral is bounded provided that all of the following three conditions hold,

i) \( A-I-m \geq 0 \), ii) \( T-A-m \geq 0 \), iii) \( T-I-m \geq 1 \).

For the second order moments the first two conditions remain the same as above, but the third condition should be \( T-I-m \geq 2 \).

It seems that a negative value for \( k \) does not give rise to any troubles in this respect. In the particular case of \( \alpha^2=1 \), i.e., \( k=0 \), \( P+Q \) is distributed in the \( m \)-dimensional noncentral Wishart distribution with \( (A-I-m)+(T-A) = T-I-m \) degrees of freedom and the means sigma matrix \( \Sigma \). \( (A1) \) may be replaced by

\(\dagger\) This is the multivariate generalization of the process to derive a Beta distribution from two independent Gamma distributions. See Kshirsagar [10] for example.

\(\ddagger\) The last factor of the integrand in \( (A1) \) arises from the loss of homoskedasticity between \( P \) and \( \alpha^2 Q \). While integrating with respect to \( A_0, A_0 \) \( \textrm{tr} C^{+1}B_0 A_0 B_0 \) in this factor may be replaced by its upper bound \( \textrm{tr} C^{+1}B_0 \) and combined with the factor, \( \exp \left( -\frac{1}{2}(2-\alpha^2)T \operatorname{tr} C^{+1}B_0 \right) \). A similar treatment may be applied to "some integral."
Thus the condition for the boundedness of the integral is \( T - l - m \geq 1 \) for the first order moment, and \( T - l - m \geq 2 \) for the second order moment.

The reasoning in section 7 is also applicable. We conclude that Nagar’s moment formulae for \( k \neq 1 \) can be justified under the following assumption, in addition to Assumptions V, VI, and VIII.

**Assumption VII’**
(a) \( k = 1 + \frac{\kappa}{T} \), where \( \kappa < 0 \); (b) \( A - l - m \geq 0 \), and \( T - l - m \geq \text{Max}(\nu, A - l) \) where \( \nu \) is the order of moments that one seeks.

**REFERENCES**


August 1973  M. Hatanaka: On the Existence and the Approximation Formulae


