IDENTIFICATION AND ESTIMATION

OF NONPARAMETRIC FINITE MIXTURES

(ECO 2403)

Victor Aguirregabiria

Winter 2016

1. Introduction and Examples of NPFM models

2. ML Estimation of FM models: EM Algorithm

3. Identification of NPFM: Basic Concepts

4. Identification under Conditional Independence

5. Estimation Methods

6. Identification and Tests of the Number of Mixtures

7. Identification of Markov NPFM

8. Identification using Exclusion Restrictions

9. Applications to Games

REFERENCES

EM algorithm:

- Dempster, Laird, and Rubin (JRSS, 1977)
- Wu (AS, 1983)
- Arcidiacono and Jones (ECMA, 2003)

Identification (Cross-section):

- Hall, and Zhou (AS, 2003)
- Allman, Matias, and Rhodes (AS, 2009)
- Bonhomme, Jochmans, and Robin (JRSS, 2016)
- Compiani and Kitamura (2015)

REFERENCES

Identification: Number of Mixtures

- Kasahara and H., and K. Shimotsu (JRSS, 2014)
- Kasahara and Shimotsu (JASA, 2015)

Identification: Markov Models

- Kasahara and Shimotsu (ECMA, 2009)

REFERENCES

Estimation

- Arcidiacono and Jones (ECMA, 2003)
- Arcidiacono and Miller (ECMA, 2011)
- Bonhomme, Jochmans, and Robin (JRSS, 2016)

Applications to Games

- Bajari, Hong, and Ridder (IER, 2011)
- Aguirregabiria and Mira (2015)

1. INTRODUCTION.

• **Unobserved heterogeneity** is pervasive in economic applications. Heterogeneity across individuals, households, firms, markets, etc.

• Not accounting for unobserved heterogeneity may imply important biases in the estimation of parameters of interest, and in our understanding of economic phenomena.

• The key feature of **Finite Mixture models** is that the variables that represent unobserved heterogeneity have finite support. There is a finite number of unobserved types.

• As we will see, this finite support structure can be without loss of generality.

INTRODUCTION.

• FM models have been extensively applied in statistics (e.g., medical science, biology) to identify and deal with unobserved heterogeneity in the description of data.

• These models are currently receiving substantial attention in **Structural Econometrics** in the estimation of dynamic structural models and empirical games.

• **Two-step estimation procedures in Structural Econometrics**. The first step in these methods involves nonparametric estimation of agents' choice probabilities conditional not only on observable state variables but also on time-invariant individual unobserved heterogeneity (dynamic models) or market-level unobserved heterogeneity in games.

INTRODUCTION: Example. Dynamic structural model

• $y_{nt} \in \{0, 1\}$ Firm n's decision to invest in a certain asset (equipment) at period t. Model:

$$y_{nt} = 1\left\{\varepsilon_{nt} \leq v\left(y_{n,t-1}, \omega_n\right)\right\}$$

where ε_{nt} is unobservable and i.i.d. with CDF F_ε, and ω_n is unobservable, time invariant, and heterogeneous across firms.

• The conditional choice probability (CCP) for a firm is:

$$\Pr(y_{nt} = 1 \mid y_{n,t-1}, \omega_n = \omega) \equiv P_\omega\left(y_{nt-1}\right) = F_\varepsilon\left[v\left(y_{nt-1}, \omega\right)\right]$$

Example. Dynamic structural model [2]

• Given panel data of N firms over T periods of time, $\{y_{nt} : t = 1, 2, ..., T;$ $n = 1, 2, ..., N\}$, the Markov structure of the model, and a Finite Mixture structure for ω_n, we have that:

$$\Pr(y_{n1}, y_{n2}, ..., y_{nT}) = \sum_{\omega=1}^{L} \pi_\omega \left[P_\omega^*(y_{n1}) \prod_{t=2}^{T} P_\omega(y_{nt-1})^{y_{nt}} \left[1 - P_\omega(y_{nt-1})\right]^{1-y_{nt}} \right]$$

• We present conditions under which the "type-specific" CCPs $P_\omega(y_{nt-1})$ are NP identified from these data.

• These estimates can be used to construct value functions, and this approach can facilitate very substantially the estimation of structural parameters in a second step.

Example. Static Game of Market Entry

• T firms, indexed by $t = 1, 2, ..., T$, have to decide whether to be active or not in a market m. $y_{mt} \in \{0, 1\}$ is firm t's decision to be active in market m.

