Econometrics II - SOLUTIONS EXAM

Answer each question in separate sheets in three hours

1. *In the model*

$$y_t = \beta' \mathbf{z}_t + v_t, \quad t = 1, 2, \ldots$$

suppose that

$$v_t = \alpha v_{t-1} + e_t, \quad |\alpha| < 1,$$

where e_t *is* $WN(0, \sigma_e^2)$.

a. [40%] *Study the asymptotic properties of* $\widehat{\beta}_T$, *the OLSE of* β, *when:*

(1) *The regressors* z_t *are non-stochastic and bounded.*

(2) $z_t = (1, t)'$.

(3) *The regressors* z_t *are stationary and ergodic, and* $E[\mathbf{z}_t e_t] = 0$ *but* $E[\mathbf{z}_t e_{t-1}] \neq 0$.

Specify clearly any additional conditions you use for the analysis.

We have that

$$\widehat{\beta}_T - \beta = \left(\frac{1}{T} \sum_{t=1}^{T} \mathbf{z}_t \mathbf{z}_t' \right) \frac{1}{T} \sum_{t=1}^{T} \mathbf{z}_t v_t = \mathbf{M}_T^{-1} \mathbf{V}_T, \quad \text{say.}$$

In Case (1), we need to assume that $\mathbf{M}_T \to \mathbf{M} \ > 0$ as $T \to \infty$. Then we can check that

$$\begin{aligned} E \left\| \mathbf{V}_T \right\|^2 &= \frac{1}{T^2} \sum_t \sum_s \mathbf{z}_t' \mathbf{z}_s E\left[v_s v_t \right] \\ &\leq \frac{C}{T} \sup_t \left\| \mathbf{z}_t \right\|^2 \frac{1}{T} \sum_t \sum_s \left| \boldsymbol{\gamma}_v \left(t - s \right) \right| \\ &= O\left(T^{-1} \right) = o(1), \end{aligned}$$

so $\mathbf{V}_T \to_p 0$ and $\widehat{\beta}_T$ is consistent. For the asymptotic normality we need a CLT for $T^{1/2} \mathbf{V}_T$, for which we need further to assume that the e_t are IID or homoskedastic MD [and/or some UI conditions], so that

$$T^{1/2} \mathbf{V}_T \to_d N\left(0, \mathbf{S} \right)$$

where

$$\mathbf{S} = \sum_{j=-\infty}^{\infty} \mathbf{R}\left(j \right) \boldsymbol{\gamma}_v \left(j \right)$$

and $\mathbf{R}\left(-j \right) = \mathbf{R}\left(j \right), \ j = 0, 1, \ldots,$ and

$$\mathbf{R}\left(j \right) = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T-j} \mathbf{z}_t \mathbf{z}_{t+j}'.$$

Then

$$T^{1/2} \left(\widehat{\beta}_T - \beta \right) \to_d N\left(0, \mathbf{M}^{-1} \mathbf{S} \mathbf{M}^{-1} \right).$$

Alternatively we could have assumed that,

$$\mathbf{D}_T^{-1} \sum_{t=1}^{T-j} \mathbf{z}_t \mathbf{z}_{t+j}' \mathbf{D}_T^{-1} \to \mathbf{R}^{+}\left(j \right), \quad j = 0, 1, 2, \ldots$$

where

$$\mathbf{D}_T = diag\,\{d_{i,T}\}_{i=0}^{k}, \quad d_{i,T}^2 = \sum_{t=1}^{T} z_{i,t}^2$$

and for Grenander conditions to hold we only need that

$$d_{i,T} \to \infty, \quad \forall i \text{ as } \quad T \to \infty,$$

since by boundedness of $\mathbf{z}_t$ we have that

$$\lim_{T\to\infty} \max_{1\le t\le T} \frac{|z_{it}|}{d_i} = 0, \quad \forall i.$$

Then we can obtain that

$$\mathbf{D}_T T \mathbf{V}_T \to_d N\left(0, \mathbf{S}^+\right)$$

where

$$\mathbf{S}^+ = \sum_{j=-\infty}^{\infty} \mathbf{R}^+(j)\,\boldsymbol{\gamma}_v(j),$$

and finally

$$\mathbf{D}_T\left(\widehat{\beta}_T - \beta\right) \to_d N\left(0, \mathbf{R}^+(0)^{-1}\mathbf{S}^+\mathbf{R}^+(0)^{-1}\right).$$

Case (2) is a particular case of this situation, with

$$\mathbf{D}_T = \begin{pmatrix} \left(\sum_t 1\right)^{1/2} & 0 \\ 0 & \left(\sum_t t^2\right)^{1/2} \end{pmatrix} \sim \begin{pmatrix} T^{1/2} & 0 \\ 0 & \left(\frac{1}{3}T^3\right)^{1/2} \end{pmatrix}.$$

In case (3) OLS is inconsistent if $\alpha \neq 0$, because $\mathrm{E}[\mathbf{z}_t v_t] \neq 0$ since

$$E[\mathbf{z}_t v_t] = \alpha E[\mathbf{z}_t e_{t-1}] + E[\mathbf{z}_t e_t] = \alpha E[\mathbf{z}_t e_{t-1}] \neq 0.$$

b. [30%] *Suppose that α is known, and consider*

$$\widetilde{\beta}_T = \left(\sum_{t=2}^{T}(\mathbf{z}_t - \alpha\mathbf{z}_{t-1})(\mathbf{z}_t - \alpha\mathbf{z}_{t-1})'\right)^{-1}\left(\sum_{t=2}^{T}(\mathbf{z}_t - \alpha\mathbf{z}_{t-1})(y_t - \alpha y_{t-1})\right).$$

Find the asymptotic distribution of $T^{1/2}(\widetilde{\beta}_T - \beta)$ under the previous conditions, specifying all the conditions used.

