研究雜談

蔡文禎

NTU, Econ, June 4, 2019

雜談什麼

1. 整理一些過去研究的結果
2. 當時想法
3. 和目前文獻的關係

虛假迴歸: 一

1. 博士論文
2. Define $y_T = \sum_{t=1}^{T} v_t$, $x_T = \sum_{t=1}^{T} w_t$.
3. v_t and w_t are both stationary, $I(d_1)$ and $I(d_2)$ processes, $-1/2 < d_1, d_2 < 1/2$.
4. The variances of $y_{_T}$ and $x_{_T}$:

$$\sigma_y^2 = \mathrm{Var}(y_{_T}) = \mathrm{Var}\left(\sum_{t=1}^{T} v_t\right)$$

$$\sigma_x^2 = \mathrm{Var}(x_{_T}) = Var\left(\sum_{t=1}^{T} w_t\right)$$

虛假迴歸: 二

1. Sowell (1990, Theorem 1) proves that

$$\sigma_y^2 = O(T^{1+2d_1}) \qquad \text{and} \qquad \sigma_x^2 = O(T^{1+2d_2}).$$

2. Davydov (1970) shows that as $T \to \infty$,

$$\frac{1}{\sigma_y} y_{[Tr]} \;\Rightarrow\; B_{d_1}(r) \qquad \text{and} \qquad \frac{1}{\sigma_x} x_{[Tr]} \;\Rightarrow\; B_{d_2}(r),$$

虛假迴歸: 三

1. Model 1: y_t is regressed on an intercept and x_t.
2. Model 2: v_t is regressed on an intercept and w_t, where $d_1 + d_2 > 0.5$.
3. Model 3: y_t is regressed on an intercept and w_t, where $d_2 > 0$.
4. Model 4: v_t is regressed on an intercept and x_t, where $d_1 > 0$.
5. Model 5: y_t is regressed on an intercept and t.
6. Model 6: v_t is regressed on an intercept and t.

虛假迴歸: 四

1. Model 1: $t_\beta = O_p(T^{1/2})$.
2. Model 2: $t_\beta = O_p(T^{d_1+d_2-0.5})$.
3. Model 3: $t_\beta = O_p(T^{d_2})$.
4. Model 4: $t_\beta = O_p(T^{d_1})$.
5. Model 5: $t_\beta = O_p(T^{1/2})$.
6. Model 6: $t_\beta = O_p(T^{d_1})$.

緩長記憶的有用模型: VARFIMA

1. Consider the maximum likelihood estimation (MLE) of a class of stationary and invertible vector autoregressive fractionally integrated moving-average (VARFIMA) processes:

$$\Phi(B)\mathrm{diag}(\nabla^{d})Y_t = \Theta(B)Z_t,$$

where

$$Y_t = (y_{1,t}, \ldots, y_{r,t})^{\top}, \quad , t = 1, 2, \ldots, T,$$

is an $r-$dimensional vector of observations of interest, and $\Phi(B)$ and $\Theta(B)$ are finite order matrix polynomials in B (usual lag operator), such that:

$$\Phi(B) = \Phi_0 - \Phi_1 B - \cdots - \Phi_p B^p, \qquad \Theta(B) = \Theta_0 + \Theta_1 B + \cdots + \Theta_q B^q,$$

VARFIMA: 二

1. The diagonal matrix $\text{diag}(\nabla^d)$ is defined as:

$$\text{diag}(\nabla^d) = \begin{bmatrix} \nabla^{d_1} & 0 & \dots & 0 \\ 0 & \nabla^{d_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \nabla^{d_r} \end{bmatrix},$$

with $\nabla = 1 - B$.

VARFIMA: 三

1. The contribution of this paper is to show that the conditional likelihood function of the VARFIMA process can be evaluated *exactly* and efficiently if the model can be represented as:

$$\text{diag}(\nabla^d)\Phi(B)Y_t = \Theta(B)Z_t. \qquad (VARIMA.1)$$

2. **Assumption 1**. Given that the data is generated by (1), we assume (i) $\Phi(B)$ is diagonal, or (ii) the values of differencing parameters d_i remain intact across $i = 1, \ldots, r$.
3. However, The parameters of the unrestricted VARFIMA models are not identified, due to the non-uniqueness of VARMA models, discussed in Lütkepohl (2005).