• Given observable market characteristics x_m and unobserved market characteristics ω_m, the probability of entry of firm t in a market of "type" ω is:

$$\Pr(y_{mt} = 1 \mid x_m, \omega_n = \omega) \equiv P_{\omega,t}(x_m)$$

Example. Static Game of Market Entry [2]

• In a game of incomplete information with independent private values, we have that:

$$\Pr\left(y_{m1}, y_{m2}, ..., y_{mT} \mid x_m\right) = \sum_{\omega=1}^{L} \pi_\omega \left[\prod_{t=1}^{T} P_{\omega,t}\left(x_m\right)^{y_{mt}} \quad \left[1 - P_{\omega,t}\left(x_m\right)\right]^{1-y_{mt}}\right]$$

• Given a random sample of M markets, we provide conditions under which it is possible to use these data to identify NP firms' CCPs $P_{\omega,t}\left(x_m\right)$ for every firm t and every market type ω.

• These estimates can be used to construct firms' expected profits and best response functions, and this approach can facilitate very substantially the estimation of structural parameters of the game in a second step.

INTRODUCTION: Variables and Data

• Let $\mathbf{Y}$ be a vector of T random variables: $\mathbf{Y} = (Y_1, Y_2, ..., Y_T)$. We index these random variables by $t \in \{1, 2, ..., T\}$. We use small letters, $\mathbf{y} = (y_1, y_2, ..., y_T)$ to represent a realization of $\mathbf{Y}$.

• The researcher observes a random sample with N i.i.d. realizations of $\mathbf{Y}$, indexed by n, $\{\mathbf{y}_n : n = 1, 2, ..., N\}$.

• EXAMPLES:

(1) Standard longitudinal data. $\mathbf{Y}$ is the history **over** T **periods of time** of a variable measured at the individual level (or firm., or market, level). N is the number of individuals in the sample.

• EXAMPLES:

(2) $\mathbf{Y}$ is the vector of prices **of** T **firms in a market**. N is the number of markets in the sample.

(3) $\mathbf{Y}$ is the vector with the characteristics **of** T **members of a family**. N is the number of families in the sample.

(4) $\mathbf{Y}$ is the vector with the academic outcomes **of** T **students in a classroom**. N is the number of classrooms in the sample.

(5) $\mathbf{Y}$ is the vector of actions **of** T **players in a game**. N is the number of realizations of the game in the sample.

INTRODUCTION: Conditioning Exogenous Variables

• In most applications, the econometric model includes also a vector of observable exogenous variables $\mathbf{X}$, such that the data is a random sample, $\{\mathbf{y}_n, \mathbf{x}_n : n = 1, 2, ..., N\}$.

• The researcher is interested in the estimation of a model for $P(\mathbf{Y} \mid \mathbf{X})$.

• For notational simplicity, we will omit $\mathbf{X}$ as an argument and use $P(\mathbf{Y})$.

• Now, incorporating exogenous conditioning variables in NPFM models is not always trivial. I will be explicit when omitting $\mathbf{X}$ is without loss of generality and when it is not.

INTRODUCTION: Mixture Models

• **Mixture models** are econometric models where the observable variable is the convolution or mixture of multiple probability distributions with different parameters, and the parameters themselves follow a probability distribution.

$$P(\mathbf{Y}) = \int \pi(\omega) \; f_\omega(\mathbf{Y}) \; d\omega$$

- $\mathbf{Y}$ is the observable variable(s)

- $P(\mathbf{Y})$ is the mixture distribution

- ω is the unobserved (or mixing) variable (unobserved type)

- $f_\omega(\mathbf{Y})$ are the type-specific density

- $\pi(\omega)$ is the mixing distribution

INTRODUCTION: Nonparametric Finite Mixture models

• **Nonparametric Finite Mixture models** are mixture models where:

[1] The mixing distribution $\pi(\omega)$ has finite support;

$$\omega \in \Omega = \{1, 2, ..., L\}$$

such that:

$$P(\mathbf{Y}) = \sum_{\omega=1}^{L} \pi_\omega \; f_\omega(\mathbf{Y})$$

with $\sum_{\omega=1}^{L} \pi_\omega = 1$.

