Indirect Inference

Thierry Kamionka

2021-2022

Let us consider that the true model is a moving average

$$y_t = \epsilon_t - \theta\epsilon_{t-1},$$

where ϵ_t are i.i.d.(0,1) and $\theta \in [-1;1]$.

Let us denote $\underline{y}_T = (y_0, \dots, y_T)$, where T is the sample size.

Let us consider the following objective function

$$Q_T(\underline{y}_T;\lambda) = \frac{1}{T}\sum_{t=1}^{T}(y_t - \lambda y_{t-1})^2 \tag{1}$$

where ϵ_t are i.i.d.(0,1) and $\theta \in [-1;1]$.

This objective function corresponds to the likelihood function of an **auxiliary model**. Here, this auxiliary model is an AR(1).

Remark : The value of λ that maximizes the objective function (1) is an OLS estimator (given the auxiliary model).

The OLS estimator of λ is

$$\hat{\lambda}_T = \frac{c\hat{o}v(y_t, y_{t-1})}{v\hat{a}r(y_{t-1})} = \frac{\sum_{t=1}^{T}(y_t - \bar{y})(y_{t-1} - \bar{y})}{\sum_{t=1}^{T}(y_{t-1} - \bar{y})^2}$$

$$\frac{\frac{1}{T}\sum_{t=1}^{T}(y_t - \bar{y})(y_{t-1} - \bar{y})}{\frac{1}{T}\sum_{t=1}^{T}(y_{t-1} - \bar{y})^2} \xrightarrow{a.s.} \frac{cov(y_t, y_{t-1})}{var(y_{t-1})} = \frac{-\theta_0}{(1+\theta_0^2)}$$

Then, the **binding function** is $\lambda_0(\theta_0) = \frac{-\theta_0}{(1+\theta_0^2)}$.

If the likelihood function of the true model is complicated to write, we can use an other model that is relatively easy to manipulate : it is the **auxiliary model**.

Doing so, the model we use is miss specified and the MLE is generally not consistent (except under some assumptions, see Gourieroux, Monfort, Trognon, 1984). We consider the case such that the conditional moments of the dependent variable are difficult to obtain.

The main idea is that we can deduce the value of θ_0 (indeed an approximation $\hat{\theta}$) from observing the value of $\hat{\lambda}$ obtained from the auxiliary model on the observed sample.

In this example, we obtain the **analytical expression** of the biding function. More generally, we have to use an approach that allow to link λ_0 and θ_0 numerically.

Let us consider the following objective function

$$\hat{\lambda}_T = \underset{\lambda}{argmax} \sum_{t=1}^{T} \ln(g(y_t \mid x_t; \lambda))$$

where g is the conditional pdf of the auxiliary model. Formally, $\hat{\lambda}_T$ is a **QML estimator** of λ.

Let us remark that is the model we consider is dynamic, then $x_t = (\underline{y}_{t-1}, z_t)$, where z_t is a set of exogenous variables and $\underline{y}_{t-1} = (y_0, \ldots, y_{t-1})$.

Let $y_t^r(\theta)$ **denote** a random draw for the **dependent variable** y_t that is obtained using the true model and the value θ of the vector of parameters.

Let us denote $x_t^r(\theta) = (\underline{y}_{t-1}^r(\theta), z_t)$.

We can estimate the parameter λ using the auxiliary model and H i.i.d. random draws

$$\hat{\lambda}_{T,H}(\theta) = \underset{\lambda}{argmax} \sum_{t=1}^{T} \sum_{r=1}^{H} \ln(g(y_t^r(\theta) \mid x_t^r(\theta); \lambda))$$

Then, we should choose the value of θ, namely $\hat{\theta}_{T,H}$, such that $\hat{\lambda}_{T,H}(\hat{\theta}_{T,H})$ is **as close as possible** to $\hat{\lambda}_T$.
This can be done using the program

$$\hat{\theta}_{T,H} = \underset{\theta \in \Theta}{argmin} \left(\hat{\lambda}_T - \hat{\lambda}_{T,H}(\theta)\right)' \; \Omega \; \left(\hat{\lambda}_T - \hat{\lambda}_{T,H}(\theta)\right)$$

where Ω is a nonnegative symmetric matrix.

Monte Carlo experiments : Let us consider the previous example for $\theta_0 = 0.5$, $T = 1000$ observations and let us assume ϵ_t are i.i.d. $N(0, 1)$.

Let us assume we observe $y = (y_0, y_1, \ldots, y_T)'$ and we estimate an auxiliary model corresponding to an AR(1). The parameter of the **auxiliary model** is denoted λ and the expression of the objective function is (1). We obtain then $\hat{\lambda}$ the OLS estimator of λ. The OLS estimator tends to $\lambda_0 = -\frac{\theta_0}{(1+\theta_0^2)} = -0.4$ (the pseudo true value of λ) when T becomes large.

In the following table, the number of Monte Carlo replications is 1000. Each one of the three panels corresponds to a Monte Carlo experiment. The **number of simulations** H is set to 100 (first panel), to 1000 (second panel) and to 2000 (third panel). For λ, RMSE is based on the pseudo true value λ_0 and for θ, RMSE is based on the true value θ_0.

Table: **Monte Carlo experiments (True model : MA(1)).**

	OLS (True data) λ	Indirect Inference θ
	$H = 100$	
Mean	-0.40287	0.58393
Std. dev.	0.02817	0.06723
RMSE	0.02818	0.10733
	$H = 1000$	
Mean	-0.40167	0.46334
Std. dev.	0.02565	0.04795
RMSE	0.02558	0.06168
	$H = 2000$	
Mean	-0.39946	0.51796
Std. dev.	0.01795	0.03436
RMSE	0.01787	0.03862

Note : T=1000. Monte Carlo replications=1000. $\theta_0 = 0.5$.

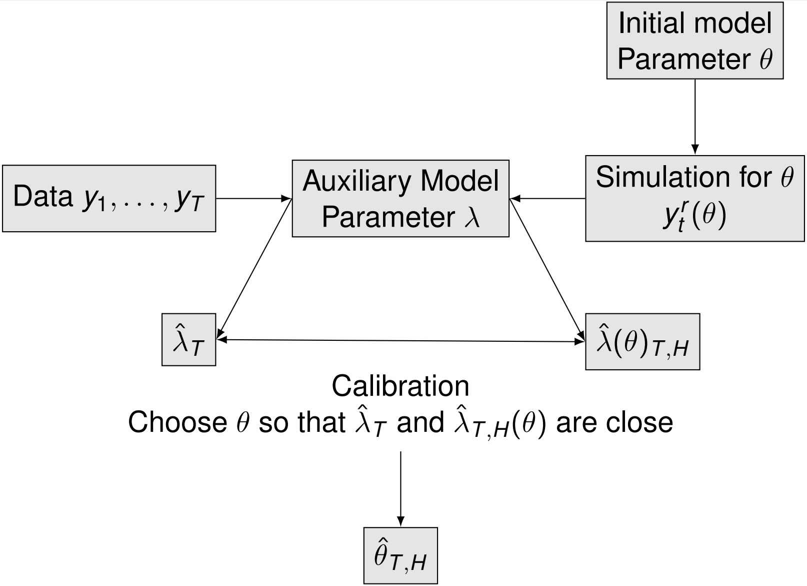

Figure: Indirect Inference (minimization of a norm)