We now have that

$$T^{1/2}(\widetilde{\beta}_T - \beta) = \left(\sum_{t=2}^{T}\mathbf{z}_t^*\mathbf{z}_t^{*\prime}\right)^{-1}\sum_{t=2}^{T}\mathbf{z}_t^* e_t$$

where $\mathbf{z}_t^* = \mathbf{z}_t - \alpha\mathbf{z}_{t-1}$.

The analysis for case (1) is similar as before, but now we have that

$$T^{1/2}\left(\widetilde{\beta}_T - \beta\right) \to_d N\left(0, \sigma^2\mathbf{M}^{*-1}\right),$$

where we assume that

$$\mathbf{M}_T^* := \frac{1}{T}\sum_{t=2}^{T}\mathbf{z}_t^*\mathbf{z}_t^{*\prime} \to \mathbf{M}^* > 0$$

as $T \to \infty$.

In case (2) we would obtain similar results with the appropriate normalization, that for the linear trend is not $T^{1/2}$, but $T^{3/2}$,

$$\mathbf{D}_T^* \left(\widetilde{\beta}_T - \beta \right) \rightarrow_d N \left(0, \sigma^2 \mathbf{R}^* \left(0 \right)^{-1} \right).$$

where, $j = 0, 1, 2, \ldots$

$$\mathbf{R}^* \left(j \right) = \lim_{T \rightarrow \infty} \mathbf{D}_T^{*-1} \sum_{t=1}^{T-j} \mathbf{z}_t^* \mathbf{z}_{t+j}^{*\prime} \mathbf{D}_T^{*-1}$$

and

$$\mathbf{D}_T^* = diag \left\{ d_{i,T}^* \right\}_{i=0}^k, \quad d_{i,T}^{*2} = \sum_{t=1}^T z_{i,t}^{*2}.$$

In Case (3) we have to check whether $\mathrm{E}[\mathbf{z}_t^* e_t] = 0$:

$$E \left[\mathbf{z}_t^* e_t \right] = \alpha E \left[\mathbf{z}_{t-1} e_t \right] + E \left[\mathbf{z}_t e_t \right] = 0,$$

because $E \left[\mathbf{z}_{t-1} e_t \right] = 0$. In this case, under appropriate conditions on the distribution of $\mathbf{z}_t^* e_t$ (ergodic MD) we have a similar result to (1), where now $\mathbf{M}^* := E \left[\mathbf{z}_t^* \mathbf{z}_t^{*\prime} \right]$, if $E \left[e_t^2 | \mathbf{z}_t^* \right] = \sigma^2$ and if $E \left[e_t e_j | \mathbf{z}_t^*, \mathbf{z}_j^* \right] = 0$, $t \neq j$. Otherwise, if $\mathbf{z}_t^* e_t$ is not a MD but satisfies a CLT, we have that

$$T^{1/2} \left(\widetilde{\beta}_T - \beta \right) \rightarrow_d N \left(0, \mathbf{M}^{*-1} \mathbf{W} \mathbf{M}^{*-1} \right)$$

where

$$\mathbf{W} = \sum_{j=-\infty}^{\infty} E \left[e_t e_{t+j} \mathbf{z}_t^* \mathbf{z}_{t+j}^{*\prime} \right].$$

c. [30%] *Consider now the following estimate of α,*

$$\widehat{\alpha}_T = \left(\sum_{t=2}^T \widehat{v}_t^{\,2} \right)^{-1} \sum_{t=2}^T \widehat{v}_t \widehat{v}_{t-1}, \quad \widehat{v}_t = y_t - \breve{\beta}_T^{\prime} \mathbf{z}_t,$$

where $\breve{\beta}_T$ is a consistent estimate of β, and let $\bar{\beta}_T$ be the estimate obtained when replacing α by $\widehat{\alpha}_T$ in $\widetilde{\beta}_T$.

Study the relationship between $\widetilde{\beta}_T$ and $\bar{\beta}_T$ under (1), (2) and (3). Then deduce if the above estimates of α and of β are asymptotically related or not in each case.

The first issue is the convergence rate of $\widehat{\alpha}_T$, which depends on that of the initial estimate of β, $\breve{\beta}_T$ (this can not be the OLSE in case (3): IV's estimates can be used instead). If the latter is $T^{1/2}$ consistent (or faster), then $\widehat{\alpha}_T$ will be so as well.