VARFIMA: 三

1. The r-dimensional VARMA(p, q) representation is said to be in final equations form if $\Theta_0 = I$ and $\Phi(B) = \phi_0 - \phi_1 B - \cdots - \phi_p B^p$ is a scalar operator if $\phi_p \neq 0$.
2. Another justification is the realized volatility literature where the value of d are very close to each other and at the range between 0.35 and 0.45.

VARFIMA: 四

1. When Y_t is a VARFIMA$(0, d, q)$ process, i.e.:

$$\text{diag}(\nabla^d) Y_t = \Theta(B) Z_t,$$

the (m, n)th element of its corresponding autocovariance function, $\Omega(h)$, is $\Omega_{m,n}(h)$:

$$\begin{aligned}
&\Omega^* \left(\sigma_{mn}\Theta_{mm,0}\Theta_{nn,0}\right) \\
&+\Omega^* \left\{ \sum_{f=1}^{q}\sum_{g=1}^{q}\sum_{u=1}^{r}\sum_{v=1}^{r} \sigma_{uv}\Theta_{mu,f}\Theta_{nv,g}\frac{\Gamma(h+d_n+g-f)\Gamma(h+1-d_m)}{\Gamma(h+d_n)\Gamma(h+1-d_m+g-f)} \right\} \\
&+\Omega^* \left\{ \sum_{f=1}^{q}\sum_{u=1}^{r} \sigma_{nu}\Theta_{mu,f}\frac{\Gamma(h+d_n-f)}{\Gamma(h+d_n)}\frac{\Gamma(h+1-d_m)}{\Gamma(h+1-d_m-f)} \right\} \\
&+\Omega^* \left\{ \sum_{f=1}^{q}\sum_{u=1}^{r} \sigma_{mu}\Theta_{nu,f}\frac{\Gamma(h+d_n+f)}{\Gamma(h+d_n)}\frac{\Gamma(h+1-d_m)}{\Gamma(h+1-d_m+f)} \right\},
\end{aligned}$$

VARFIMA: 五

1. Note that
$$\Omega^* = \frac{\Gamma(1-d_m-d_n)}{\Gamma(d_n)\Gamma(1-d_n)} \frac{\Gamma(h+d_n)}{\Gamma(h+1-d_m)}$$
2. $\Theta_{mn,k}$ denotes the (m,n)th element of Θ_k
3. Note that there are $(rq+1)^2$ terms. With the autocovariance functions in (13), we apply the multivariate Durbin-Levinson algorithm of Whittle to the VARFIMA$(0,d,q)$ processes.

VARFIMA: 六

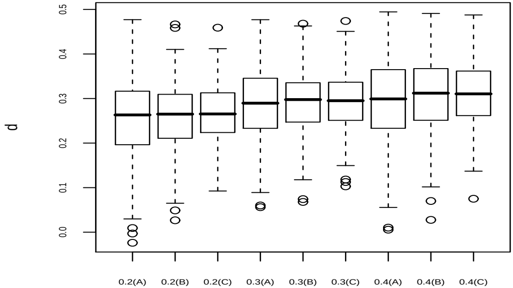

Figure: Box-plots of the estimated d from the 3-dimensional VARFIMA$(0, d, 1)$ model and $\rho = 0$. The value $f(g)$ denotes the model specification where $f = d$, g denotes the sample size, such that g=A=200, g=B=300, and g=C=400.

VARFIMA: 七

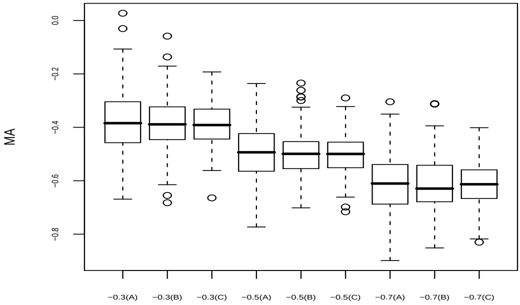

Figure: Box-plots of the estimated MA parameter from the 3-dimensional VARFIMA$(0, d, 1)$ model and $\rho = 0.5$ based on the CLDL algorithm with 250 replications. The value $f(g)$ denotes the model specification where f represents the value of MA parameter, and g denotes the sample size, such that g=A=200, g=B=300, and g=C=400.