In order to make the calibration, an estimator based on another objective function can be used (M-estimator). For instance the **score function** can be minimized with respect to θ.
Gallant and Tauchen (1996) propose to select θ such that the following function is close to zero as much as possible

$$\sum_{t=1}^{T} \sum_{r=1}^{H} \frac{\partial}{\partial \lambda} \ln(g(y_t^r(\theta) \mid x_t^r(\theta); \hat{\lambda}_T))$$

where $\hat{\lambda}_T$ is the QML estimator of λ (auxiliary model).
This can be done using the program

$$\hat{\theta}_{T,H} = \underset{\theta \in \Theta}{argmin} \left(\sum_{t=1}^{T} \sum_{r=1}^{H} \frac{\partial}{\partial \lambda} \ln(g(y_t^r(\theta) \mid x_t^r(\theta); \hat{\lambda}_T)) \right)' \Omega \left(\sum_{t=1}^{T} \sum_{r=1}^{H} \frac{\partial}{\partial \lambda} \ln(g(y_t^r(\theta) \mid x_t^r(\theta); \hat{\lambda}_T)) \right)$$

where Ω is a nonnegative symmetric matrix.

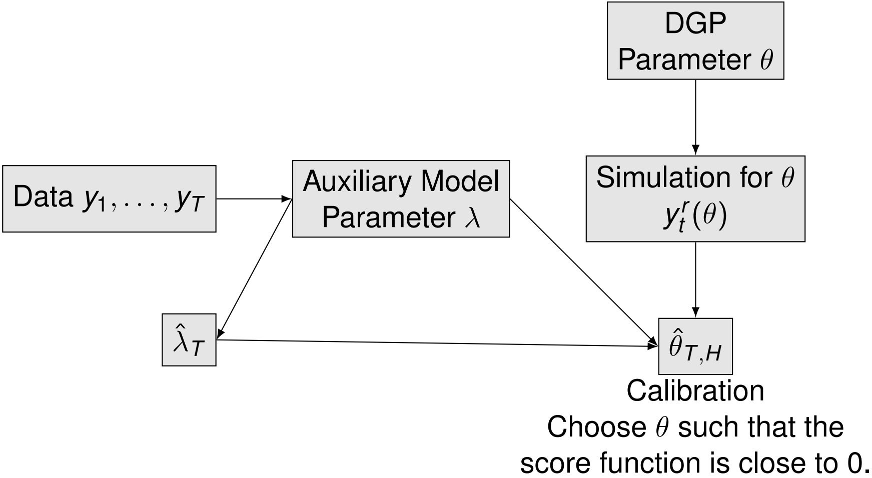

Figure: Indirect Inference (minimization of the score function)

In first approach, the calibration is based on the comparison of the **QML estimators.** One can use other estimators among the category of M-estimators.

Let us denote $f(y_t \mid x_t; \theta)$ the conditional density function of the "true model" (**the model we would like to estimate**) indexed by a parameter θ. This model is complicated to estimate directly but it is **easy to simulate**.

We consider an **auxiliary model**, indexed by an **auxiliary parameter** λ and an estimator of this last parameter based on the auxiliary model. The estimator of λ is such that

$$\hat{\lambda}_T = \max_{\lambda} Q_T(\underline{y}_T, \underline{z}_T; \lambda)$$

where $\underline{y}_T = (y_1, \ldots, Y_T)$ and $\underline{z}_T = (z_1, \ldots, z_T)$ is a vector of exogenous variables.

Let us assume that $y_t^r(\theta)$ are random draws of the endogenous variable for a given value of θ, $r = 1, \ldots, H$ and $t = 1, \ldots, T$. Let us remark that, in the case of a dynamic model, we have random paths.

Let us assume that the objective function $Q_T(\underline{y}_T^r(\theta), \underline{z}_T; \lambda)$ tends to the asymptotic objective function $Q_\infty(\lambda, \theta)$ when the sample size $T \longrightarrow \infty$ uniformly in (λ, θ).

For example, if the process Y_t is stationary, ergodic and the objective function in an empirical average:

$$Q_T(\underline{y}_T, \underline{z}_T; \lambda) = \frac{1}{T} \sum_{t=1}^{T} Q_t(y_t, x_t; \lambda)$$

$$Q_\infty(\lambda, \theta) = E_{P_\theta}[\, Q_t(y_t^r(\theta), x_t^r(\theta); \lambda)\,]$$

where $x_t^r(\theta) = (y_{t-1}^r(\theta), z_t)$ and P_θ is the distribution corresponding to the **initial model**.

Let us assume that the maximization of the asymptotic objective function with respect to λ is achieved for a unique value of λ

$$\lambda(\theta) = \underset{\lambda}{argmax}\, Q_\infty(\lambda, \theta)$$

where $\lambda(\theta)$ is the **binding function.**

Let us assume that $Q_\infty(\lambda, \theta)$ and $Q_T(\underline{Y}_T, \underline{z}_T; \lambda)$ are both differentiable with respect to λ and

$$\lim_{T\to\infty} \frac{\partial}{\partial\lambda} Q_T(\underline{Y}^r_T(\theta), \underline{z}_T; \lambda) = \frac{\partial}{\partial\lambda} Q_\infty(\lambda, \theta)$$

Let us assume $\lambda(\theta)$ is an **injective function** ($\lambda = \lambda(\theta)$ as a unique solution with respect to θ).

Let us assume that the asymptotic objective function is such that

$$\frac{\partial}{\partial\lambda} Q_\infty(\lambda, \theta) = 0 \Longrightarrow \lambda = \lambda(\theta)$$

and, consequently, $\lambda(\theta)$ is the unique solution of the first order conditions of the asymptotic optimization program.

Let us denote

$$\hat{\lambda}_T = \underset{\lambda}{argmax}\, Q_T(\underline{Y}_T, \underline{z}_T; \lambda)$$

As for comparison based on the quasi likelihood function as auxiliary objective function, we can use, alternatively, two method for **calibration**: comparison of estimators or the score function.

a) Calibration using the **comparison of estimators**:

Let us define

$$\hat{\lambda}_{T,H} = \underset{\lambda}{argmax} \sum_{r=1}^{H} Q_T(\underline{Y}_T^r(\theta), \underline{z}_T; \lambda)$$

The indirect inference estimator of θ is such that

$$\hat{\theta}_{T,H} = \underset{\theta}{argmin} \left(\hat{\lambda}_T - \hat{\lambda}_{T,H} \right)' \Omega \left(\hat{\lambda}_T - \hat{\lambda}_{T,H} \right) \tag{2}$$

where Ω is a nonnegative symmetric matrix.

b) Calibration using the **score function**:

The indirect inference estimator of θ is such that

$$\hat{\theta}_{T,H} = \underset{\theta}{argmin} \left(\sum_{r=1}^{H} \frac{\partial}{\lambda} Q_T(\underline{Y}_T^r(\theta), \underline{z}_T; \hat{\lambda}_T) \right)' \Omega \qquad (3)$$

$$\left(\sum_{r=1}^{H} \frac{\partial}{\lambda} Q_T(\underline{Y}_T^r(\theta), \underline{z}_T; \hat{\lambda}_T) \right)$$

where Ω s a nonnegative symmetric matrix.

