ALTERNATIVE PROOF OF RUBIN'S RESULT (I. I. 3) EXPANDING THE LIKELIHOOD FUNCTION FOR AN INCOMPLETE LINEAR SYSTEM

by Tjalling C. Koopmans

The alternative proof is based on replacing Rubin's restrictions:

\[ \frac{A_{u|x}}{u} = ( \frac{A_{u|y}}{u} = \frac{I_{u|x}}{u} - \frac{1}{2} \frac{u^2}{u} x ) \]

on the "unknown" part of the system by the restrictions:

\[ \frac{A_{u|x}}{u} = ( \frac{A_{u|y}}{u} = \frac{I_{u|x}}{u} ) \]

Since any matrix

\[ A_{u|x} = \begin{pmatrix} \frac{A_{u|x}}{u} \\ \frac{A_{u|x}}{u} \end{pmatrix} \]

can be made to satisfy condition (2) by a transformation

\[ \begin{align*}
\frac{A_{u|x}}{u} &= \Gamma_{u|u} A_{u|x} \\
\frac{I_{u|x}}{u} &= \begin{pmatrix} \frac{I_{u|x}}{u} & 0 \end{pmatrix}
\end{align*} \]

which leaves \( A_{u|x} \) unaffected and preserves the form of the likelihood function:

\[ \log L = \log | A_{u|y} | - \frac{1}{2} \log | A_{u|u} | - \frac{1}{2} \text{tr} E_{u|x}^{-1} A_{u|x} M_{x} A_{u|x} \]

the addition of (2) to whatever a priori restrictions are imposed on \( A_{u|x} \)

does not constitute a further restriction of the form of the probability

distribution of the observations.

\* omitting irrelevant additive constants

+ This paper has been delayed through temporary loss of the stencil. It was prepared in the Autumn of 1947.
Following Rubin, we first maximize (6) with respect to $I_{uu}$ and insert the result to obtain the (footnote on page 1)

$$
\log \hat{L} = \log |A_{uv}| - \frac{1}{2} \log |A_{xx} M_{xx} A_{xx}^t |.
$$

Next we maximize (6) with respect to $A_{u \overline{I} \overline{I}}$ for a given value of $A_{u \overline{I} \overline{I}}$.

Without the restrictions (2) on $A_{u \overline{I} \overline{I}}$ the first order conditions for a maximum would be, if $I_{u \overline{I} \overline{I}} = (A_{u \overline{I} u \overline{I}} \quad I_{u \overline{I} \overline{I}})$,

$$
\frac{\partial \log \hat{L}}{\partial A_{u \overline{I} u \overline{I}}} = I_{\overline{u} \overline{u}}, \quad \frac{\partial \log \hat{L}}{\partial A_{u \overline{I} \overline{I}}} = I_{\overline{u} \overline{u}} \left[ A_{u \overline{I} \overline{I}}^{-1} I_{u \overline{I} \overline{I}} - \left( A_{u \overline{I} \overline{I}}^{-1} M_{\overline{I} \overline{I}} A_{u \overline{I} \overline{I}}^{-1} \right)^{-1} A_{u \overline{I} \overline{I}} M_{\overline{I} \overline{I}} \right] = 0
$$

However, since (2) does not further restrict the distribution function, any point where (6) is stationary with respect to variations of $A_{u \overline{I} \overline{I}}$ subject to (2) must also be a point where (6) is stationary with respect to any variations of $A_{u \overline{I} \overline{I}}$ and hence satisfy $^t \bar{M}(7)$. Using (2) in (7) we have

$$
I_{u \overline{I} u \overline{I}} A_{u \overline{I} \overline{I}}^{-1} I_{u \overline{I} \overline{I}} - A_{u \overline{I} \overline{I}} M_{\overline{I} \overline{I}} = 0,
$$

or

$$
A_{u \overline{I} \overline{I}} = I_{u \overline{I} u \overline{I}} A_{u \overline{I} \overline{I}}^{-1} I_{u \overline{I} \overline{I}} M_{\overline{I} \overline{I}}^{-1} .
$$

Without evaluating the maximizing value $A_{u \overline{I} \overline{I}}$ (which occurs also in the right hand member of (9)) we eliminate it from (9) and (8) as follows.

Writing

$$
W_{\overline{I} \overline{I}} = (I_{\overline{I} \overline{I}} M_{\overline{I} \overline{I}}^{-1} I_{\overline{I} \overline{I}})^{-1},
$$

Footnote: For a full justification of this statement see the last para. of this note.
we have

\[ A_{uv} = \begin{pmatrix} A_{u,v} \\ A_{v,u} \end{pmatrix} = \begin{pmatrix} A_{u,v} \\ I_{u,u} \cdot A_{v,v}^{-1} \cdot A_{v,v} \end{pmatrix}, \]

and hence

\[ A_{uv} \cdot W_{uv} \cdot A_{uv}^T = \begin{pmatrix} A_{u,u} \cdot W_{uv} \cdot A_{v,v}^T \\ I_{u,u} \end{pmatrix}. \]

From (8), (2) and (12), it follows that the likelihood function after maximization with respect to \( A_{u,I}^2 \) can be written as

\[
\log L = \frac{1}{2} \log \left| A_{uv} W_{uv} A_{uv}^T \right| - \frac{1}{2} \log \left| W_{uv} \right| - \frac{1}{2} \log \left| A_{ux} W_{ux} A_{ux}^T \right| = \\
= \frac{1}{2} \log \left| A_{uv} W_{uv} A_{uv}^T \right| - \frac{1}{2} \log \left| W_{uv} \right| - \frac{1}{2} \log \left| A_{ux} W_{ux} A_{ux}^T \right|,
\]

which is Rubin's equation (8, 23).

This proof avoids the use of "perpendiculars" of the type

\[ \mathbf{a}_{x}^{\times} \mathbf{a}_{y}^{\times}, \quad (\operatorname{except \ that} \mathbf{w}_{x}^{\times} = \mathbf{w}_{y}^{\times}). \]

However, to make the statement following (7) above quite rigorous, it is necessary to specify \( \mathbf{T}_{u,I}^{\times u} \) in (4) as a continuous function of \( A_{ux} \) which assumes the value \( I_{u,I} \) whenever \( A_{ux} \) satisfies the restrictions (2). One possible choice is given by

\[
A_{u,I}^T \mathbf{a}_{x}^{\times} = -A_{u,I}^T \mathbf{w}_{x}^{\times} \mathbf{a}_{x}^{\times} A_{u,I} (A_{u,I}^T \mathbf{w}_{x}^{\times} A_{u,I}^{-1})^{-1} A_{u,I}^T + A_{u,I}^T \mathbf{a}_{x}^{\times} = \left( A_{I,I}^T \mathbf{w}_{x}^{\times} A_{I,I}^T \right)^{-1} \frac{1}{2} A_{I,I}^T \mathbf{a}_{x}^{\times},
\]

taking the positive definite square root.