Econometric Theory, **20**, 2004, 805–810. Printed in the United States of America.
DOI: 10.1017/S026646660420408X

PROBLEMS AND SOLUTIONS

SOLUTIONS

03.3.1. Normal's Deconvolution and the Independence of Sample Mean and Variance—Solution[1]

Karim M. Abadir
University of York, UK

Jan R. Magnus
Tilburg University, The Netherlands

(the posers of the problem)

(a) The "if" part is easy to prove. If $x_i \sim \mathrm{N}(\mu_i, \sigma_i^2)$ for $i = 1,2$, then their independence simplifies the moment generating function (m.g.f.) of y to

$$m_y(t) \equiv \mathrm{E}(\mathrm{e}^{t(x_1+x_2)}) = \mathrm{E}(\mathrm{e}^{tx_1})\mathrm{E}(\mathrm{e}^{tx_2}) = \mathrm{e}^{(\mu_1+\mu_2)t+(\sigma_1^2+\sigma_2^2)t^2/2},$$

so that $y \sim \mathrm{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

The "only if" part is less obvious. We will assume that $y \sim \mathrm{N}(0,1)$, without loss of generality (the usual extension to $\mu + \sigma y$ applies). Then, the characteristic function (c.f.) of y is

$$\mathrm{e}^{-t^2/2} = \mathrm{E}(\mathrm{e}^{\mathrm{i}ty}) = \mathrm{E}(\mathrm{e}^{\mathrm{i}t(x_1+x_2)}) = \mathrm{E}(\mathrm{e}^{\mathrm{i}tx_1})\mathrm{E}(\mathrm{e}^{\mathrm{i}tx_2}) \equiv m_1(\mathrm{i}t)m_2(\mathrm{i}t),$$

by independence of x_1 from x_2. Note that, for t real-valued,

$$|\mathrm{E}(\mathrm{e}^{\mathrm{i}tx_2})| \le \mathrm{E}(|\mathrm{e}^{\mathrm{i}tx_2}|) = \mathrm{E}(1) = 1,$$

so we have $\mathrm{e}^{-t^2/2} \le |m_1(\mathrm{i}t)|$ or equivalently $-2\log|m_1(\mathrm{i}t)| \le t^2$. Because the m.g.f. of x_1 exists, all the derivatives of $m_1(\mathrm{i}t)$ are finite at $t = 0$ and $\log|m_1(\mathrm{i}t)|$ has a Taylor-series representation as a polynomial in t. From the previous inequality, the maximal power of this polynomial is 2. As a result, $m_1(t) = \exp(\alpha_1 t + \alpha_2 t^2)$ for suitably chosen constants α_1 and α_2 (recall that $m_1(0)$ is set to 1, by definition). This establishes normality for x_1 and, by symmetry of the argument, for x_2 too.

Cramér's (1936) deconvolution theorem is actually more general than is stated in part (a), because it does not presume the existence of m.g.f.s for x_1 and x_2, at the cost of a further complication of the proof. In our proof, we have used (without needing to resort to the language of complex analysis) the fact that the existence of the m.g.f. implies that it is analytic (satisfies the Cauchy–Riemann equations) and is thus differentiable infinitely many times in an open neighborhood of $t = 0$ in the complex plane. On the other hand, if one did not assume the existence of m.g.f.s, then one would require some theorem from

 0266-4666/04 $12.00

complex function theory. One such requisite would be the "principle of isolated zeros" or "uniqueness theorem for analytic functions." Another alternative requisite would be "Hadamard's factorization theorem," used in Loève (1977, p. 284).

(b) For $n < \infty$, Cramér's deconvolution theorem (see part (a)) can be used $n-1$ times to tell us that $\bar{x} \sim \mathrm{N}(\mu, \sigma^2/n)$ decomposes into the sum of n independent normals, so that $\operatorname{var}(\boldsymbol{x}) = \boldsymbol{\Sigma}$ is a diagonal matrix satisfying $\operatorname{tr}(\boldsymbol{\Sigma}) = n\sigma^2$. However, the theorem does not imply that the components of the decomposition have identical variances and means, and we need to derive these two results, respectively.