The auxiliary parameters:

Unicity of the solution : In order to have a unique solution to the optimization problem (2) and (3), the dimension of the auxiliary parameter λ must be at least equal to the dimension of θ.

If $dim(\theta) = dim(\lambda)$ and the sample size T is large enough, then the value of the indirect inference estimator $\hat{\theta}_{T,H}$ obtained using the calibration based on the comparison of estimators do not depend on the matrix Ω, the value of the indirect inference estimator obtained using the calibration based on the score function do not depend on Ω and **both estimators are equal** (Gourieroux and Monfort, 1995).

The binding function $\lambda(\theta)$ is injective. In the case such that $dim(\theta) = dim(\lambda)$ a consistent estimation of θ_0 is the value of θ such that $\lambda(\theta) = \hat{\lambda}_T$. As the expression of the binding function is unknown in general and as $\hat{\lambda}_{T,H}(\theta) \underset{T\to\infty}{\longrightarrow} \lambda(\theta)$, this solution can be replaced by **the value of** θ **such that** $\hat{\lambda}_{T,H}(\theta) = \hat{\lambda}_T$. This particular value is the **indirect inference estimator** of θ, namely $\hat{\theta}_{T,H}$. Consequently, we use the indirect inference estimator as an approximation of the value of θ such that $\lambda(\theta) = \hat{\lambda}_T$.

Let us remind that

$$\lambda(\theta) = \underset{\lambda}{argmax}\, Q_{\infty}(\lambda, \theta)$$

where $\lambda(\theta)$ the **binding function.**

Let us denote

$$I_0 = \lim_{T\to\infty} V_{P_{\theta_0}}\left[\sqrt{T}\frac{\partial}{\partial\lambda}Q_T(\underline{y}_T, \underline{z}_T; \lambda_0) - E_{P_{\theta_0}}\left[\sqrt{T}\frac{\partial}{\partial\lambda}Q_T(\underline{y}_T, \underline{z}_T; \lambda_0) | \underline{z}_T\right]\right]$$

$$J_0 = \underset{T\to\infty}{plim} -\frac{\partial^2}{\partial\lambda\partial\lambda'}Q_T(\underline{y}_T, \underline{z}_T; \lambda_0)$$

Let us assume Ω is a non negative symmetric matrix.

Theorem (Gouriéroux, Monfort, Renault, 1993, Gouriéroux, Monfort, 1995)

Under some regularity conditions, the indirect inference estimator is consistent and asymptotically normal (H is fixed and T tends to ∞)

$$\sqrt{T}(\hat{\theta}_{T,H} - \theta_0) \xrightarrow[n \to \infty]{d} N(0, V)$$

where

$$V = \left(1 + \frac{1}{H}\right) \left(\frac{\partial}{\partial \theta} \lambda(\theta_0)' \, \Omega \, \frac{\partial}{\partial \theta'} \lambda(\theta_0) \right)^{-1}$$

$$\left(\frac{\partial}{\partial \theta} \lambda(\theta_0)' \, \Omega \, J_0^{-1} \, I_0 \, J_0^{-1} \, \Omega \, \frac{\partial}{\partial \theta'} \lambda(\theta_0) \right)$$

$$\left(\frac{\partial}{\partial \theta} \lambda(\theta_0)' \, \Omega \, \frac{\partial}{\partial \theta'} \lambda(\theta_0) \right)^{-1}$$

Theorem

Under some regularity conditions, for the optimal matrix $\Omega^* = J_0 I_0^{-1} J_0$, the Indirect inference estimator is asymptotically normal

$$\sqrt{T}(\hat{\theta}_{T,H} - \theta_0) \xrightarrow[n\to\infty]{d} N(0, V^*)$$

where

$$V^* = \left(1 + \frac{1}{H}\right) \left(\frac{\partial}{\partial \theta} \lambda(\theta_0)' \, J_0 \, I_0^{-1} \, J_0 \, \frac{\partial}{\partial \theta'} \lambda(\theta_0) \right)^{-1}$$

The expression of the asymptotic covariance matrix V^* includes the derivative of the binding function $\lambda(\theta)$ calculated for the true value of θ.

It is possible possible to show (Gourieroux and Monfort, 1995) that

$$V^* = \left(1+\frac{1}{H}\right)\left(\frac{\partial^2}{\partial\theta\partial\lambda'}Q_\infty(\lambda_0,\theta_0)\ I_0^{-1}\frac{\partial^2}{\partial\lambda\partial\theta'}Q_\infty(\lambda_0,\theta_0)\right)^{-1}$$

where $\lambda_0 = \lambda(\theta_0)$.

Example : Let us consider an i.i.d. sample and $y_i \mid x_i \sim$ exponential distribution with parameter $\exp(\beta_0 + \sum_{k=1}^{p} \beta_k \mathbb{1}[x_i = k])$, where $x_i \in \{0, 1, \ldots, p\}$.

Let us remark that

$$E[Y_i \mid x_i] = \exp(-\beta_0 - \sum_{k=1}^{p} \beta_k \mathbb{1}[x_i = k]).$$

Let us denote

$$\kappa(y_i, x_i) = \begin{pmatrix} \frac{n}{\sum_{i=1}^{n} \mathbb{1}[x_i = 0]} \; y_i \; \mathbb{1}[x_i = 0] \\ \frac{n}{\sum_{i=1}^{n} \mathbb{1}[x_i = 1]} \; y_i \; \mathbb{1}[x_i = 1] \\ \vdots \\ \frac{n}{\sum_{i=1}^{n} \mathbb{1}[x_i = p]} \; y_i \; \mathbb{1}[x_i = p] \end{pmatrix}$$

Let us consider the following objective function is

$$Q_n(\underline{y}_n, \underline{x}_n; \lambda) = -\left(\frac{1}{n}\sum_{i=1}^{n}\kappa(y_i, x_i) - \lambda\right)' \left(\frac{1}{n}\sum_{i=1}^{n}\kappa(y_i, x_i) - \lambda\right)$$

where $\lambda = (\lambda_0, \lambda_1, \ldots, \lambda_p)'$.

Here, for the "auxiliary model" we use **moment conditions.**

The objective criterium

$$Q_n(\underline{y}_n, \underline{x}_n; \lambda) \xrightarrow[n\to\infty]{a.s.} Q_\infty(\lambda; \theta_0)$$

where $\theta_0 = (\beta_0, \ldots, \beta_p)'$.