Then we have that

$$\begin{aligned}
& T^{1/2} \left(\widetilde{\beta}_T - \bar{\beta}_T \right) \\
= & \left\{ \left(\frac{1}{T} \sum_{t=2}^T (\mathbf{z}_t - \alpha \mathbf{z}_{t-1})(\mathbf{z}_t - \alpha \mathbf{z}_{t-1})' \right)^{-1} - \left(\frac{1}{T} \sum_{t=2}^T (\mathbf{z}_t - \widehat{\alpha}_T \mathbf{z}_{t-1})(\mathbf{z}_t - \widehat{\alpha}_T \mathbf{z}_{t-1})' \right)^{-1} \right\} \\
& \times \left(\frac{1}{T^{1/2}} \sum_{t=2}^T \overbrace{(\mathbf{z}_t - \alpha \mathbf{z}_{t-1})}^{=\mathbf{z}_t^*} \overbrace{(v_t - \alpha v_{t-1})}^{=e_t} \right) \\
& + \left(\frac{1}{T} \sum_{t=2}^T (\mathbf{z}_t - \widehat{\alpha}_T \mathbf{z}_{t-1})(\mathbf{z}_t - \widehat{\alpha}_T \mathbf{z}_{t-1})' \right)^{-1} \\
& \times \left\{ \frac{1}{T^{1/2}} \sum_{t=2}^T (\mathbf{z}_t - \alpha \mathbf{z}_{t-1})(v_t - \alpha v_{t-1}) - \frac{1}{T^{1/2}} \sum_{t=2}^T (\mathbf{z}_t - \widehat{\alpha}_T \mathbf{z}_{t-1})(v_t - \widehat{\alpha}_T v_{t-1}) \right\}.
\end{aligned}$$

The first term in braces can be showed to be $o_p(1)$ if $\widehat{\alpha}_T - \alpha = o_p(1)$, and the term $T^{-1/2}\sum_{t=2}^{T}\mathbf{z}_t^* e_t$ is $O_p(1)$, so the first term is asymptotically negligible. The leading term in the second factor converges to $\mathbf{M}^{*-1}$. The key is the second term in braces, which is equal to

$$\begin{aligned}
& \frac{1}{T^{1/2}}\sum_{t=2}^{T}(\mathbf{z}_t - \alpha\mathbf{z}_{t-1})(v_t - \alpha v_{t-1}) - \frac{1}{T^{1/2}}\sum_{t=2}^{T}(\mathbf{z}_t - \widehat{\alpha}_T\mathbf{z}_{t-1})(v_t - \widehat{\alpha}_T v_{t-1}) \\
= \; & \frac{1}{T^{1/2}}(\widehat{\alpha}_T - \alpha)\sum_{t=2}^{T}(\mathbf{z}_t - \alpha\mathbf{z}_{t-1})v_{t-1} - \frac{1}{T^{1/2}}(\widehat{\alpha}_T - \alpha)\sum_{t=2}^{T}\mathbf{z}_{t-1}(v_t - \widehat{\alpha}_T v_{t-1}) \\
= \; & T^{1/2}(\widehat{\alpha}_T - \alpha)\left\{E\left[\mathbf{z}_t^* v_{t-1}\right] + o_p(1)\right\} - T^{1/2}(\widehat{\alpha}_T - \alpha)\left\{E\left[\mathbf{z}_{t-1}e_t\right] + o_p(1)\right\}.
\end{aligned}$$

In Case (1) we have that $E\left[\mathbf{z}_t^* v_{t-1}\right], E\left[\mathbf{z}_{t-1}e_t\right] = o_p(1)$, so $T^{1/2}\left(\widetilde{\beta}_T - \bar{\beta}_T\right) = o_p(1)$ and both estimates have the same asymptotic distribution, and therefore estimation of β and of α are asymptotically independent.

In Case (2) the same conclusion applies (even despite that the estimates of some elements of β have faster convergence rate).

If in Case (1) the convergence of $\widetilde{\beta}_T$ is slower that $T^{1/2}$, then estimation of α is affected if $\breve{\beta}_T$ has simila properties, but $\widetilde{\beta}_T$ and $\bar{\beta}_T$ will be asymptotically equivalent (at the appropriate rate), so estimation of β is not affected by that of α (but not the way round)

In Case (3) we have that $E\left[\mathbf{z}_{t-1}e_t\right] = 0,$ but $E\left[\mathbf{z}_t^* v_{t-1}\right] \neq 0$, because $E\left[\mathbf{z}_t e_{t-1}\right] \neq 0$. Therefore $T^{1/2}\left(\widetilde{\beta}_T - \bar{\beta}_T\right) = O_p(1)$, and in fact we can find that

$$T^{1/2}\left(\widetilde{\beta}_T - \bar{\beta}_T\right) = T^{1/2}(\widehat{\alpha}_T - \alpha)\,E\left[\mathbf{z}_t^*\mathbf{z}_t^{*\prime}\right]^{-1}E\left[\mathbf{z}_{t-1}e_t\right] + o_p(1),$$

which shows that estimation of β and of α are not independent, and their asymptotic distributions are correlated, since under the given assumptions the $\mathbf{z}_t$ are not strictly exogeneous, only weakly exogenous.