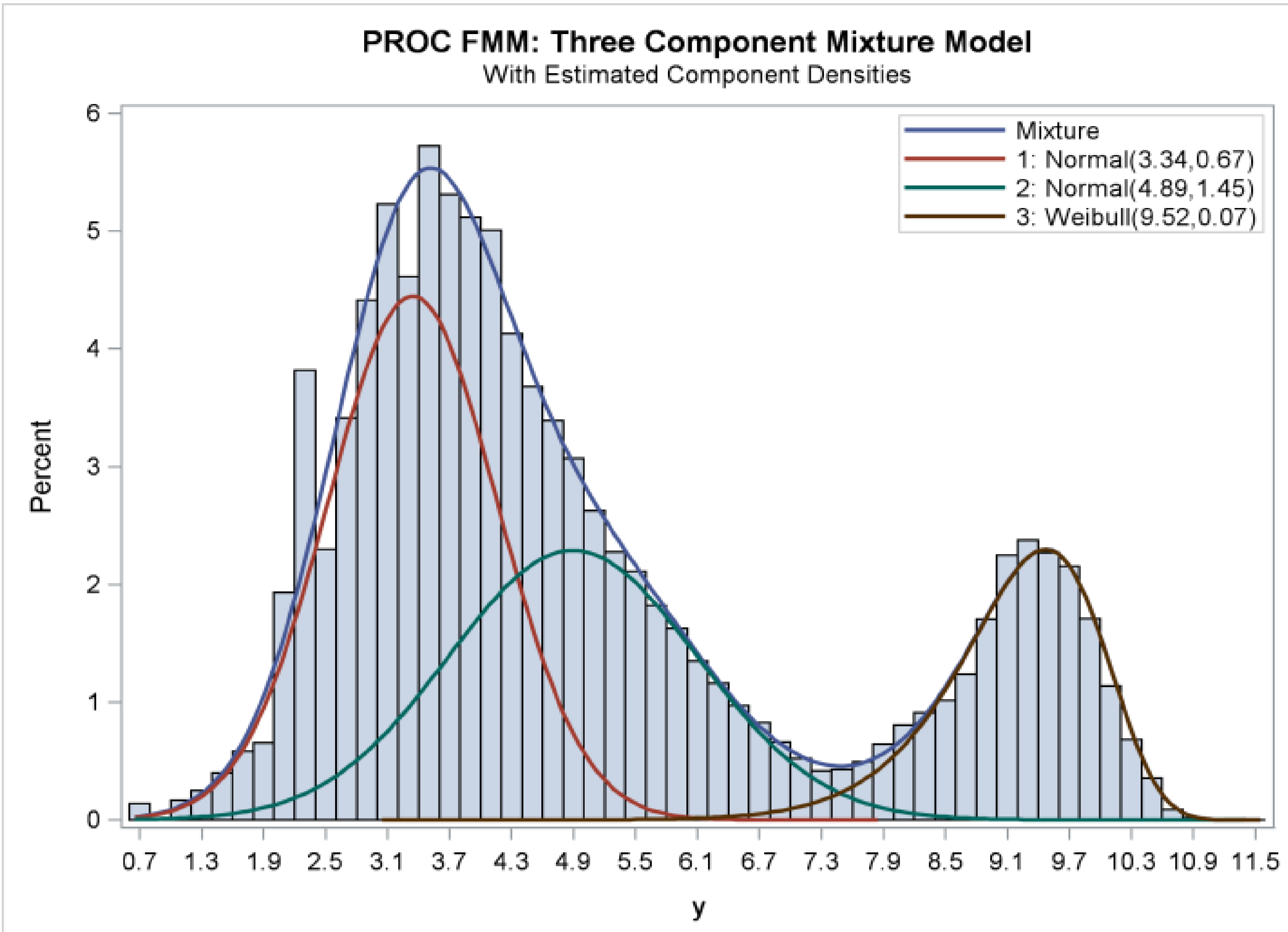
[2] Both the type-specific distributions $f(\mathbf{Y} \mid \omega)$ and the mixing distributions $\pi(\omega)$ are nonparametrically specified.

INTRODUCTION: Example. Finite Mixture of Normals (Parametric)

- $\mathbf{Y} = Y_1$ (single variable).

$$P\left(Y_1\right) = \sum_{\omega=1}^{L} \pi_\omega \, \frac{1}{\sigma_\omega} \phi\left(\frac{Y_1 - \mu_\omega}{\sigma_\omega}\right)$$

In this case, the identification is based on the shape of the distribution $P\left(Y_1\right)$.



PROC FMM: Three Component Mixture Model
With Estimated Component Densities
Mixture
1: Normal(3.34,0.67)
2: Normal(4.89,1.45)
3: Weibull(9.52,0.07)
Percent
0
1
2
3
4
5
6
0.7 1.3 1.9 2.5 3.1 3.7 4.3 4.9 5.5 6.1 6.7 7.3 7.9 8.5 9.1 9.7 10.3 10.9 11.5
y

INTRODUCTION: Example. Panel data.

• $\mathbf{Y} = (Y_1, Y_2, ..., Y_T)$ is the history of log-earnings of an individual over T periods of time.

• There are L types of individuals according to the stochastic process for the history of earnings:

$$P(Y_1, Y_2, ..., Y_T) = \sum_{\omega=1}^{L} \pi_\omega \ f_\omega(Y_1, Y_2, ..., Y_T)$$

INTRODUCTION: Example. Market entry

• There are T firms that are potential entrants in a market. $\mathbf{Y} = (Y_1, Y_2, ..., Y_T)$ with $Y_t \in \{0, 1\}$ is the vector with the entry decisions of the T firms.