We have

$$\hat{\lambda}_n = \frac{1}{n}\sum_{i=1}^{n}\kappa(y_i, x_i) \xrightarrow[n\to\infty]{a.s.} \lambda_0 = \lambda(\theta_0) = E_{x_i, y_i}[\kappa(y_i, x_i)]$$

Let $p_k = Prob[x_i = k]$. We get

$$E_{x_i,y_i}[\kappa(y_i, x_i)] = \begin{pmatrix} \vdots \\ p_k^{-1} \sum_x [p_x E_{\theta_0}[\, 1\!\!1[x = k]\, y_i \mid x_i\,]] \\ \vdots \end{pmatrix}$$

$$= \begin{pmatrix} \vdots \\ p_k^{-1} \sum_x [p_x\, 1\!\!1[x = k]\, E_{\theta_0}[\, y_i \mid x_i\,]] \\ \vdots \end{pmatrix}$$

$$= \begin{pmatrix} E_{\theta_0}[\, y_i \mid x_i = 0\,] \\ \vdots \\ E_{\theta_0}[\, y_i \mid x_i = k\,] \\ \vdots \\ E_{\theta_0}[\, y_i \mid x_i = p\,] \end{pmatrix}.$$

Let us consider i.i.d. random draws of the dependent variable $y_i^r(\theta)$ obtained using the initial distribution with parameter $\theta = (\beta_0, \ldots, \beta_p)'$.

$$\hat{\lambda}_{n,H}(\theta) = \underset{\lambda}{argmin} \sum_{r=1}^{H} \left(\frac{1}{n} \sum_{i=1}^{n} \kappa(y_i^r(\theta), x_i) - \lambda \right)' \left(\frac{1}{n} \sum_{i=1}^{n} \kappa(y_i^r(\theta), x_i) - \lambda \right)$$

we obtain

$$\hat{\lambda}_{n,H}(\theta) = \frac{1}{nH} \sum_{i=1}^{n} \sum_{r=1}^{H} \kappa(y_i^r(\theta), x_i)$$

$$\hat{\theta}_{n,H} = \underset{\theta \in \Theta}{argmin} \left(\hat{\lambda}_n - \hat{\lambda}_{nH}(\theta) \right)' \Omega \left(\hat{\lambda}_n - \hat{\lambda}_{nH}(\theta) \right)$$

Finaly, the objective function we use in order to obtain the indirect inference estimator of the θ is

$$\hat{\theta}_{n,H} = \underset{\theta \in \Theta}{argmin} \left(\frac{1}{n} \sum_{i=1}^{n} \kappa(y_i, x_i) - \frac{1}{nH} \sum_{i=1}^{n} \sum_{r=1}^{H} \kappa(y_i^r(\theta), x_i) \right)' \Omega$$

$$\left(\frac{1}{n} \sum_{i=1}^{n} \kappa(y_i, x_i) - \frac{1}{nH} \sum_{i=1}^{n} \sum_{r=1}^{H} \kappa(y_i^r(\theta), x_i) \right)$$

One can remark that it is a MSM estimator that relies on the conditional moments $E[y_i \mid x_i]$.

We consider an important application of the method in the field of labor economics. Authors use **indirect inference** to estimate a joint model **earnings**, **employment**, **job changes**, **wage rates**, and **work hours** over a career (see Altonji, Smith and Vidangos, 2013).

Some questions addressed by the paper:

The experience profile of wages.

The response of job changes to outside wage offers.

The effect of seniority on job changes.

A study of the dynamic response of wage rates, hours, and earnings to various shocks.

Measure the relative contributions of the shocks to the variance of earnings in a given year and over the lifetime.

The model is estimated using data on male households heads from the Panel Study on Income Dynamics (PSID).

The model combine discrete and continuous variables, multiple equations, state dependence, unobserved heterogeneity **measurement error**, unbalanced panel. In practice, maximum likelihood and method of moment cannot be implemented.

In the literature, we can find **other papers** who use panel data and indirect inference : Bagger et al. (2015), Nagypal (2007).

In presence of both discrete and continuous endogenous variables, the authors use a **smoothing procedure**. This method allows to use optimization methods based on the gradient (Kean and Smith, 2003).

The method of estimation consists to compare the distribution of simulated data generated using the **structural model** (initial model) for a given value of the parameters and the characteristics of the observed data.

The data : The 1975-1997 waves of the PSID. Individuals ages from 18 to 62 years old.

Sample restricted to single or married male household heads. Less than 40 years of potential experience and not self employed. Authors do not considers observations such that the individual report to be retired, disables, a housewife, a student.

Potential experience is $t_i = age_{it} - max(Educ, 10) - 5$.

ED_{it} is the number of years in a row that an individual is employed at the survey date.

In 1975 and for individuals who join the survey after 1975, ED_{it} is set to the tenure in the current employment (TEN).

UD_{t-1} is the number of consecutive years, calculated at $t-1$, that the individual has not been employed (set to 0 the first time a person is observed).

Wage is the hourly wage rate at the time of the survey (observed only for employed individuals).

Initial conditions :

Employment : $E_{i1} = 1\!\!1[b_{0g} + \delta_{\mu}^{EE}\mu_i + \delta_{\eta}^{EE}\eta_i + \epsilon_{i1}^{EE} > 0],$

Wages : $wage_{i1}^{lat} = X_{i1}\gamma_x^w + \gamma_{t^3}^w 1^3 + \delta_{\mu}^w \mu_i + \omega_{i1} + v_{ij(1)},$

General Productivity: $\omega_{i1} \sim N(0, \sigma_{\omega_1,g}^2),$

Wage Job Match: $v_{ij(1)} \sim N(0, \sigma_{v_1}^2),$

Earnings error: $e_{i1} \sim N(0, \sigma_e^2),$

Other initial conditions: $TEN_{i1}{=}0, ED_{i1}{=}E_{i1}, UD_{i1}{=}1{-}E_{i1}, JC_{i1}{=}0,$

where *JC* means "Job change". $j(1)$ means the firm occupied by person i at time 1.

ω_{i1}, $v_{ij(1)}$ and e_{i1} are assumed to be independent and independent of the shocks.

The intercept b_{0g} and the variance $\sigma^2_{w_1,g}$ depend on **education group** g (these groupe defined in the paper).

Initial conditions : two approaches in the literature (Heckman 1981, Wooldridge 2005).

The observed variables are:

$$\begin{array}{ll} \text{Wage :} & wage^*_{it} = E_{it}\,(wage^{lat}_{it} + m^w_{it}), \\ \text{Hours:} & hours^*_{it} = hours_{it} + m^h_{it}, \\ \text{Earnings} & earn^*_{it} = earn_{it} + m^e_{it}, \end{array}$$

where **measurement errors** m^w_{it}, m^h_{it} and m^e_{it} are i.i.d. $N(0, \sigma^2_{m\tau})$, $\tau = w, h, e$. These errors are mutually independent and independent from all other error terms.