2. *Let the scalar dependent variable y be a binary variable, $y \in \{0,1\}$, and let z be some explanatory variables. The conditional probability of y given the vector of variables z is given by*

$$\begin{aligned}\Pr(y=1|\mathbf{z}) &= \Lambda(\mathbf{z}'\boldsymbol{\theta}_0),\\ \Pr(y=0|\mathbf{z}) &= 1-\Lambda(\mathbf{z}'\boldsymbol{\theta}_0),\end{aligned}$$

where $\Lambda(\cdot)$ is the logit cumulative distribution function,

$$\Lambda(x) = \frac{\exp(x)}{1+\exp(x)}$$

and θ_0 is a vector of parameters.

a. [15%] *Find $E(y|\mathbf{z})$ and $V(y|\mathbf{z})$. Give the objective function for the Nonlinear Least Squares (NLS) estimates of θ_0 given a sample of size n of independent observations of $(y,\mathbf{z})$.*

$$\begin{aligned}E(y|\mathbf{z}) &= \Lambda(\mathbf{z}'\boldsymbol{\theta}_0)\\ V(y|\mathbf{z}) &= \Lambda(\mathbf{z}'\boldsymbol{\theta}_0)(1-\Lambda(\mathbf{z}'\boldsymbol{\theta}_0))\\ Q_n(\boldsymbol{\theta}) &= \frac{1}{2}E_n\left[(y-\Lambda(\mathbf{z}'\boldsymbol{\theta}))^2\right].\end{aligned}$$

b. [35%] *Assuming consistency, deduce the asymptotic distribution of the NLS estimates $\widehat{\theta}_n$ and obtain the Gauss-Newton recursion to approximate the NLS estimates given an initial value of θ. Propose a consistent estimate of the asymptotic variance of the NLE $\widehat{\theta}_n$. State any further assumption you use.*

We have that

$$\begin{aligned}\frac{\partial}{\partial\theta}Q_n(\boldsymbol{\theta}) &= -E_n\left[(y-\Lambda(\mathbf{z}'\boldsymbol{\theta}))\,\lambda(\mathbf{z}'\boldsymbol{\theta})\mathbf{z}\right]\\ \frac{\partial^2}{\partial\theta\partial\theta'}Q_n(\boldsymbol{\theta}) &= E_n\left[\lambda(\mathbf{z}'\boldsymbol{\theta})^2\mathbf{z}\mathbf{z}'\right] - E_n\left[(y-\Lambda(\mathbf{z}'\boldsymbol{\theta}))\,\dot{\lambda}(\mathbf{z}'\boldsymbol{\theta})\mathbf{z}\mathbf{z}'\right].\end{aligned}$$

Therefore, $u = u(\boldsymbol{\theta}_0) := y - \Lambda(\mathbf{z}'\boldsymbol{\theta}_0)$,

$$\begin{aligned}n^{1/2}\frac{\partial}{\partial\theta}Q_n(\boldsymbol{\theta}_0) &= -n^{1/2}E_n\left[u\lambda(\mathbf{z}'\boldsymbol{\theta})\mathbf{z}\right]\\ &\to_d N\left(0, E\left[\Lambda(\mathbf{z}'\boldsymbol{\theta}_0)(1-\Lambda(\mathbf{z}'\boldsymbol{\theta}_0))\,\lambda(\mathbf{z}'\boldsymbol{\theta}_0)^2\mathbf{z}\mathbf{z}'\right]\right)\\ &:= N(0,\mathbf{D})\end{aligned}$$

and, if $\bar{\boldsymbol{\theta}}_n \to_p \boldsymbol{\theta}_0$,

$$\begin{aligned}\frac{\partial^2}{\partial\theta\partial\theta'}Q_n(\bar{\boldsymbol{\theta}}_n) &= E_n\left[\lambda(\mathbf{z}'\bar{\boldsymbol{\theta}}_n)^2\mathbf{z}\mathbf{z}'\right] - E_n\left[u(\bar{\boldsymbol{\theta}}_n)\,\dot{\lambda}(\mathbf{z}'\bar{\boldsymbol{\theta}}_n)\mathbf{z}\mathbf{z}'\right]\\ &\to_p E_n\left[\lambda(\mathbf{z}'\boldsymbol{\theta}_0)^2\mathbf{z}\mathbf{z}'\right] - E_n\left[u\dot{\lambda}(\mathbf{z}'\boldsymbol{\theta}_0)\mathbf{z}\mathbf{z}'\right]\\ &\to_p E\left[\lambda(\mathbf{z}'\boldsymbol{\theta}_0)^2\mathbf{z}\mathbf{z}'\right] := \mathbf{E},\end{aligned}$$

so that we obtaint the usual result after linearizing $Q_n\left(\hat{\boldsymbol{\theta}}_n\right)$ around $\boldsymbol{\theta}_0$, if the true value $\boldsymbol{\theta}_0$ is in the interior of the allowed set Θ,

$$n^{1/2}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) \to_d N\left(0, \mathbf{E}^{-1}\mathbf{D}\mathbf{E}^{-1}\right).$$

