3. Vector Autoregressive Moving Average (VARMA) Models

On-Line Supp. 3A. (**MLE vs OLS**) As we have previously seen in the Chapter, following the notation in Lütkepohl (2005), we can write (3.11) as

$$\mathbf{Y} = \mathbf{B}\mathbf{Z} + \mathbf{U},$$

where $\mathbf{Y} \equiv \left[\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_t\right]$, $\mathbf{B} \equiv \left[\mathbf{a}_0, \mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_p\right]$ $\mathbf{U} \equiv \left[\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_T\right]$, and $\mathbf{Z} \equiv \left[\mathbf{Z}_0, \mathbf{Z}_1, \ldots, \mathbf{Z}_{T\text{-}1}\right]$ with $\mathbf{Z}_t \equiv [\mathbf{1}', \mathbf{y}'_{t-1}, \mathbf{y}'_{t-2}, \ldots, \mathbf{y}'_{t-p+1}]'$.

The MLE of **B**, which contains a constant term and the autoregressive coefficients, turns out to be $\hat{\mathbf{B}} = (\mathbf{Z'Z})^{-1}\mathbf{Z'Y}$ which can be viewed as the sample analog of the population linear projection of **Y** on **Z**.

Therefore, the *jth* row of $\hat{\mathbf{B}}$ is

$$\widehat{\boldsymbol{b}_j} = \left[\sum_t^T y_{jt}\boldsymbol{z}'_t\right]\left[\sum_t^T \boldsymbol{z}_t\boldsymbol{z}'_t\right]^{-1},$$

which is just the estimated coefficient vector from an OLS regression of y_{jt} on $\boldsymbol{z}_t$.

Thus, maximum likelihood estimates of the coefficients for the *jth* equation of a VAR are found by an OLS regression of y_{jt} on a constant term and p lags of all the variables in the system.

To verify this relation, first we have to consider that

$$\sum_t^T \left[(\boldsymbol{y}_t - \widehat{\boldsymbol{B}}'\boldsymbol{z}_t)'\boldsymbol{\Sigma}_u^{-1}(\boldsymbol{y}_t - \widehat{\boldsymbol{B}}'\boldsymbol{z}_t)\right]$$

$$= \sum_t^T \left[(\boldsymbol{y}_t - \widehat{\boldsymbol{B}}'\boldsymbol{z}_t + \widehat{\boldsymbol{B}}'\boldsymbol{z}_t - \boldsymbol{B}'\boldsymbol{z}_t)'\boldsymbol{\Sigma}_u^{-1}(\boldsymbol{y}_t - \widehat{\boldsymbol{B}}'\boldsymbol{z}_t + \widehat{\boldsymbol{B}}'\boldsymbol{z}_t - \boldsymbol{B}'\boldsymbol{z}_t)\right]$$

$$= \sum_t^T \left[\left[\widehat{\boldsymbol{u}_t} + (\widehat{\boldsymbol{B}} - \boldsymbol{B})'\boldsymbol{z}_t\right]'\boldsymbol{\Sigma}_u^{-1}\left[\widehat{\boldsymbol{u}_t} + (\widehat{\boldsymbol{B}} - \boldsymbol{B})'\boldsymbol{z}_t\right]\right],$$

where the *jth* element of the vector $\widehat{\boldsymbol{u}_t}$ is the sample residual for the observation t from an OLS regression of y_{jt} on $\boldsymbol{z}_t$:

$$\widehat{\boldsymbol{u}_t} = \boldsymbol{y}_t - \widehat{\boldsymbol{B}}'\boldsymbol{z}_t.$$

It is worth noting that $\sum_t^T\left[(\boldsymbol{y}_t - \widehat{\boldsymbol{B}}'\boldsymbol{z}_t)'\boldsymbol{\Sigma}_u^{-1}(\boldsymbol{y}_t - \widehat{\boldsymbol{B}}'\boldsymbol{z}_t)\right]$ can be expanded as

$$= \sum_t^T \widehat{\boldsymbol{u}_t}'\boldsymbol{\Sigma}_u^{-1}\,\widehat{\boldsymbol{u}_t} +$$

$$+2\sum_t^T \widehat{\boldsymbol{u}_t}'\boldsymbol{\Sigma}_u^{-1}\left(\widehat{\boldsymbol{B}} - \boldsymbol{B}\right)'\boldsymbol{z}_t + \sum_t^T \boldsymbol{z}_t'\left(\widehat{\boldsymbol{B}} - \boldsymbol{B}\right)\boldsymbol{\Sigma}_u^{-1}\,(\widehat{\boldsymbol{B}} - \boldsymbol{B})'\,\boldsymbol{z}_t.$$