Log wage rate :

$$
\begin{aligned}
& wage_{it} = E_{it}\ wage_{it}^{lat}, \\
& wage_{it}^{lat} = X_{it}\gamma_X^w + \gamma_{t^3}^w t^3 + P(TEN_{it})\ \gamma_{TEN}^w + \delta_\mu^w \mu_i + \omega_{it} + v_{ij(t)}, \\
& \omega_{it} = \rho_\omega\, \omega_{i,t-1} + \gamma_{1-E_t}^\omega (1-E_{it}) + \gamma_{1-E_{t-1}}^\omega (1-E_{i,t-1}) + \epsilon_{it}^\omega, \\
& v_{ij(t)} = (1-S_{it}) v_{ij(t-1)} + S_{it} v'_{ij'(t)}, \\
& v'_{ij'(t)} = \rho_v\, v_{ij(t-1)} + \epsilon_{ij(t)}^v,
\end{aligned} \tag{4}
$$

where S_{it} is the employer change indicator. ϵ_{it}^k are i.i.d. $N(0, \sigma_k^2)$, where k is the dependent variable depending directly from ϵ_{it}^k. $P(TEN_{it})$ is a vector consisting of the first four powers of the employer tenure TEN_{it}. $v'_{ij'(t)}$ is a potential new match component.
$wage_{it}^{lat}$ is $wage_{it}$ for employed individual and job offer otherwise.

Employment Transitions (EE_t) : remaining employed

$$EE_{it} = \mathbb{1}[X_{i,t-1}\gamma_X^{EE} + \gamma_{ED}^{EE}\min(ED_{i,t-1}, 9) + \gamma_{TEN}^{EE} TEN_{i,t-1} + \gamma_{w^s}^{EE} wage_{it}^s$$

$$+\gamma_\mu^{EE}\mu_i + \delta_\eta^{EE}\eta_i + \epsilon_{it}^{EE} > 0], \text{ given } E_{i,t-1} = 1,$$

where $ED_{i,t-1}$ is the lagged value of employment duration, $wage_{it}^s$ is the value of $wage_{it}^{lat}$ with $E_{it} = 1$ (employed at time t), $E_{i,t-1} = 1$, $TEN_{it} = TEN_{i,t-1} + 1$ and the employer change indicator $S_{it} = 0$.

For a practical reason, the vector $X_{i,t-1}$ is the same as $X_{i,t}$ except it includes $(t-1)$ and $(t-1)^2$ instead of t and t^2.

$wage_{it}^s$ is a proxy of current wage opportunity. μ_i can be interpreted as a **permanent ability component.** η_i can interpreted as a **preference component relative to labour supply and job and employment mobility.**

Job change equation (JC_t) : conditional on remaining employed

$$JC_{it} = \mathbb{1}[X_{i,t-1}\gamma_X^{JC} + \gamma_{TEN}^{JC} TEN_{i,t-1} + \delta_{v'j'(t)}^{JC} v'_{ij'(t)} + \delta_{vj(t-1)}^{JC} v_{ij(t-1)} + \delta_{\mu}^{JC}\mu_i + \delta_{\eta}^{JC}\eta_i + \epsilon_{it}^{JC} > 0], \text{ given } E_{i,t-1} = E_{i,t} = 1,$$

where $v_{ij(t-1)}$ is the job match component of the **current job**. Job search model and Matching model predict that there is a negative impact of this component on the conditional probability to change job (lower search activity and higher reservation wage).

A person employed is assigned a potential draw, namely $v'_{ij'(t-1)}$ (see (4)). Search models predict that there is a positive impact of this component on the conditional probability to change of employer. $TEN_{i,t-1}$ is included to take into account the decline of the separation rate with tenure.

Unemployment to Employment transition (UE_t):

$$UE_{it} = \mathbb{1}[X_{i,t-1}\,\gamma_X^{UE} + \gamma_{UD}^{UE}\,UD_{i,t-1} + \delta_\mu^{UE}\,\mu_i + \delta_\eta^{UE}\eta_i + \epsilon_{it}^{UE} > 0],\ \text{given } E_{i,t} = 0,$$

where $UD_{i,t-1}$ is the number of years unemployed at the time of the survey and $UD_{i,t} = (1 - E_{i,t})(UD_{i,t-1} + 1)$.

Let us remark that

$$E_{i,t} = E_{i,t-1} \times EE_{i,t} + (1 - E_{i,t-1}) \times UE_{i,t-1}$$

Part of the individuals employed at time t were employed previously and part were unemployed at time $t - 1$.

Log annual hours (*$hours_{it}$*):

$$hours_{it} = X_{i,t}\,\gamma_X^h + \gamma_{t^3}^{h'} t^3 + (\gamma_E^h + \xi_{ij(t)})\,E_{it} + \gamma_\mu^h wage_{it}^{lat} + \delta_\mu^h \mu_i + \delta_\eta^h \eta_i + \epsilon_{it}^h$$

where $\xi_{ij(t)}$ is a job-specific component. This component can reflect the job schedules. For unemployed individuals, $wage_{it}^{lat}$ is the wage the individual would receive. For many individuals, $wage_{it}^{lat}$ is the wage they receive. Let us assume that $t^3 = ((t-1), (t-1)^2/10, (t-1)^3/1000)'$.

Log earnings (*$earn_{it}$*):

$$\begin{aligned} earn_{it} = &\ \gamma_0^e + \gamma_w^e\, wage_{it}^{lat} + \gamma_h^e\, hours_{it} + e_{it} \\ e_{it} = &\ \rho_e\, e_{i,t-1} + \epsilon_{it}^e. \end{aligned}$$

Let us remark that the parameters γ_w^e and γ_h^e can be different from 1, singularly due to the presence of overtime and bonuses.

Individual effects: μ_i and η_i are i.i.d. $N(0,1)$.

Shocks: ϵ_{it}^{EE}, ϵ_{it}^{UE} and ϵ_{it}^{JC} are i.i.d. $N(0,1)$.

Other assumptions: The job components $\xi_{ij(t)}$ are i.i.d. $N(0,\sigma_\xi^2)$ and the innovation terms $\epsilon_{ij(t)}^{v}$ are i.i.d. $N(0,\sigma_v^2)$.

These variables are independent from one another. The authors impose that $\delta_\mu^w > 0$ and $\delta_\eta^{JC} > 0$.

Estimation:

The **initial model** has k parameters. Let us denote β the vector of parameters of the initial model (the one we are interested in).

The **auxiliary model** has $p \geq k$ parameters. Let us denote θ the vector of parameters of the auxiliary model.

Let us assume we have n individuals in the sample and T periods of observation. Let us assume that y_{it} is a vector of endogenous variables and x_{it} is a vector of exogenous variables.

The vector of parameters θ of the **auxiliary model** can be estimated using the observed data as a solution of the objective function

$$\hat{\theta} = \underset{\theta}{argmax}\, L(y; x; \theta)$$

where $L(y; x; \theta)$ is the likelihood function of the auxiliary model (quasi likelihood).

Given x and β we can use the **initial model** in order to generate H i.i.d. simulated data sets, denoted $y_{it}^{r}(\beta)$, $r = 1, \ldots, H$.