We have that

$$\begin{aligned}\hat{\mathbf{D}}_n &= E\left[\Lambda(\mathbf{z}'\hat{\boldsymbol{\theta}}_n)\left(1-\Lambda(\mathbf{z}'\hat{\boldsymbol{\theta}}_n)\right)\lambda(\mathbf{z}'\hat{\boldsymbol{\theta}}_n)^2\mathbf{z}\mathbf{z}'\right] \to_p \mathbf{D}.\\ \hat{\mathbf{E}}_n &= E_n\left[\lambda(\mathbf{z}'\hat{\boldsymbol{\theta}}_n)^2\mathbf{z}\mathbf{z}'\right] \to_p \mathbf{E}.\end{aligned}$$

Gauss-Newton recursion:

$$\hat{\boldsymbol{\theta}}_n^{(k+1)} = \hat{\boldsymbol{\theta}}_n^{(k)} - \mathbf{G}_n^{-1}\left(\hat{\boldsymbol{\theta}}_n^{(k)}\right)\mathbf{g}_n\left(\hat{\boldsymbol{\theta}}_n^{(k)}\right)$$

where

$$\begin{aligned}\mathbf{g}_n(\boldsymbol{\theta}) &= -E_n\left[(y-\Lambda(\mathbf{z}'\boldsymbol{\theta}))\,\lambda(\mathbf{z}'\boldsymbol{\theta})\mathbf{z}\right]\\ \mathbf{G}_n(\boldsymbol{\theta}) &= E_n\left[\lambda(\mathbf{z}'\boldsymbol{\theta})^2\mathbf{z}\mathbf{z}'\right].\end{aligned}$$

c. [30%] *Consider alternatively the estimate obtained through this recursion,*

$$\widetilde{\boldsymbol{\theta}}_n = \widehat{\boldsymbol{\theta}}_n - \mathbf{F}_n(\widehat{\boldsymbol{\theta}}_n)^{-1}\mathbf{f}_n(\widehat{\boldsymbol{\theta}}_n)$$

where $\widehat{\theta}_n$ *is the NLS estimate and*

$$\begin{aligned}\mathbf{f}_n(\boldsymbol{\theta}) &:= -\mathbb{E}_n\left[\frac{\lambda(\mathbf{z}'\boldsymbol{\theta})\{y-\Lambda(\mathbf{z}'\boldsymbol{\theta})\}}{\Lambda(\mathbf{z}'\boldsymbol{\theta})(1-\Lambda(\mathbf{z}'\boldsymbol{\theta}))}\mathbf{z}\right]\\ \mathbf{F}_n(\boldsymbol{\theta}) &:= \mathbb{E}_n\left[\frac{\{\lambda(\mathbf{z}'\boldsymbol{\theta})\}^2}{\Lambda(\mathbf{z}'\boldsymbol{\theta})(1-\Lambda(\mathbf{z}'\boldsymbol{\theta}))}\mathbf{z}\mathbf{z}'\right]\end{aligned}$$

with $\lambda(x) = (d/dx)\Lambda(x) = \exp(x)/(1+\exp(x))^2$. *Compare these iterations with Gauss-Newton method.*

$\widetilde{\boldsymbol{\theta}}_n$ corrects for the conditional variance of u, so is a sort of Weighted or Generalized NLS.

Which objective function is minimizing $\widetilde{\theta}_n$ *asymptotically?*

$$Q_n^*(\boldsymbol{\theta}) = \frac{1}{2}E_n\left[\frac{(y-\Lambda(\mathbf{z}'\boldsymbol{\theta}))^2}{\Lambda(\mathbf{z}'\boldsymbol{\theta}_0)(1-\Lambda(\mathbf{z}'\boldsymbol{\theta}_0))}\right].$$

Deduce the asymptotic distribution of $\widetilde{\theta}_n$ *using that*

$$n^{1/2}\left(\widetilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) = -n^{1/2}\mathbf{F}_n(\widehat{\boldsymbol{\theta}}_n)^{-1}\mathbf{f}_n(\boldsymbol{\theta}_0) + o_p(1).$$

We have that

$$\begin{aligned}n^{1/2}\mathbf{f}_n(\boldsymbol{\theta}_0) &\to_d N\left(0, \mathbb{E}\left[\frac{\lambda^2(\mathbf{z}'\boldsymbol{\theta}_0)}{\Lambda(\mathbf{z}'\boldsymbol{\theta})(1-\Lambda(\mathbf{z}'\boldsymbol{\theta}))}\mathbf{z}\mathbf{z}'\right]\right)\\ \mathbf{F}_n(\widehat{\boldsymbol{\theta}}_n) &\to_p \mathbf{F}_n(\boldsymbol{\theta_0})\\ &\to_p \mathbb{E}\left[\frac{\lambda(\mathbf{z}'\boldsymbol{\theta}_0)^2}{\Lambda(\mathbf{z}'\boldsymbol{\theta}_0)(1-\Lambda(\mathbf{z}'\boldsymbol{\theta}_0))}\mathbf{z}\mathbf{z}'\right]\end{aligned}$$

and therefore

$$n^{1/2}\left(\widetilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) \to_d N\left(0, \mathbb{E}\left[\frac{\lambda^2(\mathbf{z}'\boldsymbol{\theta}_0)}{\Lambda(\mathbf{z}'\boldsymbol{\theta}_0)(1-\Lambda(\mathbf{z}'\boldsymbol{\theta}_0))}\mathbf{z}\mathbf{z}'\right]^{-1}\right),$$

displaying the usual result for efficient estimates. In fact, it can be showed that this is also the asymptotic distribution of the MLE estimate, and that $\widetilde{\boldsymbol{\theta}}_n$ is an approximation to it.