• The researcher observes these T firms making entry decisions at N independent markets.

• There are L types of markets according to unobservable market characteristics affecting entry decisions.

$$P\left(Y_1, Y_2, ..., Y_T\right) = \sum_{\omega=1}^{L} \pi_\omega \; f_\omega \left(Y_1, Y_2, ..., Y_T\right)$$

2. ML ESTIMATION OF FM MODELS

• Consider a (semiparametric) FM model with $P(\mathbf{Y}_n) = \sum_{\omega=1}^{L} \pi_\omega \, f_\omega \left(\mathbf{Y}_n; \beta_\omega\right)$. The vector of parameters $\boldsymbol{\theta} \equiv (\boldsymbol{\pi}, \boldsymbol{\beta}) = (\pi_\omega, \beta_\omega : \omega = 1, 2, ..., L)$. And the log-likelihood function is:

$$\ell\left(\boldsymbol{\theta}\right) = \sum_{n=1}^{N} \ell_n\left(y_n, \boldsymbol{\theta}\right)$$

where $\ell_n\left(y_n, \boldsymbol{\theta}\right)$ is the contribution of observation n to the log-likelihood.

$$\ell_n\left(y_n, \boldsymbol{\theta}\right) = \sum_{\omega=1}^{L} \pi_\omega \; \log f_\omega\left(y_n, \beta_\omega\right)$$

• Maximization of this function w.r.t. $\boldsymbol{\theta}$ is a computationally complex task, i.e., many local maxima.

MLE ESTIMATION: EM ALGORITHM

• The EM (Expectation-Maximization) algorithm is an iterative method for the maximization of the MLE in finite mixture models. It is a very robust method in the sense that, under very mild conditions, each iteration improves the LF.

• To describe the EM algorithm an its properties, it is convenient to obtain an alternative description of the log-likelihood function.

• First, for arbitrary parameters $\boldsymbol{\theta}$, define the posterior probabilities $\pi_{\omega,n}^{post}(\boldsymbol{\theta})$, such that:

$$\pi_{\omega,n}^{post}(\boldsymbol{\theta}) \equiv P(\omega|y_n, \boldsymbol{\theta}) = \frac{\pi_\omega \; f_\omega(y_n; \beta_\omega)}{\sum_{\omega'=1}^{L} \pi_{\omega'} \; f_{\omega'}(y_n; \beta_{\omega'})}$$

MLE ESTIMATION: EM ALGORITHM [2]

• Second, note that $P(\omega_n, y_n|\boldsymbol{\theta}) = P(\omega_n|y_n, \boldsymbol{\theta})\ P(y_n|\boldsymbol{\theta})$. Therefore,

$$\ell_n(y_n, \boldsymbol{\theta}) \equiv \log P(y_n|\boldsymbol{\theta}) = \log P(\omega_n, y_n|\boldsymbol{\theta}) - \log \pi_{\omega,n}^{post}(\boldsymbol{\theta})$$

• Integrating the RHS over the posterior distribution $\{\pi_{\omega,n}^{post}(\boldsymbol{\theta}) : \omega = 1, 2, ..., L\}$, we get:

$$\begin{aligned} \ell_n(y_n, \boldsymbol{\theta}) &= \left(\sum_{\omega=1}^{L} \pi_{\omega,n}^{post}(\boldsymbol{\theta})\ \log P(\omega, y_n|\boldsymbol{\theta})\right) \\ &- \left(\sum_{\omega=1}^{L} \pi_{\omega,n}^{post}(\boldsymbol{\theta})\ \log \pi_{\omega,n}^{post}(\boldsymbol{\theta})\right) \end{aligned}$$

MLE ESTIMATION: EM ALGORITHM [3]

- And the log-likelihood function can be written as:

$$\ell\left(\boldsymbol{\theta}\right) = \left(\sum_{n=1}^{N}\sum_{\omega=1}^{L}\pi_{\omega,n}^{post}\left(\boldsymbol{\theta}\right) \left[\log \pi_{\omega} + \log f_{\omega}\left(y_n, \theta_{\omega}\right)\right]\right)$$

$$- \left(\sum_{n=1}^{N}\sum_{\omega=1}^{L}\pi_{\omega,n}^{post}\left(\boldsymbol{\theta}\right) \log \pi_{\omega,n}^{post}\left(\boldsymbol{\theta}\right)\right)$$

MLE ESTIMATION: EM ALGORITHM [4]