VARFIMA: 八

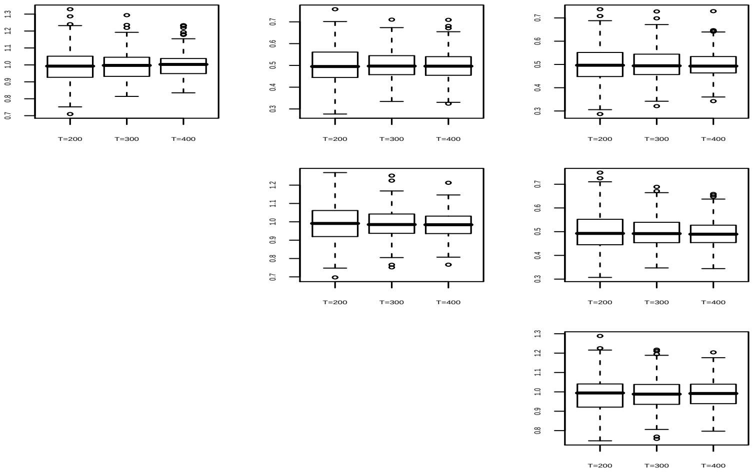

Figure: Box-plots of the estimated Σ from the 3-dimensional VARFIMA$(0, d, 1)$ model and $\rho = 0.5$ based on the CLDL algorithm with 250 replications.

緩長記憶的有用模型: VAR

1.
$$\text{diag}(\nabla^d)Y_t = \varepsilon_t,$$
where we allow the integration of each component of Y_t to be different across $i = 1, 2, \cdots, r$.

2. **Assumption 1**. (i) $\varepsilon_t = (\varepsilon_{1,t}, \ldots, \varepsilon_{r,t})'$ is an $r-$dimensional disturbance vector for the t-th observation; (ii) ε_t is an i.i.d. process with $\mathrm{E}(\varepsilon_t) = 0$, and $\mathrm{E}(\varepsilon_t \varepsilon_t') = \Sigma$, where the off-diagonal elements of Σ are allowed to be non-zero; (iii) each element of e_t has a finite fourth moment, i.e., $\mathrm{E}(\varepsilon_{i,t}^4) = \kappa_i < \infty$ for all $i = 1, 2, \ldots, r$.

VAR: 二

1. We consider the problem of predicting the $(T+1)$th value of Y_t, i.e., Y_{T+1}.
2. The predictor has the VAR(k)form:

$$\widehat{Y}_{T,k}(1) = \sum_{j=1}^{k} \widehat{A}_{jk} Y_{T-j+1}.$$

We are interested in the asymptotic properties of the prediction error $Y_{T+1} - \widehat{Y}_{T,k}(1)$.

VAR: 三

1. We estimate A_{jk}, $j = 1, \ldots, k$, by the multivariate LS coefficient estimator $\widehat{A}(k)$:

$$\widehat{A}(k) = \left(\widehat{A}_{1k}, \widehat{A}_{2k}, \ldots, \widehat{A}_{kk}\right) = \widehat{\Gamma}_{1,k}^{\top} \widehat{\Gamma}_{k}^{-1},$$

where

$$\widehat{\Gamma}_{1,k} = (T-k)^{-1} \sum_{t=k}^{T-1} Y_{t,k} Y_{t+1}^{\top}, \qquad \widehat{\Gamma}_{k} = (T-k)^{-1} \sum_{t=k}^{T-1} Y_{t,k} Y_{t,k}^{\top},$$

and

$$Y_{t,k} = \left(Y_t^{\top}, Y_{t-1}^{\top}, \ldots, Y_{t-k+1}^{\top}\right)^{\top}.$$

VAR: 四

1. **Theorem 2**. Given that Y_t is generated an multivariate fractional white noise and Assumption 1 holds, $d_i \in (0, 0.5), i = 1, 2, \cdots, r$,
2. $d = max\{d_1, d_2, \cdots, d_r\}$ and $d^* = min\{d_1, d_2, \cdots, d_r\}$,
3. $k/T^{1-2d} \to 0$ as $k, T \to \infty$ when $0.25 < d < 0.5$,
4. $k/T^{0.5}(\log\ T)^{-0.5} \to 0$ as $k, T \to \infty$ when $0 < d \leq 0.25$,
5. As $T \to \infty$, $E[(Y_{T+1} - \widehat{Y}_{T,k}(1))(Y_{T+1} - \widehat{Y}_{T,k}(1))'] = \Sigma + o_p(1)$, where $\widehat{Y}_{T,k}(1)$ is the same-realization prediction and $\Sigma = E(\varepsilon_{T+1}\varepsilon_{T+1}^{\top})$.