Can you say anything about the efficiency of $\widetilde{\theta}_n$ *and* $\widehat{\theta}_n$*?*

d. [20%] *Suppose that* z *contains the constant term, so* $z'\theta = \theta_1 + z_2'\theta_2$. *Consider the null hypothesis that the coefficients of all the other variables in* z *are zero,* $H_0 : \theta_{2,0} = 0$. *Which will be your LS estimate of* $\Pr(y = 1|z)$ *under the null?*

In the case that $\mathbf{z}'\boldsymbol{\theta} = \theta_1$ we have that our model is simply that $E[y] = E[y|\mathbf{z}] = \Pr(y = 1|\mathbf{z}) = \Lambda(\theta_1) := p$, and the LS of a model with only an intercept is the sample mean, $\hat{p} = \bar{y}_n$, and the restricted estimate is defined by $\Lambda\left(\hat{\boldsymbol{\theta}}^*_{1n}\right) = \bar{y}_n$.

Propose a LM test for H_0.

We need to calculate, where the restricted estimates are $\hat{\boldsymbol{\theta}}^*_n = \left(\hat{\boldsymbol{\theta}}^*_{1n}, \mathbf{0}'\right)'$, [here I work with the general case of vector $\mathbf{z}_2$]

$$\begin{aligned}\frac{\partial}{\partial\theta_2}Q_n\left(\hat{\boldsymbol{\theta}}^*_n\right) &= -E_n\left[\left(y - \Lambda(\hat{\boldsymbol{\theta}}^*_{1n})\right)\lambda(\hat{\boldsymbol{\theta}}^*_{1n})\mathbf{z}_2\right] \\ &= -E_n\left[\left(y - \bar{y}_n\right)\lambda(\hat{\boldsymbol{\theta}}^*_{1n})\mathbf{z}_2\right],\end{aligned}$$

which looks like a covariance between the errors of the restricted model, $y - \bar{y}_n$, and the new regressors, $\mathbf{z}_2$.

and we have that the LM statistic is given by (using Q_n)

$$\begin{aligned}LM_n &= n\frac{\partial}{\partial\theta'}Q_n\left(\hat{\boldsymbol{\theta}}^*_n\right)\mathbf{H}_n\left(\hat{\boldsymbol{\theta}}^*_n\right)^{-1}\frac{\partial}{\partial\theta}Q_n\left(\hat{\boldsymbol{\theta}}^*_n\right) \\ &= n\frac{\partial}{\partial\theta_2'}Q_n\left(\hat{\boldsymbol{\theta}}^*_n\right)\left\{\mathbf{H}_n\left(\hat{\boldsymbol{\theta}}^*_n\right)^{-1}\right\}_{2,2}\frac{\partial}{\partial\theta_2}Q_n\left(\hat{\boldsymbol{\theta}}^*_n\right)\end{aligned}$$

where, for example,

$$\mathbf{H}_n(\boldsymbol{\theta}) = \frac{\partial^2}{\partial\theta\partial\theta'}Q_n(\boldsymbol{\theta}) = E_n\left[\lambda(\mathbf{z}'\boldsymbol{\theta})^2\mathbf{z}\mathbf{z}'\right] - E_n\left[\left(y - \Lambda(\mathbf{z}'\boldsymbol{\theta})\right)\dot{\lambda}(\mathbf{z}'\boldsymbol{\theta})\mathbf{z}\mathbf{z}'\right].$$

The asymptotic distribution is χ^2_k where k is the dimension of z_2.

3. *Consider the following system, where the first equation is a labour supply function for working, married women, and the second equation is a wage offer function, with equilibrium conditions imposed,*

$$\begin{aligned} hours &= \gamma_{12}\log(wage) + \delta_{10} + \delta_{11}educ + \delta_{12}age \\ &\quad + \delta_{13}kidslt6 + \delta_{14}kidsge6 + \delta_{15}mwifeinc + u_1 \\ \log(wage) &= \gamma_{21}hours + \delta_{20} + \delta_{21}educ + \delta_{22}exper + \delta_{23}exper^2 + u_2, \end{aligned}$$

where $kidstl6$ *is number of children less than 6,* $kidsge6$ *is the number of children between 6 and 18 and* $mwifeinc$ *is income other than the woman's labor income. We assume that* u_1 *and* u_2 *have zero mean conditional on* $educ$, age, $kidslt6$, $kidsge6$, $nwifeinc$ *and* $exper$.

a. [30%] *Study the identification of the labor supply function when* $exper$ *and* $exper^2$ *have not direct effect on current annual hours (*$hours$*). Is this equation no, just or over-identified?*

It is over-identified as far as $\delta_{22} \neq 0$ and $\delta_{23} \neq 0$:

$$\mathbf{R}_1 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

which is of rank $J_1 = 2 > G - 1$ in this case, and

$$\mathbf{R}_1\mathbf{B} = \begin{pmatrix} 0 & \delta_{22} \\ 0 & \delta_{23} \end{pmatrix}$$

which is of rank $1 = G - 1$.