Define the idempotent matrix $\boldsymbol{A} = (a_{ij}) := \boldsymbol{I}_n - \boldsymbol{\iota\iota}'/n$. Then, because $\boldsymbol{x}'(\sigma^{-2}\boldsymbol{A})\boldsymbol{x} \sim \chi^2(n-1)$, we have $\boldsymbol{A} = \sigma^{-2}\boldsymbol{A\Sigma A}$. The fact that $\boldsymbol{A}$ is idempotent implies that $\boldsymbol{ADA} = \mathbf{O}$, where

$$\boldsymbol{D} = \operatorname{diag}(d_1, \ldots, d_n) := \boldsymbol{I}_n - \sigma^{-2}\boldsymbol{\Sigma}$$

with

$$\operatorname{tr}(\boldsymbol{D}) = n - \sigma^{-2}\operatorname{tr}(\boldsymbol{\Sigma}) = n - \sigma^{-2}n\sigma^2 = 0. \tag{1}$$

The diagonal elements of $\boldsymbol{ADA}$ are given by

$$(\boldsymbol{ADA})_{jj} = d_j - \frac{2}{n}d_j + \frac{1}{n^2}\operatorname{tr}(\boldsymbol{D}) = \left(1 - \frac{2}{n}\right)d_j. \tag{2}$$

For $n \geq 3$, the equation $\boldsymbol{ADA} = \mathbf{O}$ thus gives $d_j = 0$ for $j = 1, \ldots, n$, and hence $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{I}_n$.

To obtain the mean, we note that the noncentrality parameter of $\boldsymbol{x}'(\sigma^{-2}\boldsymbol{A})\boldsymbol{x}$ is given by $\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1/2}(\sigma^{-2}\boldsymbol{A})\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu}$. Because our quadratic form has a central χ^2-distribution and $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{I}_n$, we obtain $\boldsymbol{A\mu} = \mathbf{0}$ and hence

$$\boldsymbol{\mu} = \boldsymbol{\iota}\frac{\boldsymbol{\iota}'\boldsymbol{\mu}}{n}.$$

Then, $\mathrm{E}(\boldsymbol{x}) = \mu\boldsymbol{\iota}$ follows by

$$\mu = \mathrm{E}(\bar{x}) = \mathrm{E}\left(\frac{\boldsymbol{\iota}'\boldsymbol{x}}{n}\right) = \frac{\boldsymbol{\iota}'\boldsymbol{\mu}}{n}.$$

(c) When $n = 2$,

$$\boldsymbol{A} = \frac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

https://doi.org/10.1017/S0266466604214086 Published online by Cambridge University Press

and

$$\boldsymbol{ADA} = \frac{1}{4}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} d_1 & 0 \\ 0 & d_2 \end{pmatrix}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} = \frac{d_1 + d_2}{4}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$$

Equating the latter to zero, as in (2), provides no further information on the variance of the two normal components of $\boldsymbol{x}$, beyond what was already known from (1). In this case, result (b) does not hold.

As a counterexample, let

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{3}{2} \end{pmatrix}\right).$$

Then, it is still the case that $\bar{x} \sim \mathrm{N}(0, \frac{1}{2})$ and

$$z = \frac{1}{2}(x_1, \quad x_2)\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \frac{1}{2}(x_1 - x_2)^2 \sim \chi^2(1).$$

Notice, however, that $\mathrm{cov}(x_1 + x_2, x_1 - x_2) = \mathrm{var}(x_1) - \mathrm{var}(x_2) \neq 0$, so that $\bar{x}$ is not independent of z. We will now show that assuming independence of $\bar{x}$ from z makes the statement in (b) hold for $n = 2$ also.

Independence of the linear form $\boldsymbol{\imath}'\boldsymbol{x}/n$ from the quadratic form $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}/\sigma^2$ occurs if and only if $\boldsymbol{A\Sigma\imath} = \boldsymbol{0}$. For $n = 2$, setting

$$\frac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} \sigma_1^2 - \sigma_2^2 \\ \sigma_2^2 - \sigma_1^2 \end{pmatrix}$$

equal to zero ensures that $\sigma_1^2 = \sigma_2^2$.

A variation on part (c) is proved by a different approach in Zinger (1958, Theorem 6). There, independence of $\bar{x}$ from z is assumed but not the normality of $\boldsymbol{x}$. In fact, for $2 \leq n < \infty$, normality of $\boldsymbol{x}$ is obtained there as a result of one of two alternative assumptions on the components of $\boldsymbol{x}$ being pairwise identically distributed or being decomposable further as independent and identically distributed (i.i.d.) variates.

NOTE

1. An independent solution has been proposed by Luc Lauwers, KU Leuven, Belgium.

REFERENCES

Cramér, H. (1936) Über eine eigenschaft der normalen verteilungfunktion. *Mathematische Zeitschrift* 41, 405–414.

