Instrumental Variable Methods (IVM) (Ch. 8)

Main Idea: Modify the LS method to be consistent also for correlated disturbances.

The least squares estimate
\[
\hat{\theta} = \left[ \frac{1}{N} \sum_{t=1}^{N} \varphi(t)\varphi^T(t) \right]^{-1} \left[ \frac{1}{N} \sum_{t=1}^{N} \varphi(t)y(t) \right]
\]
has the estimation error (when \( N \to \infty \))
\[
\hat{\theta} - \theta_0 = E \left[ \varphi(t)\varphi^T(t) \right]^{-1} E \left[ \varphi(t)\varepsilon(t) \right]
\]
Consequently, for \( \hat{\theta} - \theta_0 = 0 \) to hold, we must have
\[
E \left[ \varphi(t)\varepsilon(t) \right] = 0,
\]
which is satisfied if, and essentially only if, \( \varepsilon(t) \) is white noise. Hence, the least squares estimate is not consistent for correlated noise sources.

Consider the ARX model,
\[
A(q^{-1})y(t) = B(q^{-1})u(t) + \varepsilon(t)
\]
or, equivalently,
\[
y(t) = \varphi^T(t)\theta + \varepsilon(t)
\]
where \( \varepsilon(t) \) is the equation error \( (y(t) - y_m(t)) \), and
\[
\varphi(t) = [y(t - 1) \ldots - y(t - n_a)u(t - 1) \ldots u(t - n_b)]^T
\]
\[
\theta = [a_1 \ldots a_{n_a} b_1 \ldots b_{n_b}]^T
\]

Cure:
- PEM (last lecture). Model the noise.
  - Applicable to general model structures.
  - Generally very good properties of the estimates.
  - Computationally quite demanding.
- Instrumental variable methods (IVM). Do not model the noise.
  - Retain the simple LS structure.
  - Simple and computationally efficient approach.
  - Consistent for correlated noise.
  - Less robust and statistically less effective than PEM.
### The IV method

Introduce a vector \( z(t) \in \mathbb{R}^{n_t} \) with entries uncorrelated with \( \varepsilon(t) \). Then (for large values of \( N \))

\[
0 = \frac{1}{N} \sum_{t=1}^{N} z(t) \varepsilon(t) - \frac{1}{N} \sum_{t=1}^{N} z(t) [y(t) - \varphi^T(t)\theta]
\]

which yields (if the inverse exists)

\[
\dot{\theta} = \frac{1}{N} \sum_{t=1}^{N} z(t) \varphi^T(t) \left( \frac{1}{N} \sum_{t=1}^{N} z(t)\varepsilon(t) \right)^{-1}
\]

The elements of \( z(t) \) are usually called the **instruments**. Note that if \( z(t) = \varphi(t) \), the IV estimate reduces to the LS estimate.

---

### Choice of Instruments

Obviously, the choice of instruments is very important. They have to be chosen

(i) such that \( z(t) \) is uncorrelated with \( \varepsilon(t) \) \( \langle E z(t)\varepsilon(t) \rangle = 0 \), and

(ii) such that the matrix

\[
\frac{1}{N} \sum_{t=1}^{N} z(t)\varphi^T(t) \rightarrow E z(t)\varphi^T(t)
\]

has full rank. In other words it is essential that \( z(t) \) and \( \varphi(t) \) are correlated.

---

### Extended IV methods

Recall that the basic IV estimate is derived from

\[
\min_{\theta} \sum_{t=1}^{N} z(t)\varepsilon(t)^2
\]

More flexibility is obtained if the instrument vector \( z(t) \) is augmented to dimension \( n_z \) \( (n_z \geq n_\theta) \), and if we allow for a weighting and a prefiltering of the residuals by some stable filter \( F(q^{-1}) \), i.e.,

\[
\min_{\theta} \sum_{t=1}^{N} z(t)F(q^{-1})\varepsilon(t)^2
\]

where \( ||x||_Q^2 = x^TQx \) and \( Q \) is a positive definite weighting matrix.
Inserting

\[ z(t) = y(t) - \varphi^T(t)\theta \]

yields the so-called extended IV method

\[
\hat{\theta} = \arg \min_{\theta} \left\{ \frac{1}{N} \sum_{t=1}^{N} z(t) F(q^{-1}) \varphi(t) \right\}^2
\]

When \( F(q^{-1}) = 1 \) and \( Q = I \), the basic IV method is obtained.

Introduce

\[
R_N = \frac{1}{N} \sum_{t=1}^{N} z(t) F(q^{-1}) \varphi(t)
\]

\[
r_N = \frac{1}{N} \sum_{t=1}^{N} z(t) F(q^{-1}) y(t)
\]

Then

\[
\hat{\theta} = \arg \min_{\theta} \left\{ R_N \theta - r_N \right\}_Q^2
\]

\[
= \arg \min_{\theta} \left\{ R_N \theta - r_N \right\}^T Q (R_N \theta - r_N)
\]

\[
= \left[ R_N^T Q R_N \right]^{-1} R_N^T Q r_N
\]

Note that due to numerical instability the algorithm should not be implemented in this manner.

Rem: Notice that \( R_N \) is in general not a square matrix.

---

Assumptions

(i) The system is strictly causal and asymptotically stable.

(ii) The input is persistently exciting of a sufficiently high order.