The size of each simulated data set is n.

For each simulated data set we calculate

$$\tilde{\theta}_r(\beta) = \underset{\theta}{argmax}\, L(\tilde{y}_r(\beta); x; \theta),$$

where $\tilde{y}_r(\beta) = \{y^r_{it}(\beta)\}$ is a simulated sample (under the initial model and given x and θ).
Let us denote

$$\tilde{\theta}(\beta) = \frac{1}{H}\sum_{r=1}^{H}\tilde{\theta}_r(\beta)$$

the average of the estimated vectors of parameters.

We have presented two ways to estimate β (**calibration**).
Altonji et al. (2013) use a **third method**

$$\hat{\beta} = \underset{\beta}{argmin}\left(L(y; x; \hat{\theta}) - L(y; x; \tilde{\theta}(\beta))\right)$$

Remarks: The objective function is discontinuous with respect to β due to the presence of discrete random variables (E_{it}, JC_{it}, UE_{it}, EE_{it}). Indeed, the simulated choice $\tilde{y}_{it}^{r}(\beta)$ is discontinuous with respect to β. Then, the estimated vector of auxiliary parameters $\tilde{\theta}(\beta)$ is discontinuous with respect to β.

To illustrate, let us assume that we have only a binary dependent variable. The binary variable $\tilde{y}_{it}^{r}(\beta) = 1$ if the latent utility $\tilde{u}_{it}^{r}(\beta)$ is positive and $y_{it}^{r}(\beta) = 0$ otherwise.

The idea is to replace when we estimate the auxiliary model $\tilde{y}_{it}^{r}(\beta) = 1$ by a continuous function $g(\tilde{u}_{it}^{r}(\beta); \lambda)$ such that $g(\tilde{u}_{it}^{r}(\beta); \lambda)$ converges to $\tilde{y}_{it}^{r}(\beta)$ when the smoothing parameter λ tends to 0 (see Lecture 6).

They fix $H = 20$ but choose $\lambda = 0.05$ and they argue the bias is small using simulation experiments.

The auxiliary model:

The auxiliary model has two parts. The first one allows to obtain information on the parameters γ_X^w, $\gamma_{t^3}^w$, γ_X^h, $\gamma_{t^3}^w$ of the **initial model**.

The corresponding equations of the **auxiliary model** are

$$\begin{aligned} wage_{it}^* = & \quad [X_{it}, t^3]\theta_1^w + u_{it}^w, \\ hours_{it}^* = & \quad [X_{it}, t^3]\theta_1^h + u_{it}^h, \end{aligned}$$

The corresponding objective function is

$$L_1(\theta_1^w, \theta_1^h) = \sum_{i,t}(wage_{it}^* - [X_{it}, t^3]\theta_1^w)^2 + \sum_{i,t}(hours_{it}^* - [X_{it}, t^3]\theta_1^h)^2$$

Let us consider the following dependent variables

$$\widetilde{wage}^*_{it} = \quad wage^*_{it} - [X_{it}, t_i^3]\theta_1^w,$$

$$\widetilde{hours}^*_{it} = \quad hours^*_{it} - [X_{it}, t_i^3]\theta_1^h,$$

$$\widetilde{earn}^*_{it} = \quad earn^*_{it} - [X_{it}, t_i^3]\theta_1^w - [X_{it}, t_i^3]\theta_1^h,$$

$$Y_{it} = [E_{it}E_{i,t-1}, E_{it}(1-E_{i,t-1}), JC_{it}E_{it}E_{i,t-1}, \widetilde{wage}^*_{it}, \widetilde{hours}^*_{it}, \widetilde{earn}^*_{it}, \ln(1+\widetilde{wage}^{*2}_{it})]'$$

Let us consider the following vector of explanatory variables

$$
\begin{aligned}
Z_{it} = \; & [Const, (t_i - 1), (t_i - 1)^2, Black_i, EDUC_i, ED_{i,t-1}, UD_{i,t-1}, \\
& TEN_{i,t-1}, E_{i,t-1}E_{i,t-2}, E_{i,t-2}E_{i,t-3}, E_{i,t-1}(1 - E_{i,t-2}), \\
& E_{i,t-2}(1 - E_{i,t-3}), JC_{i,t-1}E_{i,t-1}E_{i,t-2}, JC_{i,t-2}E_{i,t-2}E_{i,t-3}, \\
& \widetilde{wage}^*_{i,t-1}, \widetilde{wage}^*_{i,t-2}, \widetilde{hours}^*_{i,t-1}, \widetilde{hours}^*_{i,t-2}, \widetilde{earn}^*_{i,t-1}, \widetilde{earn}^*_{i,t-2}, \\
& \widetilde{wage}^*_{i,t-1}(t_i - 1), \widetilde{wage}^*_{i,t-1}(t_i - 1)^2, \widetilde{wage}^*_{i,t-1}JC_{i,t}, \\
& \widetilde{wage}^*_{i,t-2}JC_{i,t-1}, \widetilde{wage}^*_{i,t-2}E_{i,t-1}]'.
\end{aligned}
$$

where $\widetilde{wage}^*_{i,t-1}(t_i - 1)$ and $\widetilde{wage}^*_{i,t-1}(t_i - 1)^2$ take account of change with potential experience on the level of persistence of wages.

Terms $\widetilde{wage}^{*}_{i,t-1}JC_{i,t}$, $\widetilde{wage}^{*}_{i,t-2}JC_{i,t-1}$ and $\widetilde{wage}^{*}_{i,t-2}E_{i,t-1}$ capture the impact of job mobility and past employment on state dependence in the wage equation.

This part of the **auxiliary model** consists to estimate

$$Y_{it} = \Pi Z_{it} + u_{it}, \tag{5}$$

where u_{it} are i.i.d. $N(0, \Sigma)$.

We consider the likelihood $L_2(\theta_2)$ that correspond to the system (5), where $\theta_2 = (\Pi, \Sigma)$.

Π has 7×25 elements (7 **dependent variables** and 25 explanatory variables). Σ is a 7×7 variance-covariance matrix (28 unique elements). Consequently, this part of the auxiliary model has 203 parameters. The quasi likelihood L_1 has 10 slope parameters and two variances.

The model consists in 55 parameters plus measurement error parameters, tenure coefficients and ρ_w.

Altonji et al. (2013) add additional moment conditions in order to identify some extra parameters (these conditions are incorporated in the objective function). For instance, the variance $\sigma^2_{w_1,g}$ is estimated as "the variance of (residual) wage observations in the PSID corresponding to $t \leq 5$".

The number of characteristics of the data used to estimate the initial model is much larger than the number of parameters of the initial model.

The objective function is $\alpha\, L_1(\theta_1^w, \theta_1^h) + L_2(\theta_2)$ where α is a weight (fixed to a large value in order to identify the parameters of the corresponding equations). The indirect inference estimator is consistent for any positive value of α.