Loève, M. (1977) *Probability Theory I*. 4th ed. Springer-Verlag.

Zinger, A.A. (1958) Independence of quasi-polynomial statistics and analytical properties of distributions. *Theory of Probability and Its Applications* 3, 247–265.

https://doi.org/10.1017/S0266466604214086 Published online by Cambridge University Press

03.3.2. The Asymptotic Distribution of the Dickey–Fuller Statistic under Nonnegativity Constraint—Solution[1]

Giuseppe Cavaliere (the poser of the problem)
University of Bologna, Italy

It is convenient to express the Dickey–Fuller (DF) statistic as

$$Z_T = \frac{(1/T)\left(X_T^2 - \sum_{t=1}^{T} (\Delta X_t)^2\right)}{(2/T^2)\sum_{t=1}^{T} X_{t-1}^2}$$

(see Phillips, 1987). To prove that the asymptotic distribution of Z_T under the nonnegativity constraint is standard, we need to prove that $(T^{-1}X_T^2, T^{-2}\sum_{t=1}^{T} X_{t-1}^2)' \xrightarrow{w} \sigma^2(B(1)^2, \int_0^1 B(s)^2\,ds)'$ and that $T^{-1}\sum_{t=1}^{T}(\Delta X_t)^2 \xrightarrow{p} \sigma^2$.

The weak convergence result can be obtained by showing that the constrained random walk $\{X_t\}$ can be expressed as a functional of the simple random walk $S_t := \sum_{i=1}^{t} \varepsilon_t$, $S_0 = 0$. This key result is given in the following lemma, which is adapted from Proposition 4 in Cavaliere (2002).

LEMMA 1. *For each $t \geq 0$, $X_t = S_t + L_t$, where $S_t := \sum_{i=0}^{t} \varepsilon_i$ and $L_t = -\min_{i \leq t}\{S_i\}$.*

Proof. First, note that for each t the process can be written as $X_t = X_{t-1} + \varepsilon_t + l_t$, where

$$l_t = -(X_{t-1} + \varepsilon_t)\mathbb{I}\{X_{t-1} + \varepsilon_t < 0\}. \qquad \textbf{(3)}$$

Hence $X_t = X_0 + \sum_{i=1}^{t} \varepsilon_i + \sum_{i=1}^{t} l_i = X_0 + S_t + L_t$, where $L_t := \sum_{i=1}^{t} l_i$. We need to show that $L_t = -\min_{i \leq t}\{S_i\}$, which can be proved by induction. For $t = 0$, $S_0 = 0$, $L_0 = 0$ and the relation is therefore satisfied. Assume that the relation is satisfied at time t, i.e., $L_t = -\min_{i \leq t}\{S_i\}$. At time $t + 1$, if $X_t + \varepsilon_{t+1} \geq 0$, then $l_{t+1} = 0$, $L_{t+1} = L_t = -\min_{i \leq t}\{S_i\} = -\min_{i \leq t+1}\{S_i\}$, and the relation is satisfied. Conversely, if $X_t + \varepsilon_{t+1} < 0$, then $l_{t+1} = -(X_t + \varepsilon_{t+1}) = -(S_t + L_t + \varepsilon_{t+1})$, which implies $-L_{t+1} = S_{t+1}$. But because $l_{t+1} > 0$, $-L_{t+1} = S_{t+1} < -L_t = \min_{i \leq t}\{S_i\}$, which gives $-\min_{i \leq t+1}\{S_i\} = S_{t+1} = -L_{t+1}$, and the relation is proved. ■

The mapping from $\{S_t\}$ to $\{X_t\}$ is explained graphically in Figure 1, where $\{X_t\}$ is plotted along with $\{S_t\}$ and $\{-L_t\}$. It is interesting to notice that the cumulated effect of the nonnegativity constraint, i.e., the differential between the unconstrained and the constrained random walks, $S_t - X_t$, is given by the running minimum $\min_{i \leq t}\{S_i\} =: -L_t$.

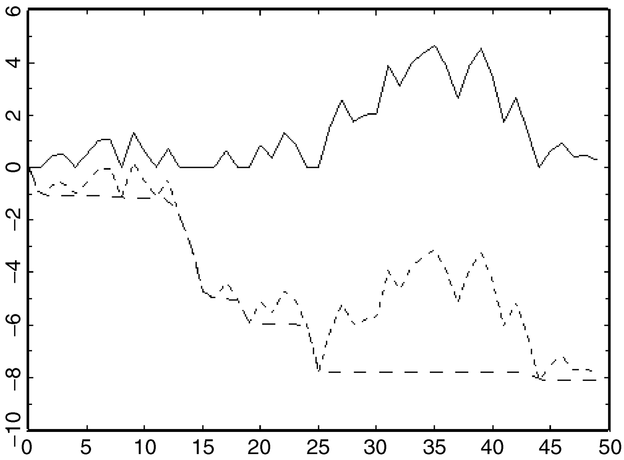

FIGURE 1. Construction of the constrained random walk: X_t (solid line), S_t (short dashed line), and $-L_t$ (dashed line), $T = 50$, $\sigma = 1$.