(iii) The disturbance is a stationary stochastic process with rational spectral density,

\[ z(t) = H(q^{-1}) r(t), \quad E r^2(t) - \lambda^2 \]

(iv) The input and the disturbance are independent.

(v) The model and the true system have the same transfer function if and only if \( \theta = \theta_0 \) (uniqueness).

(vi) The instruments and the disturbances are uncorrelated.

Consider the system

\[ y(t) = \varphi^T(t) \theta_0 + \varepsilon(t) \]

Then

\[
r_N = \frac{1}{N} \sum_{t=1}^{N} z(t) F(q^{-1}) y(t)
\]

\[
= \frac{1}{N} \sum_{t=1}^{N} z(t) F(q^{-1}) \varphi(t) \theta_0 + \frac{1}{N} \sum_{t=1}^{N} z(t) F(q^{-1}) z(t)
\]

\[
= R_N \theta_0 + q_N
\]
Thus 
\[ \hat{\theta} - \theta_0 = [R_N^T Q R_N]^{-1} R_N^T Q q_N - [R^T Q R]^{-1} R^T Q q \]
where 
\[ R \triangleq \lim_{N \to \infty} R_N = E [z(t)F(q^{-1})\varphi^T(t)] \]
\[ q \triangleq \lim_{N \to \infty} q_N = E [z(t)F(q^{-1})\varepsilon(t)] \]
Therefore, the IV estimate will be consistent \((\lim_{N \to \infty} \hat{\theta} - \theta_0)\) if
(i) \( R \) has full rank (inaccurate estimates will be obtained if \( R \) is nearly rank deficient),
(ii) \( E [z(t)F(q^{-1})\varepsilon(t)] = 0. \)

Furthermore, the parameter estimation errors are asymptotically Gaussian distributed with zero mean and variance \( P_{IV} \)

\[ \sqrt{N}(\hat{\theta} - \theta_0) \to N(0, P_{IV}) \]
where 
\[ P_{IV} = \lambda^2 (R^T Q R)^{-1} R^T Q S Q R (R^T Q R)^{-1} \]
where 
\[ S = E [F(q^{-1})H(q^{-1})\varepsilon(t)] [F(q^{-1})H(q^{-1})\varepsilon(t)]^T \]

Rem: For multivariable systems \( S \) must be modified.

**Optimal IVM**

The main usefulness in being able to express \( P_{IV} \) lies in the comparison to \( P_{PEM} \) (recall that PEM is efficient for Gaussian disturbances). An “appropriate” choice of parameters leads to the optimal IVM. For example, (single output)

\[ z(t) = H^{-1}(q^{-1})\tilde{\varphi}(t) \]
\[ F(q^{-1}) = H^{-1}(q^{-1}) \]
\[ Q = I \]

where \( \tilde{\varphi}(t) \) is the noise-free part of \( \varphi(t) \). Then,

\[ P_{IV}^{opt} = \lambda^2 \left\{ E [H(q^{-1})\tilde{\varphi}(t)H(q^{-1})\tilde{\varphi}^T(t)] \right\}^{-1} \]

and \( P_{IV} \geq P_{IV}^{opt} \geq P_{PEM} \).

**Approximative implementation of the optimal IVM**

Note that the optimal instruments can not be implemented as it requires knowledge of the undisturbed output, the noise variance \((\lambda^2)\), and the shaping filter \( H(q^{-1}) \). Fortunately, it is possible to find an approximate (iterative) implementations.

One way is the following four-step IV estimator:

(i) Use the least-squares estimate of

\[ y(t) = \varphi^T(t)\tilde{\theta} \quad \Rightarrow \quad \tilde{\theta}_N^{(i)} \]
(ii) Use the IV estimator with the instruments
\[ z^{(1)}(t) = \begin{bmatrix} -x^{(1)}(t-1) & \ldots & -x^{(1)}(t-n) & u(t-1) & \ldots & u(t-n_1) \end{bmatrix} \]
where \( x^{(1)}_t = \frac{\hat{\Phi}^{(1)}(q^{-1})}{\hat{A}_N^{(1)}(q^{-1})} u_t \Rightarrow \hat{\theta}^{(2)}_N. \)

(iii) Estimate \( H(q^{-1}) \). Postulate an AR model, and use the least-squares method
\[ L(q^{-1})\hat{w}^{(2)}(t) = e(t), \Rightarrow \hat{L}_N(q^{-1}) \]
where \( \hat{w}^{(2)}_t = \hat{A}^{(2)}_N(q^{-1}) y(t) - \hat{B}^{(2)}_N(q^{-1}) u(t) \)

(iv) Use the IV estimator with \( F(q^{-1}) = \hat{L}(q^{-1}) \), and
\[ z^{(2)}(t) = \hat{L}_N(q^{-1})[\begin{bmatrix} -x^{(2)}(t-1) & \ldots & -x^{(2)}(t-n) & u(t-1) & \ldots & u(t-n_1) \end{bmatrix}] \]

Summary IVM

- The implementation of the PEM is computationally complex for many model structures.
- The computationally convenient LS method is normally biased for such model structures (i.e. for correlated disturbances).
- The IV method uses instruments that are uncorrelated with the disturbances to make the “LS-like” solution consistent.
- The parameters obtained by the IV method are thus consistent (if the instruments are chosen with care) but has a (slightly) higher variance than the PEM estimates.
- Approximately optimal IV methods can be implemented in an iterative manner to achieve the lowest possible variance of the IV estimates.