Is the wage offer function identified with the exclusion restrictions imposed?

Yes because we have many potential IV's available: age, $kidslt6, kidsge6, mwifeinc$.

b. [50%] *We obtain the following output from Eviews.*

(1) Dependent Variable: HOURS

Method: Two-Stage Least Squares

Included observations: 428 after adjusting endpoints

Instrument list: EDUC AGE KIDSLT6 KIDSGE6 NWIFEINC EXPER EXPER^2

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2432.198	594.1719	4.093425	0.0001
LWAGE	1544.819	480.7387	3.213426	0.0014
EDUC	-177.4490	58.14260	-3.051961	0.0024
AGE	-10.78409	9.577347	-1.125999	0.2608
KIDSLT6	-210.8339	176.9340	-1.191596	0.2341
KIDSGE6	-47.55708	56.91786	-0.835539	0.4039
NWIFEINC	-9.249121	6.481116	-1.427088	0.1543
Prob(F-statistic)	0.002692			

(2) Dependent Variable: HOURS

Method: Two-Stage Least Squares

Included observations: 428 after adjusting endpoints

Instrument list: EDUC AGE KIDSLT6 KIDSGE6 NWIFEINC EXPER

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2478.435	655.2070	3.782674	0.0002
LWAGE	1772.323	594.1850	2.982781	0.0030
EDUC	-201.1870	69.91013	-2.877795	0.0042
AGE	-11.22885	10.53692	-1.065668	0.2872
KIDSLT6	-191.6588	195.7609	-0.979046	0.3281
KIDSGE6	-37.73248	63.63485	-0.592953	0.5535
NWIFEINC	-9.977747	7.174493	-1.390725	0.1650

Prob(F-statistic) 0.009204

(3) Estimation Method: Iterative Three-Stage Least Squares

Included observations: 428

Instruments: EDUC AGE KIDSLT6 KIDSGE6 NWIFEINC EXPER EXPER^2 C

Convergence achieved after: 5 weight matricies, 6 total coef iterations

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	2511.719	566.1989	4.436106	0.0000
C(2)	1689.389	455.0691	3.712378	0.0002
C(3)	-207.6063	54.69267	-3.795870	0.0002
C(4)	-12.42447	8.683959	-1.430738	0.1529
C(5)	-199.5608	139.0501	-1.435172	0.1516
C(6)	-48.71302	36.00547	-1.352934	0.1764
C(7)	1.259363	3.309783	0.380497	0.7037
C(8)	-0.705146	0.308275	-2.287395	0.0224
C(9)	0.000201	0.000213	0.943669	0.3456
C(10)	0.112975	0.015321	7.373960	0.0000
C(11)	0.020856	0.014085	1.480770	0.1390
C(12)	-0.000293	0.000243	-1.207872	0.2274

Determinant residual covariance 98640.45

Equation 1: HOURS = C(1)+C(2)* LWAGE+C(3)* EDUC +C(4)*AGE +C(5)*KIDSLT6 +C(6)*KIDSGE6 +C(7)*NWIFEINC

Equation 2: LWAGE =C(8) +C(9)*HOURS+C(10)* EDUC +C(11)*EXPER +C(12)* EXPER^2

(4) Estimation Method: Generalized Method of Moments

Included observations: 428

Instruments: EDUC AGE KIDSLT6 KIDSGE6 NWIFEINC EXPER EXPER^2 C

White Covariance

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	2688.763	527.2528	5.099570	0.0000
C(2)	1937.325	492.1760	3.936245	0.0001
C(3)	-230.8447	54.08686	-4.268037	0.0000
C(4)	-15.31151	9.043993	-1.693003	0.0908
C(5)	-231.0814	157.5143	-1.467050	0.1427
C(6)	-52.79678	38.84534	-1.359154	0.1745
C(7)	-1.782789	3.457134	-0.515684	0.6062
C(8)	-0.559507	0.364794	-1.533763	0.125
C(9)	0.000106	0.000247	0.430531	0.6669
C(10)	0.111260	0.014364	7.745982	0.0000
C(11)	0.020723	0.014140	1.465522	0.1432
C(12)	-0.000261	0.000268	-0.975978	0.3294

Determinant residual covariance 143243.9

J-statistic 0.053546

Equation 1: HOURS = C(1)+C(2)* LWAGE+C(3)* EDUC +C(4)*AGE +C(5)*KIDSLT6 +C(6)*KIDSGE6 +C(7)*NWIFEINC

Equation 2: LWAGE =C(8) +C(9)*HOURS+C(10)* EDUC +C(11)*EXPER +C(12)* EXPER^2

(5) Dependent Variable: LWAGE

Method: Least Squares

Included observations: 428 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.903705	0.267473	3.378678	0.0008
EXPER	0.046301	0.014007	3.305528	0.0010
EXPER^2	-0.000867	0.000421	-2.057972	0.0402
AGE	-0.007040	0.005655	-1.244817	0.2139
KIDSLT6	0.015490	0.092304	0.167816	0.8668
KIDSGE6	-0.041348	0.029031	-1.424277	0.1551
NWIFEINC	0.012434	0.003316	3.749734	0.0002