From Lemma 1, the invariance principle, and the continuous mapping theorem, as $T \uparrow \infty$

$$T^{-1/2}\sigma^{-1}X_{[sT]} = T^{-1/2}\sigma^{-1}S_t - \min_{i\le t}\{T^{-1/2}\sigma^{-1}S_i\}$$

$$\xrightarrow{w} B(s) - \inf_{r\le s} B(r) \stackrel{d}{=} |B(s)| \tag{4}$$

uniformly for all $s \in [0,1]$. The probability result given by the last distributional equality in (4) is well known in the probability literature (see, e.g., Harrison, 1985). Note that it also implies $T^{-1}\sigma^{-2}X_{[sT]}^2 \xrightarrow{w} |B(s)|^2 = B(s)^2$ and hence

$$\begin{pmatrix} T^{-1}X_T^2 \\ T^{-2}\sum_{t=1}^{T} X_{t-1}^2 \end{pmatrix} \xrightarrow{w} \sigma^2 \begin{pmatrix} B(1)^2 \\ \int_0^1 B_s^2\, ds \end{pmatrix}.$$

The proof is completed by showing that the nonnegativity constraint does not affect the estimation of the variance σ^2. Because $(1/T)\sum_{t=1}^{T}\varepsilon_t^2 \xrightarrow{p} \sigma^2$ ($\{\varepsilon_t\}$ is Gaussian and independently and identically distributed [i.i.d.]), we only need to prove that $|(1/T)\sum(\Delta X_t)^2 - (1/T)\sum\varepsilon_t^2| = o_p(1)$. Let l_t, $0 \le l_t \le |\varepsilon_t|$, be defined as in (3). As $(\Delta X_t)^2 = (\varepsilon_t + l_t)^2 = \varepsilon_t^2 + l_t(l_t + 2\varepsilon_t) \le \varepsilon_t^2 + 3l_t|\varepsilon_t|$,

https://doi.org/10.1017/S0266466604214086 Published online by Cambridge University Press

$$\left|\frac{1}{T}\sum_{t=1}^{T}(\Delta X_t)^2 - \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2\right| \le \frac{3}{T}\sum_{t=1}^{T} l_t|\varepsilon_t| \le 3\,\frac{\max_{t=1,\dots,T}|\varepsilon_t|}{\sqrt{T}}\,\frac{1}{\sqrt{T}}\sum_{t=1}^{T} l_t. \qquad \textbf{(5)}$$

But $T^{-1/2}\sum_{t=1}^{T} l_t = -T^{-1/2}\min_{t=1,\dots,T}\{S_t\} \xrightarrow{w} -\sigma\inf_{s\le 1}B(s)$ and $T^{-1/2}\max_{t=1,\dots,T}|\varepsilon_t| = o_p(1)$ ($\{\varepsilon_t\}$ is i.i.d. and Gaussian); hence, the random variable on the right-hand side of (5) is $o_p(1)$, and the desired result follows.

Final Remark. The asymptotics obtained show that the nonnegativity constraint does not affect the asymptotic distribution of the Dickey–Fuller unit root test. However, the same property does not automatically apply to other unit root tests (e.g., tests that employ deterministic corrections). See Cavaliere (2001) for further insights.

NOTE

1. An independent solution has been proposed by Paulo Rodrigues, University of Algrave, Portugal, jointly with Antonio Rubia, University of Alicante, Spain.

REFERENCES

Cavaliere, G. (2001) Asymptotics for Nonstationary Time Series under Range Constraints. Preprint 7/2001, Department of Statistics and Operations Research, University of Copenhagen.

Cavaliere, G. (2002) Bounded integrated processes and unit root tests. *Statistical Methods and Applications—Journal of the Italian Statistical Society* 11, 41–70.

Harrison, M.J. (1985) *Brownian Motion and Stochastic Flow Systems*. Wiley.

Phillips, P.C.B. (1987) Time series regression with a unit root. *Econometrica* 55, 277–301.