隨機邊界模型: Cumulative distribution function 的應用

1. Consider a linear stochastic frontier model in the usual matrix form:

$$y = X\beta + \varepsilon,$$

where y and ε are $n \times 1$ vectors of observations on dependent variable and the random disturbance, respectively; X is an $n \times k$ matrix of observations on a constant term and $k-1$ regressors; and β is a $k \times 1$ vector of unknown regression coefficients to be estimated.

2. The error specification is:

$$\varepsilon = v + u,$$

where the elements of v are independently and identically distributed (iid) as $N(0, \sigma_v^2)$, and the elements of u are the absolute value of the variables which are iid as $N(0, \sigma_u^2)$.

3. We follow the reparameterization of Aigner et al. (1977):

$$\sigma^2 = \sigma_u^2 + \sigma_v^2, \qquad \lambda = \frac{\sigma_u}{\sigma_v}.$$

CDF: 二

1. The log likelihood function for the SFA model is:

$$L_0 = \frac{n}{2}\ln(2/\pi) - n\ln(\sigma) + \sum_{i=1}^{n}\ln\left[\Phi\left(\frac{\lambda}{\sigma}\varepsilon\right)\right] - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\varepsilon_i^2,$$

where $\varepsilon_i = y_i - x_i^\top\beta$, $x_i^\top$ is the i-th row of X, and Φ is the cumulative distribution function (cdf) of $N(0,1)$.

2. The maximum likelihood estimator is obtained by the maximization of the above log-likelihood function with respect to the parameter $(\beta^\top, \sigma_u, \sigma_v)$.

CDF: 三

1. When the dependent variable is censored, i.e.,

$$\begin{cases} y_i^* = x_i^\top \beta + \varepsilon_i, \quad i = 1, 2, \ldots, n; \\ y_i = y_i^*, \quad \text{if} \quad y_i^* > 0; \\ y_i = 0, \quad \text{if} \quad y_i^* \leq 0, \end{cases} \qquad (\textit{censored SFA})$$

   where $\varepsilon_i = v_i + u_i$.

2. We should not used the likelihood function of the standard SFA model to estimate the parameters of the model, because it does not take the presence of censored dependent variable into account.

CDF: 四

1. We should use Tobit MLE procedure for the stochastic frontier models with censored dependent variable.
2. The Tobit likelihood function for the censored SFA model is:

$$L_1 = \sum_1 \ln f(\varepsilon_i) + \sum_0 \ln F(-x_i^\top \beta), \qquad (\textit{Tobit likelihood})$$

where $f(\varepsilon_i)$ is the density function of ε_i, $F(-x_i^\top \beta)$ is the cumulative distribution function of ε_i from $-\infty$ to $-x_i^\top \beta$, $\sum_1$ denotes the sum over those i for which $y_i^* > 0$, and $\sum_0$ means the sum over those i for which $y_i^* \leq 0$.

CDF: 五

1. The uncensored part of the Tobit SFA model is easily implemented, because we note that

$$f(\varepsilon_i) = \frac{2}{\sigma}\,\phi\left(\frac{\varepsilon_i}{\sigma}\right)\Phi\left(\frac{\lambda}{\sigma}\,\varepsilon_i\right),$$

   where $\phi(.)$ denotes the density function of $N(0,1)$.

2. The major difficulty is to compute $F(-x_i^\top\beta)$:

$$F(-x_i^\top\beta) = \int_{-\infty}^{-x_i^\top\beta} f(\varepsilon_i)d\varepsilon_i, \quad \text{for} \quad y_i = 0.$$

CDF: 六

1. We know $F(-x_i^\top \beta)$ can be represented as:

$$F(-x_i^\top \beta) = \frac{2}{\sigma} I(-x_i^\top \beta) = \frac{2}{\sigma} \int_{-\infty}^{Q} \left(\int_{-\infty}^{a\varepsilon} \phi\left(\zeta\right) d\zeta \right) \phi\left(b\varepsilon\right) d\varepsilon,$$

where

$$a = \frac{\lambda}{\sigma}, \quad b = \frac{1}{\sigma}, \quad Q = -x_i^\top \beta.$$