The middle term in this equation is a scalar and therefore it remains unchanged by applying the "trace" operator:

$$\sum_t^T \widehat{\boldsymbol{u}_t}'\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}\big(\widehat{\boldsymbol{B}}-\boldsymbol{B}\big)'\boldsymbol{z}_t = trace\left[\sum_t^T \widehat{\boldsymbol{u}_t}'\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}(\widehat{\boldsymbol{B}}-\boldsymbol{B})'\boldsymbol{z}_t\right]$$
$$= trace\left[\sum_t^T \boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}(\widehat{\boldsymbol{B}}-\boldsymbol{B})'\boldsymbol{z}_t\widehat{\boldsymbol{u}_t}'\right]$$
$$= trace\left[\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}(\widehat{\boldsymbol{B}}-\boldsymbol{B})'\sum_t^T \boldsymbol{z}_t\widehat{\boldsymbol{u}_t}'\right].$$

An important condition to remember is that the sample residuals from an OLS regression are by construction orthogonal to the explanatory variables, meaning that $\sum_t^T \boldsymbol{z}_t\widehat{u_{jt}}' = 0$ for all j and so $\sum_t^T \boldsymbol{z}_t\widehat{\boldsymbol{u}_t}' = 0$.
Therefore, $\sum_t^T\big[(\boldsymbol{y}_t-\widehat{\boldsymbol{B}}'\boldsymbol{z}_t)'\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}(\boldsymbol{y}_t-\widehat{\boldsymbol{B}}'\boldsymbol{z}_t)\big]$ simplifies to

$$= \sum_t^T \widehat{\boldsymbol{u}_t}'\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}\,\widehat{\boldsymbol{u}_t} + \sum_t^T \boldsymbol{z}_t'\big(\widehat{\boldsymbol{B}}-\boldsymbol{B}\big)\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}\,(\widehat{\boldsymbol{B}}-\boldsymbol{B})'\boldsymbol{z}_t.$$

Since $\boldsymbol{\Sigma}_{\boldsymbol{u}}$ is a positive definite matrix, $\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}$ is as well. Thus, defining the vector $\boldsymbol{z}_t^*$ as

$$\boldsymbol{z}_t^* \equiv (\widehat{\boldsymbol{B}}-\boldsymbol{B})'\boldsymbol{z}_t\,,$$
$$\sum_t^T \boldsymbol{z}_t'\big(\widehat{\boldsymbol{B}}-\boldsymbol{B}\big)\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}\,\big(\widehat{\boldsymbol{B}}-\boldsymbol{B}\big)'\boldsymbol{z}_t = \sum_t^T [\boldsymbol{z}_t^*]'\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}\boldsymbol{z}_t^*.$$

This is positive for any sequence $\{\boldsymbol{z}_t^*\}_t^T$ other than $\boldsymbol{z}_t^* = \boldsymbol{0}$ for all t. Thus, the smallest possible value that the resulting equation can take on is achieved when $\boldsymbol{z}_t^* = \boldsymbol{0}$ or when $\boldsymbol{B} = \widehat{\boldsymbol{B}}$.
Because $\sum_t^T \widehat{\boldsymbol{u}_t}'\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}\,\widehat{\boldsymbol{u}_t} + \sum_t^T \boldsymbol{z}_t'\big(\widehat{\boldsymbol{B}}-\boldsymbol{B}\big)\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}\,(\widehat{\boldsymbol{B}}-\boldsymbol{B})'\boldsymbol{z}_t$ is minimized by setting $\boldsymbol{B} = \widehat{\boldsymbol{B}}$, it follows that the likelihood function is maximized by setting $\boldsymbol{B} = \widehat{\boldsymbol{B}}$, establishing the claim that OLS regressions provide the maximum likelihood estimates of the coefficients of a vector autoregression.