- Then, we can write the log-likelihood function as:

$$\ell(\boldsymbol{\theta}) = Q\left(\boldsymbol{\theta};\, \boldsymbol{\pi}^{post}(\boldsymbol{\theta})\right) - R\left(\boldsymbol{\pi}^{post}(\boldsymbol{\theta})\right)$$

with

$$\begin{aligned} Q\left(\boldsymbol{\theta};\, \boldsymbol{\pi}^{post}(\boldsymbol{\theta})\right) &= \sum_{n=1}^{N}\sum_{\omega=1}^{L} \pi_{\omega,n}^{post}(\boldsymbol{\theta})\ \left[\log \pi_\omega + \log f_\omega(y_n, \theta_\omega)\right] \\ R\left(\boldsymbol{\pi}^{post}(\boldsymbol{\theta})\right) &= \sum_{n=1}^{N}\sum_{\omega=1}^{L} \pi_{\omega,n}^{post}(\boldsymbol{\theta})\ \log \pi_{\omega,n}^{post}(\boldsymbol{\theta}) \end{aligned}$$

- Keeping the posterior probabilities $\{\pi_{\omega,n}^{post}\}$ constant at arbitrary values, we have the Pseudo-Likelihood function:

$$Q\left(\boldsymbol{\theta},\, \boldsymbol{\pi}^{post}\right) = \sum_{n=1}^{N}\sum_{\omega=1}^{L} \pi_{\omega,n}^{post}\ \left[\log \pi_\omega + \log f_\omega(y_n, \beta_\omega)\right]$$

MLE ESTIMATION: EM ALGORITHM [5]

• Given initial values $\widehat{\boldsymbol{\theta}}^0$, and iteration of the EM algorithm makes two different steps in order to obtain new values $\widehat{\boldsymbol{\theta}}^1$.

(1) **Expectation Step:** Computes the posterior probabilities

$\pi_{\omega,n}^{post,0} = \pi_{\omega,n}^{post}\left(\widehat{\boldsymbol{\theta}}^0\right)$ for every ω and n.

(2) **Maximization Step:** Maximization of the pseudo log-likelihood $Q\left(\boldsymbol{\theta};\, \boldsymbol{\pi}^{post,0}\right)$ with respect to $\boldsymbol{\theta}$, keeping $\boldsymbol{\pi}^{post,0}$ fixed.

EM ALGORITHM: Expectation Step

• Given initial values $\widehat{\boldsymbol{\theta}}^{0}$, we construct the posterior mixing probabilities $\pi_{\omega,n}^{post}$ for any ω and any observation n in the sample:

$$\pi_{\omega,n}^{post} = \frac{\widehat{\pi}_{\omega}^{0} \; f_{\omega}\left(y_n; \widehat{\beta}_{\omega}^{0}\right)}{\sum_{\omega'=1}^{L} \widehat{\pi}_{\omega'}^{0} \; f_{\omega'}\left(y_n; \widehat{\beta}_{\omega'}^{0}\right)}$$

EM ALGORITHM: Maximization Step w.r.t. $\boldsymbol{\pi}$

• Taking the posterior probabilities $\{\pi_{\omega,n}^{post}\}$ fixed, we maximize $Q\left(\boldsymbol{\theta};\, \boldsymbol{\pi}^{post}\right) = \sum_{n=1}^{N} \sum_{\omega=1}^{L} \pi_{\omega,n}^{post} \left[\log \pi_\omega + \log f_\omega\left(y_n, \beta_\omega\right)\right]$ with respect to $\boldsymbol{\pi}$.

• It is straightforward to show that the vector $\widehat{\boldsymbol{\pi}}^1$ that maximizes $Q\left(\boldsymbol{\theta};\, \boldsymbol{\pi}^{post}\right)$ with respect to $\boldsymbol{\pi}$ is:

$$\widehat{\pi}_\omega^1 = \frac{1}{N} \sum_{n=1}^{N} \pi_{\omega,n}^{post} = \frac{1}{N} \sum_{n=1}^{N} \frac{\widehat{\pi}_\omega^0 \; f_\omega\left(y_n; \widehat{\beta}_\omega^0\right)}{\sum_{\omega'=1}^{L} \widehat{\pi}_{\omega'}^0 \; f_{\omega'}\left(y_n; \widehat{\beta}_{\omega'}^0\right)}$$

EM ALGORITHM: Maximization Step w.r.t. $\boldsymbol{\beta}$

• Taking the posterior probabilities $\{\pi_{\omega,n}^{post}\}$ fixed, we maximize $Q\left(\boldsymbol{\theta};\, \boldsymbol{\pi}^{post}\right) = \sum_{n=1}^{N} \sum_{\omega=1}^{L} \pi_{\omega,n}^{post} \left[\log \pi_{\omega} + \log f_{\omega}\left(y_{n}, \beta_{\omega}\right)\right]$ with respect to $\boldsymbol{\beta}$.