2. We derive an approximate formula I_{app} for the component I.

CDF: 七

1. To derive the approximation, we define

$$Erf(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt = 2 \int_0^{\sqrt{2}z} \phi(t)\, dt,$$

and

$$sign(x) = \begin{cases} 1, & \text{if} \quad x > 0; \\ 0, & \text{if} \quad x = 0; \\ -1, & \text{if} \quad x < 0, \end{cases}$$

CDF: 八

1. **Theorem 1**. *Under* $(Q, a, b) \in$ *finite* R, $a \geq 0$, *and* $b \neq 0$, I *is approximated by* I_{app}:

$$I_{app} = \frac{e^{\frac{a^2 c_1^2}{4b^2 - 4a^2 c_2}} \left[1 - Erf\left(\frac{-ac_1 + \sqrt{2}Q\left(b^2 - a^2 c_2\right) \times sign(Q)}{2\sqrt{b^2 - a^2 c_2}}\right)\right]}{4\sqrt{b^2 - a^2 c_2}}$$

$$+ \frac{Erf\left(\frac{bQ}{\sqrt{2}}\right)}{2b} \times \frac{1 + sign\left(Q\right)}{2},$$

where $c_1 = -1.0950081470333$ *and* $c_2 = -0.75651138383854$.

CDF: 九

1. Since $a = \lambda/\sigma \geq 0$, we divide the derivation into two parts: for $(Q \geq 0, a \geq 0)$ and $(Q \leq 0, a \geq 0)$.
2. let us emphasize two equations, $(7.1.1)$ and $(7.4.32)$, of Abramowitz and Stegun (1970) for later applications:

$$
\begin{aligned}
Erf(z) &= \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt = 2 \int_0^{\sqrt{2}z} \phi(t)\, dt, \\
\int e^{-\left(kx^2 + 2mx + n\right)} dx &= \frac{1}{2}\sqrt{\frac{\pi}{k}} e^{\frac{m^2 - kn}{k}} Erf\left(\sqrt{k}x + \frac{m}{\sqrt{k}}\right) + C, \quad k \neq 0,
\end{aligned}
$$

where C denotes some finite constant.

CDF: 十

1. Given that $(Q, a, b) \in$ finite R, $b \neq 0$, $Erf(-x) = -Erf(x)$, and define $\varepsilon = \sqrt{2}\nu/a$, we have:

$$
\begin{aligned}
I_{a\geq 0} &= \frac{\sqrt{2}}{a}\int_{-\infty}^{\frac{a}{\sqrt{2}}Q}\left(\int_{-\infty}^{\sqrt{2}\nu}\phi(\zeta)\,d\zeta\right)\phi\left(\sqrt{2}\nu\frac{b}{a}\right)d\nu \\
&= \frac{\sqrt{2}}{2a}\int_{-\infty}^{\frac{a}{\sqrt{2}}Q}(1+Erf(\nu))\,\phi\left(\sqrt{2}\nu\frac{b}{a}\right)d\nu \\
&= \frac{\sqrt{2}}{2a}\int_{-\infty}^{0}(1+Erf(\nu))\,\phi\left(\sqrt{2}\nu\frac{b}{a}\right)d\nu \\
&\quad +\frac{\sqrt{2}}{2a}\int_{0}^{\frac{a}{\sqrt{2}}Q}(1+Erf(\nu))\,\phi\left(\sqrt{2}\nu\frac{b}{a}\right)d\nu.
\end{aligned}
$$

CDF: 十一

1. Note that $Erf(x)$ can be well approximated by a function, $g(x) = 1 - e^{c_1 x + c_2 x^2}$ for $x \geq 0$, where c_1 and c_2 are chosen to ensure that $g(x)$ is as close to $Erf(x)$ as possible. The choice of c_1 and c_2 is discussed above.

Accurancy of F_{app} in Computing CDF

$\lambda = 1$			Q	
Method	-16	-12	12	16
F_{app}	$1.3037e-105$	$1.9947e-61$	$1-3.5530e-33$	$1-2.778e-57$
AST	$4.0776e-115$	$3.1534e-66$	$1-3.6401e-39$	$1-1.7856e-75$
Exact	$4.0816e-115$	$3.1559e-66$	$1-3.5530e-33$	$1-1.2778e-57$

Notes: AST denotes the simulation method proposed by Amsler, Schmidt, and Tsay (2019).