Assume that $y_t$ is an n-dimensional I(1) process with VEC form:

$$\Delta y_t = C_1 \Delta y_{t-1} + \ldots + C_{p-1} \Delta y_{t-p+1} + C_0 y_{t-1} + \varepsilon_t$$

where

$$\varepsilon_t \sim \text{w.n.} \ (\Omega)$$

$$C_0 = -BA',$$

A is an nxh matrix, $h < n$, which spans the CI space of $y$ (i.e., A has rank $h$ and $A'y_t \sim I(0)$),

B is an nxh “factor-loading” matrix.
Notes –

1. The nxn matrix $C_0$ has rank $h$
2. This existence of this representation of a CI process is an implication of the “Granger Representation Theorem” (Engle and Granger, *Ecta*, 1987.)
3. Consider the special case with $n = 2$, $p = 1$, $h = 1$.

\[
\begin{align*}
\Delta y_{1t} &= b_1 (a_1 y_{1t-1} + a_2 y_{2t-1}) + \varepsilon_{1t} \\
\Delta y_{2t} &= b_2 (a_1 y_{1t-1} + a_2 y_{2t-1}) + \varepsilon_{2t}
\end{align*}
\]

where $a_1 y_{1t} + a_2 y_{2t}$ is a zero-mean I(0) process.
How to estimate the VECM?

OLS applied to each equation? The problem is that the OLS estimator does not constrain the matrix $C_0$ to be rank $h$.

Quasi-MLE: Act under the assumption that the $\varepsilon$’s are normally distributed, i.e., $\varepsilon_t \sim \text{i.i.d. } N(0, \Omega)$ and then maximize the log-likelihood function with respect to the elements of $C_1, \ldots, C_{p-1}, A, B, \text{ and } \Omega$. (How to select $h$?)

The QMLE approach appears problematic because of the apparent need to estimate $A$ and $B$ (and, possibly, the other parameters) numerically.
Johansen, in a series of papers, provided:

- A simple procedure based on the principles of “reduced-rank” regressions to compute the MLE of the VECM for a given $h$.
- The asymptotic distributions of likelihood ratio statistics for testing the size of $h$ (including $H_0: h = 0$)
- The asymptotic distributions of the MLEs of $A$ and $B$. (The distributions of the C-hats are standard.)
Johansen’s MLE for Estimating the VECM –

Assume h and p are known.

The log-likelihood function for the sample $y_1, \ldots, y_T$, conditional on the initial observations $y_0, \ldots, y_{p-1}$ and the assumption that the $\varepsilon$’s are normally distributed is:

$$L(D, C_1, \ldots, C_{p-1}, B, A, \Omega) =$$

$$-\frac{Tn}{2} \log(2\pi) - \frac{T}{2} \log \det \Omega - \frac{1}{2} \sum_{t=1}^{T} \varepsilon_t^\prime \Omega^{-1} \varepsilon_t$$

where

$$\varepsilon_t = \Delta y_t - D - C_1 \Delta y_{t-1} - \ldots - C_{p-1} \Delta y_{t-p+1} + BA'y_{t-1}$$

(Note that we’ve added an intercept here, for generality. More on its interpretation, restrictions, etc. later.)
1. Fit Auxiliary Regressions

Fit a p-1-order VAR to $\Delta y_t$ (applying OLS equation-by-equation):

$$\Delta y_t = \hat{F}_0 + \hat{F}_1 \Delta y_{t-1} + \ldots + \hat{F}_{p-1} \Delta y_{t-p+1} + \hat{u}_t$$

Fit a regression (OLS equation by equation) of $y_t$ on 1, $\Delta y_{t-1}, \ldots, \Delta y_{t-p+1}$:

$$y_t = \hat{G}_0 + \hat{G}_1 \Delta y_{t-1} + \ldots + \hat{G}_{p-1} \Delta y_{t-p+1} + \hat{v}_t$$

2. Compute the (squared) sample canonical coefficients for the u-hats and v-hats. That is, compute the eigenvalues $\hat{\lambda}_1, \ldots, \hat{\lambda}_n$ of

$$\hat{\Sigma}_v^{-1} \hat{\Sigma}_u \hat{\Sigma}_u^{-1} \hat{\Sigma}_uv$$

where

$$\hat{\Sigma}_v = \frac{1}{T} \sum_1^T \hat{v}_t \hat{v}_t', \hat{\Sigma}_u = \frac{1}{T} \sum_1^T \hat{u}_t \hat{u}_t', \hat{\Sigma}_{vu} = \frac{1}{T} \sum_1^T \hat{v}_t \hat{u}_t' = \hat{\Sigma}_{uv}$$

and, WLOG, $\hat{\lambda}_1 > \ldots > \hat{\lambda}_n$. 
3. The MLE of the cointegrating space is:

\[ \hat{A} = [\hat{a}_1 \ldots \hat{a}_h] \]

where \( \hat{a}_i \) is the eigenvector of \( \hat{\Sigma}_{vu}^{-1} \hat{\Sigma}_{uu}^{-1} \hat{\Sigma}_{uv} \) associated with \( \hat{\lambda}_i \).

4. The MLE of the remaining parameters are:

\[ \hat{B} = \hat{\Sigma}_{uv} \hat{A} \]

\[ \hat{D} = \hat{F}_0 - \hat{B} \hat{A}' \hat{G}_0 \]

\[ \hat{C}_i = \hat{F}_i - \hat{B} \hat{A}' \hat{G}_i \quad i = 1, \ldots, p - 1 \]

\[ \hat{\Omega} = \frac{1}{T} \sum_{t=1}^{T} (\hat{u}_t - \hat{B} \hat{A}' \nu_t)(\hat{u}_t - \hat{B} \hat{A}' \nu_t)' \]
Notes –

1. Given $p$ and $h$, the maximum value of the likelihood function is:

$$L^* = \frac{-Tn}{2} \log(2\pi) - \frac{Tn}{2} - \frac{T}{2} \log \det \hat{\Sigma}_{uu} - \frac{T}{2} \sum_{i}^{h} (1 - \lambda_i)$$

This statistic is the basis for testing the size of $h$ and $p$. Note that its calculation only requires steps 1 and 2 above.

2. In the VECM, we left the intercept term, $D$, unrestricted. This turns out to mean that

   i. Each of the $h$ CI relationships can have an intercept
   ii. The $n$-$h$ variables that are not CI with one another can have drifts.

   If we want to allow (i) but rule out (ii) this imposes an additional constraint on the likelihood function and requires a modification of the algorithm. (See Hamilton, pp. 643-5.)