R-squared	0.076983	Mean dependent var	1.190173
Log likelihood	-450.9589	F-statistic	5.852191
Durbin-Watson stat	1.940720	Prob(F-statistic)	0.000007

(6) Dependent Variable: LWAGE

Method: Least Squares

Included observations: 428 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.357997	0.318296	-1.124729	0.2613
EDUC	0.099884	0.015097	6.615970	0.0000
AGE	-0.003520	0.005415	-0.650176	0.5159
KIDSLT6	-0.055873	0.088603	-0.630591	0.5287
KIDSGE6	-0.017648	0.027891	-0.632765	0.5272
NWIFEINC	0.005694	0.003320	1.715373	0.0870
EXPER	0.040710	0.013372	3.044344	0.0025
EXPER^2	-0.000747	0.000402	-1.860055	0.0636

R-squared	0.164098	Mean dependent var	1.190173
Log likelihood	-429.7438	F-statistic	11.77879
Durbin-Watson stat	1.966359	Prob(F-statistic)	0.000000

i. *Does the data confirm your identification analysis for the labor supply function in a.?*
The key is to check identification is the joint significance of *exper* and *exper*^{2} in the RF for *lwage*, using a Wald test,

$$W = n\frac{R_u^2 - R_r^2}{1 - R_u^2} = 428\frac{0.164098 - 0.131258}{1 - 0.076983} = 15.228,$$

which is significative compared to a χ_2^2 (though *exper*^{2} is not significative). Note that 2SLS and 3SLS/GMM fitted models have to be taken with care because they need the equation or the system to be identified to the t- and F-tests to be valid, but they also display some doubts. In principle adding instruments makes a difference in 2SLS, so under the assumption of valid estimates, it could be some room to imagine overidentification (see also the J test below; also a Hausman test could be used, but not OLS estimation to compare is available).

ii. *Why do you expect that estimates in (1) to be more efficient than those of (2)?*
Because we are using more instruments (overidentifying restrictions).
Is always this the case?
Not always: e.g. when in the RF for *lwage*, *exper*^{2} is not significative ((6) indicates that this might be the case). In principle we don't have information about the conditional variances to design more efficient estimates with optimal instruments.

iii. *Why do you think that in (3) are needed iterations?*
Because we can use estimates of Ω from intial 3SLS iterations, not just from e.g. initial 2SLS.

iv. *Which indicates the J statistic in (4) [usual definitions requires to multiply this number given by Eviews by n]? Give an example where J should be zero.*
This is a test of specification of the model (of the moment conditions), after multiplication by n:

$$n * J = 428 * 0.053546 = 22.918,$$

it should be distributed as χ_q^2, where q is the number of overidentifying restrictions. In this case we have that the first equation has 1 o.r. and the second equation has 3 o.r., so we should compare with a χ_4^2 : it appears to be significative, and some specification problem could be present.

v. *Which estimates are more efficient asymptotically, those in (3) or (4)? Specify clearly your assumptions.*

In principle (efficient) GMM estimates are always efficient given the moment conditions (which fix the instruments to be used). But 3SLS can also be efficient given conditional homoskedasticity conditions.

c. [10%] *Discuss identification of the first equation when we add on the rhs two new terms:* $\gamma_{13}\left(\log(wage)\right)^2 + \gamma_{14}\left(\log(wage)\right)^3$.

In this case we don't have enough linear instruments: we should rely on nonlinear transformations of the IV's, which provide valid instruments for the quadratic and cubic terms in the equation. Even if $\gamma_{13} = \gamma_{14} = 0$, the equation remains identified because the reduced forms for $\left(\log(wage)\right)^2$ and $\left(\log(wage)\right)^3$ contains cuadratic and cubic terms of the $\mathbf{x}'s$ and we have two linear instruments for $\log(wage)$.

d. [10%] *Assume that educ is endogenous in the second equation, but exogenous in the first. Provide conditions which guarantee the identification of the system.*

We need to provide a RF for *educ*, but at least in principle we have enough instruments $(age, kidslt6, kidsge6, mwifeinc)$ from the first equation to instrumentalize both *educ* and *hours*, so that the respective linear projections are not collinear.

$\chi^2_{(1)}$	$\chi^2_{(2)}$	$\chi^2_{(3)}$	$\chi^2_{(4)}$
$\Pr\left(\chi^2_{(1)} > 6.63\right) = 0.01$	$\Pr\left(\chi^2_{(2)} > 9.21\right) = 0.01$	$\Pr\left(\chi^2_{(3)} > 11.34\right) = 0.01$	$\Pr\left(\chi^2_{(4)} > 13.28\right) = 0.01$
$\Pr\left(\chi^2_{(1)} > 3.84\right) = 0.05$	$\Pr\left(\chi^2_{(2)} > 5.99\right) = 0.05$	$\Pr\left(\chi^2_{(3)} > 7.81\right) = 0.05$	$\Pr\left(\chi^2_{(4)} > 9.49\right) = 0.05$
$\Pr\left(\chi^2_{(1)} > 2.71\right) = 0.10$	$\Pr\left(\chi^2_{(2)} > 4.61\right) = 0.10$	$\Pr\left(\chi^2_{(3)} > 6.25\right) = 0.10$	$\Pr\left(\chi^2_{(4)} > 7.78\right) = 0.10$