I. Linear Case. Suppose we are presented with a complete linear system of stochastic equations

\[(1) \quad \beta_{qy} y_t' + \Gamma_{qs} z_t' = q_t',\]

where \(y_t\) is jointly dependent at "time" \(t\), \(z_t\) is predetermined at "time" \(t\), and \(q_t\) are independently and normally distributed with mean 0 and constant non-singular covariance matrix \(\Sigma\). Suppose that \(q = (u \ r)\). Then (1) may be written as

\[(2) \quad \beta_{uy} y_t' + \Gamma_{uz} z_t' = u_t',\]
\[(3) \quad \beta_{ry} y_t' + \Gamma_{rz} z_t' = r_t'.\]

Then the maximum likelihood estimates of \(\alpha_{ux} = (\beta_{uy} \Gamma_{uz})\) is that matrix \(\hat{\alpha}_{ux}\) which minimizes

\[(3) \quad V = \left| \begin{array}{cc} A_{ux} & M_{xx} A_{ux} \Gamma_{ux} \\ A_{ux} W_{xx} A_{ux} & A_{ux} \end{array} \right|\]

where

\[(4) \quad W_{xx} = M_{xx} - M_{xx} M_{xx}^{-1} M_{xx},\]

or the moment matrix\(^1\) of the residuals of the coordinates of \(X_t\) about their regression on the coordinates of \(z_t\).

II. Non-Linear Case. Suppose \(q_t\) are serially independent random variables (not necessarily in a Euclidean space) and suppose there exist measurable vector

---

1. By \(M_{ab}\) we mean \(\frac{1}{T} \sum_{t=1}^{T} a_t' b_t\).
functions \( x_t = f_t \left( q_t, q_{t-1}, \ldots \right), u_t = g_t \left( q_t, q_{t-1}, \ldots \right) \) such that

\[ (5) \quad \mathbf{A}_{ux} x_t = u_t. \]

Suppose further that the vector \( z_t = h_t \left( q_{t-1}, q_{t-2}, \ldots \right) \). Then why not estimate \( \mathbf{A}_{ux} \) by the same procedure as before?

For example, suppose

\[ \begin{align*}
  y_{1t} + \alpha v_{1t} y_{2t} + \beta y_{3t-1} + \gamma = q_{1t} y_{1,t-1} \\
  y_{3t} + \delta y_{1t} y_{3t} + \varepsilon v_{2t} + \theta = q_{2t} v_{1t} v_{2t}
\end{align*} \]

and \( v_1 \) and \( v_2 \) are exogenous. Then we may take

\[ \begin{align*}
x_t &= \left( y_{1t}, y_{2t}, v_{1t}, v_{2t}, y_{3,t-1}, 1, y_{3t}, y_{3t}, v_{1t}, v_{2t} \right), \\
u_t &= \left( q_{1t} y_{1,t-1}, q_{2t} v_{1t} v_{2t} \right)
\end{align*} \]

and \( z_t = \left( v_{1t}, y_{3,t-1}, v_{2t}, 1, y_{3t} v_{2t}, v_{1t}^2 \right) \). Then

\[ \mathbf{A}_{ux} = \begin{pmatrix}
  1 & \alpha & \beta & 0 & 0 & 0 \\
  0 & 0 & 0 & \delta & 1 & \varepsilon
\end{pmatrix}. \]

The important properties of maximum likelihood estimates which cause us to use them are consistency, efficiency and asymptotic normality. We in general cannot get efficiency without a much better knowledge of the process than that contained in \( \mathbf{A}_{ux} \), but we may still hope for the other properties. We find that certain "reasonable" requirements, which are also the same in the case of maximum likelihood estimates, suffice here. Let us examine these conditions.

In the case of maximum likelihood estimates, \( \mathbb{E}(z_t' u_t) = 0 \) for all \( t \). We need an asymptotic condition similar to this.