• For every value ω, the new value $\widehat{\beta}_{\omega}^{1}$ solves the likelihood equations:

$$\sum_{n=1}^{N} \pi_{\omega,n}^{post} \frac{\partial \log f_{\omega}\left(y_{n}, \widehat{\beta}_{\omega}^{1}\right)}{\partial \beta_{\omega}} = 0$$

• In many applications, this type-specific log-likelihood is easy to maximize (e.g., it is globally concave).

EM ALGORITHM: Example 1 (Mixture of Normals)

• Suppose that Y is a FM of L normal random variables with different means and known unit variance. We want to estimate $\boldsymbol{\pi}$ and $\boldsymbol{\beta} = (\mu_1, \mu_2, ..., \mu_L)$.

$$Q\left(\boldsymbol{\theta},\, \boldsymbol{\pi}^{post}\right) = \sum_{n=1}^{N} \sum_{\omega=1}^{L} \pi_{\omega,n}^{post} \; \left[\log \pi_\omega + \log \phi\left(y_n - \mu_\omega\right)\right]$$

• Expectation Step:

$$\pi_{\omega,n}^{post} = \frac{\widehat{\pi}_{\omega}^{0} \; \phi\left(y_n - \widehat{\mu}_{\omega}^{0}\right)}{\sum_{\omega'=1}^{L} \widehat{\pi}_{\omega'}^{0} \; \phi\left(y_n - \widehat{\mu}_{\omega'}^{0}\right)}$$

EM ALGORITHM: Example 1 [cont]

- Maximization Step:

$$
\begin{aligned}
\widehat{\pi}_{\omega}^{1} &= \frac{1}{N}\sum_{n=1}^{N} \pi_{\omega,n}^{post} \\
\widehat{\mu}_{\omega}^{1} &= \frac{\sum_{n=1}^{N} \pi_{\omega,n}^{post}\ y_n}{\sum_{n=1}^{N} \pi_{\omega,n}^{post}}
\end{aligned}
$$

EM ALGORITHM: Example 2 (T Bernoullis, Mixture of i.i.d. Bernoullis)

• Suppose that $\mathbf{Y} = (Y_1, Y_2, ..., Y_T)$ is a vector of binary variables, $Y_t \in \{0, 1\}$. Conditional on ω, these T variables are $i.i, d.$ Bernoulli with probability β_ω.

$$P\left(\mathbf{y}_n\right) = \sum_{\omega=1}^{L} \pi_\omega \left[\beta_\omega\right]^{T_n^1} \; \left[1 - \beta_\omega\right]^{T-T_n^1}$$

with $T_n^1 = \sum_{t=1}^{T} y_{tn}$.

• We want to estimate $\boldsymbol{\pi}$ and $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_L)$.

$$Q\left(\boldsymbol{\theta}, \, \boldsymbol{\pi}^{post}\right) = \sum_{n=1}^{N} \sum_{\omega=1}^{L} \pi_{\omega,n}^{post} \; \left[\log \pi_\omega + T_n^1 \log \beta_\omega + \left(T - T_n^1\right) \log \left[1 - \beta_\omega\right]\right]$$

EM ALGORITHM: Example 2 [cont]

- Expectation Step:

$$\pi_{\omega,n}^{post} = \frac{\widehat{\pi}_{\omega}^{0} \left[\widehat{\beta}_{\omega}^{0}\right]^{T_n^1} \left[1 - \widehat{\beta}_{\omega}^{0}\right]^{T - T_n^1}}{\sum_{\omega'=1}^{L} \widehat{\pi}_{\omega'}^{0} \left[\widehat{\beta}_{\omega'}^{0}\right]^{T_n^1} \left[1 - \widehat{\beta}_{\omega'}^{0}\right]^{T - T_n^1}}$$

- Maximization Step:

$$\widehat{\pi}_{\omega}^{1} = \frac{1}{N} \sum_{n=1}^{N} \pi_{\omega,n}^{post}$$

$$\widehat{\beta}_{\omega}^{1} = \frac{\sum_{n=1}^{N} \pi_{\omega,n}^{post} \left[\frac{T_n^1}{T}\right]}{\sum_{n=1}^{N} \pi_{\omega,n}^{post}}$$

EM ALGORITHM: Example 3 (T Multinom., Mixture of i.i.d. Multinom.)