Let us consider the case such that the dependent variable y is such that

$$\begin{aligned} y = & \quad g(y^*;\theta), \\ y^* = & \quad m(x,\epsilon;\theta), \end{aligned}$$

where x is a vector of exogenous variables, $\theta \in \boldsymbol{R}^p$ is a vector of parameters to be estimated, ϵ is an error term i.i.d. according to the cdf F_ϵ (ϵ is independent of x).

The functions g and m are known up to the vector of parameter θ.

We consider here the case such that g is not continuous with respect to the latent variable y^*.

Example: Binary choice model with serially dependent errors

$$y_{it} = \mathbb{1}[x_{it}'\beta + \nu_{it} > 0]$$

where $\nu_{it} = \rho\nu_{it-1} + \epsilon_{it}$, $i = 1, \ldots, n$ and $t = 1, \ldots, T$.

In this example, $y_{it}^{*} = x_{it}'\beta + \rho\nu_{it-1} + \epsilon_{it}$, and $\theta = (\beta', \rho)'$. $\square$

Let H denote the number of draws per observation. Let us assume that ϵ_i^r is i.i.d. $\sim F_\epsilon$. Let us consider

$$\begin{aligned} y_i^r = & \quad g(y_i^{r*}; \theta), \\ y_i^{r*} = & \quad m(x_i, \epsilon_i^r; \theta), \end{aligned}$$

Let us consider an **auxiliary model** for y_i. In this auxiliary model we have a vector of explanatory variables z_i and a vector of parameters $\gamma \in \boldsymbol{R}^d (d \geq p)$.

The auxiliary model imply moment conditions

$$E[h(y_i, z_i; \gamma^0)] = 0$$

Let us consider the **simulated auxiliary moments**

$$M_n(\theta, \gamma) = \frac{1}{nH} \sum_{i=1}^{n} \sum_{r=1}^{H} h(y_i^r(\theta), z_i; \gamma), \tag{6}$$

1) the indirect inference estimator (based on simulated moments):

$$\hat{\theta}_{II} = \underset{\theta}{argmin} \mid\mid M_n(\theta, \hat{\gamma}) \mid\mid^2_{\Omega_n} \equiv \underset{\theta}{argmin}\ Q^{LM}(\theta) \tag{7}$$

where Ω_n is a sequence of positive definite matrices and $\hat{\gamma}$ is the solution of the sample analog of the population moments

$$\frac{1}{n}\sum_{i=1}^{n} h(y_i, z_i; \gamma) = 0 \tag{8}$$

2) the indirect inference estimator (based on comparison):

$$\hat{\theta}_{II} = \underset{\theta}{argmin} \mid\mid \tilde{\gamma}(\theta) - \hat{\gamma}) \mid\mid^2_{\Omega_n} \equiv \underset{\theta}{argmin}\ Q^{W}(\theta) \tag{9}$$

where $\tilde{\gamma}(\theta) = \frac{1}{H}\sum_{r=1}^{H} \hat{\gamma}^r(\theta)$ and $\hat{\gamma}^r(\theta)$ is the solution of the sample analog of the population moments

$$\frac{1}{n}\sum_{i=1}^{n} h(y_i^r(\theta), z_i; \gamma) = 0 \tag{10}$$

Example: (continued) Let us consider $(u^r_{i1}, \ldots, u^r_{iT})$, for $i = 1, \ldots, n$, $r = 1, \ldots, H$, a vector of simulated **uniform** random variables.

For θ fixed we proceed recursively,

$$\begin{aligned}
\nu^r_{it} = &\quad \rho\nu^r_{it-1} + F^{-1}_{\epsilon}(u^r_{it}), \\
\nu^r_{i1} = &\quad F^{-1}_{\epsilon}(u^r_{i1}), \\
y^{r*}_{it}(\theta) = &\quad x'_{it}\beta + \rho\nu^r_{it-1} + F^{-1}_{\epsilon}(u^r_{it}), \\
y^r_{it}(\theta) = &\quad 1\!\!1[y^{r*}_{it}(\theta) > 0] = 1\!\!1[F_{\epsilon}(-x'_{it}\beta - \rho\nu^r_{it-1}) < u^r_{it}],
\end{aligned}$$

For the **auxiliary model** we can consider for instance a **linear probability model**:

$$y_{it} = z'_{it}\gamma + v_{it},$$

where $z_{it} = (x'_{it}, x'_{it-1})'$ and v_{it} is an error term.

Let us consider $h(y_{it}, z_{it}; \gamma) = z_{it}(y_{it} - z_{it}'\gamma)$ to construct the moments. Then, using moment conditions on the sample similar to (8) and (10) respectively, the estimators of the auxiliary parameter are

$$\hat{\gamma} = (\sum_{i=1}^{n}\sum_{t=2}^{T} z_{it}z_{it}')^{-1}\sum_{i=1}^{n}\sum_{t=2}^{T} z_{it}y_{it}$$

and

$$\hat{\gamma}^r(\theta) = (\sum_{i=1}^{n}\sum_{t=2}^{T} z_{it}z_{it}')^{-1}\sum_{i=1}^{n}\sum_{t=2}^{T} z_{it}y_{it}(\theta)^r, \text{ for } r = 1, \ldots, H,$$

respectively.

Using the objective function (7) or (9) one may try to estimate θ. However, the relation between θ and $y_{it}(\theta)^r$ is **not continuous**. Then, the derivative of the objective function w.r.t. θ do not exists for all θ. $\square$

The idea : The objective function may not be continuous with respect to the parameters. The objective function and its derivatives are functions of the random variables u_i^r (i is the index of the observation and r is the index of the draw).

We are going to **replace these draws** by new draws $u_i^r(\theta, \theta^*)$ (when we want to calculate the derivatives of the objective with respect to θ at point θ^*) in such a way that the new derivatives exist and are continuous.

The sample counterparts of the moment conditions converge to the same asymptotic criterium (expectation). This last property comes from the fact that we use a **change of variables** and that we are going to "weight" properly the empirical moments using $\frac{\partial\ u_i^r(\theta,\theta^*)}{\partial\ u_i^r} \equiv \omega_i^r(\theta, \theta^*)$ in oder to take this change of variables into account.

Assumption: The function g can be written

$$g(y_i^*;\theta) = \sum_{j=0}^{J} \alpha_j \ 1\!\!1[c_i^j(\theta) < u_i \leq c_i^{j+1}(\theta)] \tag{11}$$

where the parameters α_j, $j{=}0,\ldots,J$, are known constants, $u_i = F_\epsilon(\epsilon)$ is a uniform variable and $c_i^j(\theta)$, $j{=}0,\ldots,J{+}1$, are twice continuous differentiable functions ($c_i^0(\theta) = 0$ and $c_i^{J+1}(\theta) = 1$). Moreover, $c_i^j : \Theta \longrightarrow [0,1]$.

Let θ^* **denote** a point at witch we want to calculate the derivatives of the function $y_i^r(\theta)$. Let us consider the change of variable

$$u_i^r(\theta,\theta^*) = c_i^j(\theta) + \frac{c_i^{j+1}(\theta) - c_i^j(\theta)}{c_i^{j+1}(\theta^*) - c_i^j(\theta^*)}\,(u_i^r - c_i^j(\theta^*)),$$

where $c_i^j(\theta^*) < u_i^r \leq c_i^{j+1}(\theta^*)$, for all $j = 0,\ldots,J$.