Condition I. \( \lim_{t \to \infty} M_{uz} M_{zu}^{-1} M_{zu} = 0, \quad \mathbb{E}(u_t' z_t) = 0. \)

I.e., for large enough \( T, M_{zu} \) is close to 0 "most of the time." This condition stated mathematically is: For any \( \varepsilon > 0 \), any any \( \delta > 0 \), for \( T > T(\varepsilon, \delta) \), the probability that \( ^3 \text{tr}(M_{zu} M_{uz}) < \delta^{-2} \) is greater than \( 1 - \varepsilon \).

Let us write \( \text{tr} \mathbf{H} H' = \| \mathbf{H} \|_2^2 \). This notation will be convenient in what follows:

---

2. A weaker condition than this can be used.

3. Any other definition of norm could be used.
For maximum likelihood estimates, plim \( M_{uu} = Z \). We need merely that the u's are not "too wild," i.e.,

Condition II: For any \( \varepsilon > 0 \), there is an \( N_1 \) such that \( ||M_{uu}|| \leq N_1 \)

with probability greater than \( 1 - \varepsilon \), and an \( N_2 \) such that for sufficiently large \( T \), \( ||M_{uu}^{-1}|| \leq N_2 \) with probability greater than \( 1 - \varepsilon \).

We also need that the z's do not get too big, i.e.,

Condition III. For every \( \varepsilon > 0 \), there is an \( N_3 \) such that \( ||M_{zz}|| < N_3 \)

with probability greater than \( 1 - \varepsilon \).

Let

\[
\mu_{xz} = \frac{1}{T} \sum_{t=1}^{T} (x_t | t-1) z_t.
\]

Then

\[
\alpha_{ux} \mu_{xz} = \frac{1}{T} \sum_{t=1}^{T} (u_t | t-1) z_t = 0.
\]

If \( M_{zz} \) approaches a singular matrix, the estimates need not be consistent.

Let \( Q \) be an orthogonal matrix such that \( Q M_{zz} Q' \) is diagonal and the roots are arranged in descending order. Let \( Q_\varepsilon \) be the matrix composed of those rows of \( Q \) for which the characteristic roots are greater than \( \varepsilon \). We also need something corresponding to identifiability, but slightly stronger (less common)

Condition IV: For every \( \varepsilon > 0 \), there is a \( \varepsilon > 0 \), such that for \( T \) sufficiently large, \( \alpha_{ux} M_{xz} Q_\varepsilon = 0 \) has a unique solution with probability greater than \( 1 - \varepsilon \).

We also need

Condition V: \( \text{plim} \ M_{xz} = \mu_{xz} = 0 \)

Condition VI: For every \( \varepsilon > 0 \), there is a bounded closed set \( \mathcal{S} \) in \( M_{xz} \) space such that \( \mu_{xz} \) is in \( \mathcal{S} \) with probability greater than \( 1 - \varepsilon \)

for \( T \) sufficiently large, and if \( H \) is in \( \mathcal{S} \), then for every neighborhood \( \mathcal{N} \) of \( \alpha_{ux} \), if \( K \) is sufficiently near \( H \) and \( AK = 0 \), then \( A \) is in \( \mathcal{N} \).
This last condition states that $\mu_{xz}$ does not "misbehave" too violently and that it stays in a region in which we may "continuously" solve for $A$ as a function of $\mu_{xz}$.

Under these conditions, $\hat{\mu}_{ux}$ is a consistent estimate of $\alpha_{ux}$.

If, in addition, the elements of $\sqrt{T}M_{uz}$ are asymptotically normally distributed, so are the elements of $\hat{\mu}_{ux}$, provided the elements of $\alpha_{ux}$ are functions of the elements of $\mu_{xz}$ with continuous first partial derivatives. If the covariance of $\sqrt{T}m_{ij}z_k$ and $\sqrt{T}m_{ij}z_e$ is approximately $m_{ij}m_{ek}z_k^2$, then the asymptotic covariance matrix of the parametric system for $\alpha_{ux}$ is twice the inverse of the second partial derivative matrix of log $V$ with respect to those parameters.