Let $\omega_i^r(\theta, \theta^*)$ **denote** the Jacobian $\frac{\partial\ u_i^r(\theta,\theta^*)}{\partial\ u_i^r}$. We have

$$\omega_i^r(\theta, \theta^*) = \frac{c_i^{j+1}(\theta) - c_i^j(\theta)}{c_i^{j+1}(\theta^*) - c_i^j(\theta^*)},$$

where $c_i^j(\theta^*) < u_i^r \leq c_i^{j+1}(\theta^*)$, for all $j = 0, \ldots, J$.

The new simulated outcomes are

$$y_i^r(\theta, \theta^*) = \sum_{j=0}^{J} \alpha_j \ 1\!\!1[c_i^j(\theta) < u_i^r(\theta, \theta^*) \leq c_i^{j+1}(\theta)]$$

The moment function $h(y_i^r(\theta), z_i, \gamma)$ (see (8) or (10)) is replaced by the following one

$$h_i^r(\theta, \theta^*, \gamma) = h(y_i^r(\theta, \theta^*), z_i, \gamma)\ \omega_i^r(\theta, \theta^*)$$

The moment conditions (6) are replaced by

$$M_n(\theta, \theta^*, \gamma) = \frac{1}{n\,H} \sum_{i=1}^{n} \sum_{r=1}^{H} h_i^r(\theta, \theta^*, \gamma)$$

Under regularity conditions, the derivative of $M_n(\theta, \theta^*, \gamma)$ with respect to θ is an unbiased and uniformly consistent estimator of the same derivative for the limit program (n tends to infinity).

Frazier et al. (2019), under some regularity conditions, show that the indirect inference estimators $\hat{\theta}_n^{LM}$ and $\hat{\theta}_n^{W}$ are both **consistent** (n tends to infinity) and, in particular, if H is fixed, are **asymptotically normal**.

Example: (continued)

Let us remind that

$$y_{it}^r(\theta) = \mathbb{1}[-x_{it}'\beta - \rho\nu_{it-1}^r < u_{it}^r.]$$

So, in this case, $J = 1$ and $\alpha_0 = 0$ and $\alpha_1 = 1$ (see (11)).

Then the **critical functions** are $c_{it}^0(\theta) = 0$, $c_{it}^1(\theta) = F_\epsilon(-x_{it}'\beta - \rho\nu_{it-1}^r)$ and $c_{it}^2(\theta) = 1$.

Let $\theta^* = (\beta^{*\prime}, \rho^*)'$ a point where we want to calculate the function $y_{it}^r(\theta)$. For the **change of variables** we consider

$$u_{it}^r(\theta, \theta^*) = \begin{cases} \frac{c_{it}^1(\theta)}{c_{it}^1(\theta^*)}\ u_{it}^r,\ \text{if } u_{it}^r \leq c_{it}^1(\theta^*), \\ \\ c_{it}^1(\theta) + \frac{1-c_{it}^1(\theta)}{1-c_{it}^1(\theta^*)}\ (u_{it}^r - c_{it}^1(\theta^*)),\ \text{if } c_{it}^1(\theta^*) < u_{it}^r, \end{cases}$$

As these variables depend on θ only via $c_{it}^1(\theta)$ that is continuously differentiable, they share the same property. It is also the case of the Jacobian $\omega_{it}^r(\theta, \theta^*)$.

The dependent variable are now simulated using the new variables $u_{it}^r(\theta, \theta^*)$ as follows

$$y_{it}^r(\theta, \theta^*) = 1\!\!1[c_{it}^1(\theta) < u_{it}^r(\theta, \theta^*)]$$

Remark : $c_{it}^1(\theta) < u_{it}^r(\theta, \theta^*)$ **iff** $c_{it}^1(\theta^*) < u_{it}^r$, then $y_{it}^r(\theta, \theta^*) = 1\!\!1[c_{it}^1(\theta^*) < u_{it}^r]$.

The approximated version of the moment function is

$$M_n(\theta, \theta^*, \gamma) = \frac{1}{nH} \sum_{i=1}^{n} \sum_{r=1}^{H} \sum_{t=2}^{T} z_{it} \left(y_{it}^r(\theta, \theta^*) - z_{it}'\gamma \right) \omega_{it}^r(\theta, \theta^*).$$

The estimator of γ is

$$\tilde{\gamma}(\theta, \theta^*) = \frac{1}{H} \sum_{r=1}^{H} \left(\sum_{i=1}^{n} \sum_{t=2}^{T} z_{it} z_{it}' \right)^{-1} \sum_{i=1}^{n} \sum_{t=2}^{T} z_{it} \ y_{it}^r(\theta, \theta^*) \ \omega_{it}^r(\theta, \theta^*)$$

Then, the approximated moment function are differentiable with respect of θ. For instance,

$$\frac{\partial}{\partial \theta} M_n(\theta, \theta^*, \gamma) = \frac{1}{nH} \sum_{i=1}^{n} \sum_{r=1}^{H} \sum_{t=2}^{T} z_{it} \left(\mathbb{1}[c_{it}^1(\theta^*) < u_{it}^r] - z_{it}'\gamma \right) \frac{\partial}{\partial \theta} \omega_{it}^r(\theta, \theta^*)$$

This derivative evaluated at the point $\theta = \theta^*$ is an unbiased and uniformly consistent estimator of $\frac{\partial}{\partial \theta} M_n(\theta^*, \gamma)$. $\square$

Altonji, J.G., Smith, A.A., Vidangos, I., 2013, Modeling Earnings Dynamics, Econometrica, 81(4), 1395-1454.

Bagger, J., Fontaine, F., Postel-Vinay, F., Robin, J.-M., 2015, Tenure, Experience, Human Capital, and Wages: A Tractable Equilibrium Search Model of Wage Dynamics, The American Economic Review , 104(6), 1551-1596.

Frazier, D.T., Oka, T., Zhu D., 2019, Indirect inference with a non-smooth criterion function, Journal of Econometrics, 212, 623-645.

Gallant, A.R., Tauchen, G., 1996, Which moments to match?, Econometric Theory, 12, 657-681.

Gourieroux, Ch., Monfort, A., Renault, E., 1993, Indirect Inference, Journal of Applied Econometrics, S85-S118.

Gourieroux, Ch., Monfort, A., 1995, Simulated Based Econometric Methods, Core Lecture Series, Université Catholique de Louvain.

Keane, M., Smith, A., 2003, Generalized indirect inference for discrete choice models, Unpublished Manuscript, Yale University.

Nagypál, É., 2007, Learning by Doing vs. Learning about Match Quality: Can We Tell Them Apart?, The Review of Economic Studies , 74(2), 537-566.

Smith, A.A., 1993, Estimating non linear time-series models using simulated vector autoregressions, Journal of Applied Econometrics, 8, S63-S84.