• Suppose that $\mathbf{Y} = (Y_1, Y_2, ..., Y_T)$ is a vector of multinomial variables, $Y_t \in \{0, 1, ..., J\}$. Conditional on ω, these T variables are $i.i, d.$ multinomial with vector of probabilities $\beta_\omega = (\beta_{\omega,1}, \beta_{\omega,2}, ..., \beta_{\omega,J})$.

$$P(\mathbf{y}_n) = \sum_{\omega=1}^{L} \pi_\omega \left[\beta_{\omega,1}\right]^{T_n^1} ... \left[\beta_{\omega,J}\right]^{T_n^J} \left[1 - \textstyle\sum_{j=1}^J \beta_{\omega,j}\right]^{T - \sum_{j=1}^J T_n^j}$$

with $T_n^j = \sum_{t=1}^T 1\{y_{tn} = j\}$.

• We want to estimate $\boldsymbol{\pi}$ and $\boldsymbol{\beta} = (\beta_{\omega,j} : \omega = 1, 2, ..., L;\ j = 1, 2, ..., J)$.

$$Q\left(\boldsymbol{\theta},\ \boldsymbol{\pi}^{post}\right) = \sum_{n=1}^{N} \sum_{\omega=1}^{L} \pi_{\omega,n}^{post} \left[\log \pi_\omega + \sum_{j=0}^{J} T_n^j \log \beta_{\omega,j}\right]$$

EM ALGORITHM: Example 3 [cont.]

- Expectation Step:

$$\pi_{\omega,n}^{post} = \frac{\widehat{\pi}_{\omega}^{0} \left[\widehat{\beta}_{\omega,0}^{0}\right]^{T_n^0} \left[\widehat{\beta}_{\omega,1}^{0}\right]^{T_n^1} \cdots \left[\widehat{\beta}_{\omega,J}^{0}\right]^{T_n^J}}{\sum_{\omega'=1}^{L} \widehat{\pi}_{\omega'}^{0} \left[\widehat{\beta}_{\omega',0}^{0}\right]^{T_n^0} \left[\widehat{\beta}_{\omega',1}^{0}\right]^{T_n^1} \cdots \left[\widehat{\beta}_{\omega',J}^{0}\right]^{T_n^J}}$$

- Maximization Step:

$$\widehat{\pi}_{\omega}^{1} = \frac{1}{N} \sum_{n=1}^{N} \pi_{\omega,n}^{post}$$

$$\widehat{\beta}_{\omega}^{j} = \frac{\sum_{n=1}^{N} \pi_{\omega,n}^{post} \left[\frac{T_n^j}{T}\right]}{\sum_{n=1}^{N} \pi_{\omega,n}^{post}}$$

EXERCISE:

• Consider a FM for $\mathbf{Y} = (Y_1, Y_2, Y_3)$, with $Y_t \in \{0, 1, 2\}$, and with $\omega \in \{1, 2\}$. Conditional on ω, the three variables (Y_1, Y_2, Y_3) are $i.i.d.$ multinomial distributed with parameters $\beta_{\omega,0}$, $\beta_{\omega,1}$, $\beta_{\omega,2}$. The values of the parameters are:

$$\pi_1 = 0.2; \quad \beta_{\omega=1,0} = 0.1; \quad \beta_{\omega=1,1} = 0.3; \quad \beta_{\omega=1,2} = 0.6;$$

$$\pi_2 = 0.8; \quad \beta_{\omega=2,0} = 0.5; \quad \beta_{\omega=1,1} = 0.4; \quad \beta_{\omega=1,2} = 0.1;$$

• Write program code that generates $N = 1000$ observations $\mathbf{y}_n = (y_{1n}, y_{2n}, y_{3n})$ from this distribution.

• Write program code that implements the EM-algorithm for these (simulated) and obtain estimates of the parameters of the model $(\pi, \beta_{\omega,j})$.

EM ALGORITHM: Monotonicity and Convergence

• Let $\{\boldsymbol{\theta}^{(k)} : k \geq 0\}$ be the sequence of parameters generated by the EM algorithm given an arbitrary initial value $\boldsymbol{\theta}^{(0)}$.

• In the original paper that proposed the EM algorithm, **Dempster, Laird, and Rubin (JRSS, 1977)** showed that [by construction] the likelihood function is monotonically increasing in this sequence:

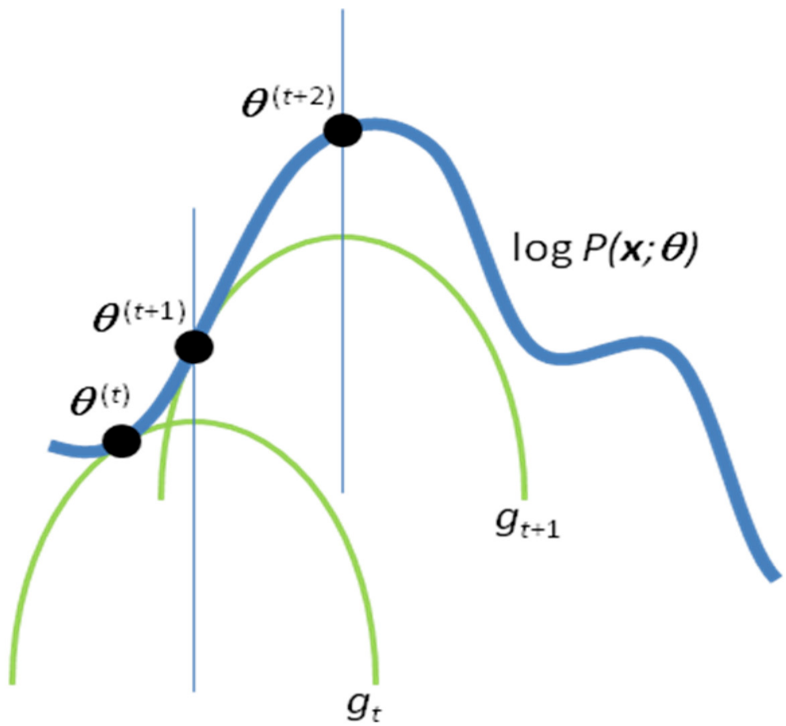
$$\ell\left(\boldsymbol{\theta}^{(k+1)}\right) \geq \ell\left(\boldsymbol{\theta}^{(k)}\right) \quad \text{for any } k \geq 0$$

• In a compact parameter space Θ, this property implies that the sequence $\{\boldsymbol{\theta}^{(k)} : k \geq 0\}$ converges to some value $\boldsymbol{\theta}^* \in \Theta$.

EM ALGORITHM: Monotonicity and Convergence [2]

- Wu (AS, 1983) shows that if the likelihood is continuous in $\boldsymbol{\theta}$, then the limit value $\boldsymbol{\theta}^*$ is a local maximum.

- Convergence to the global maximum requires stronger conditions.



Supplementary Figure 1 Convergence of the EM algorithm. Starting from initial parameters $\boldsymbol{\theta}^{(t)}$, the E-step of the EM algorithm constructs a function g_t that lower-bounds the objective function $\log P(\boldsymbol{x};\boldsymbol{\theta})$. In the M-step, $\boldsymbol{\theta}^{(t+1)}$ is computed as the maximum of g_t. In the next E-step, a new lower-bound g_{t+1} is constructed; maximization of g_{t+1} in the next M-step gives $\boldsymbol{\theta}^{(t+2)}$, etc.

3. IDENTIFICATION OF NPFM MODELS: Basics [1]

• We have implicitly assumed that the vector of parameters $\boldsymbol{\theta}$ is point identified, i.e., there is a unique value $\boldsymbol{\theta} \in \Theta$ that maximizes the likelihood function.

• This is not necessarily the case. There are many simple examples where the model is not identified.

• We concentrate on the identification of NPFM models where $\mathbf{Y}$ is discrete.

• More specifically: $\mathbf{Y} = (Y_1, Y_2, ..., Y_T)$ with $Y_t \in \{1, 2, ..., J\}$, such that $\mathbf{Y} \in \{1, 2, ..., J\}^T$ and can take J^T values.

• For discrete $\mathbf{Y}$, the NP specification of the type-specific probability functions $f_\omega(\mathbf{Y})$ implies an unrestricted multinomial distribution: $f_\omega(\mathbf{y}) = \beta_{\omega,\mathbf{y}}$.

IDENTIFICATION: Basics [2]

• **Without further assumptions this model is not identified**. To see this, note that the model can be described in terms of the following restrictions: for any $\mathbf{y} \in \{1, 2, ..., J\}^T$

$$P(\mathbf{y}) = \sum_{\omega=1}^{L} \pi_\omega \ f_\omega \left(\mathbf{y}, \boldsymbol{\beta}_\omega\right)$$

• The number of restrictions is $J^T - 1$, while the number of free parameters is $L - 1$ (from $\pi_\omega' s$) and $L \left[J^T - 1\right]$. The order condition for identification requires:

$$J^T - 1 \geq L - 1 + L \left[J^T - 1\right]$$

It is clear that this condition never holds for any $L \geq 2$.

• We need some to impose some restrictions on $f_\omega \left(\mathbf{y}, \boldsymbol{\beta}_\omega\right)$.

IDENTIFICATION: Basics [3]

• We will consider identification of NPFM models under four different types of assumptions. Let $\mathbf{Y} = (Y_1, Y_2, ..., Y_T)$

[1] **Conditional i.i.d.**

$$f_\omega(\mathbf{y}, \boldsymbol{\beta}_\omega) = \prod_{t=1}^{T} p_\omega(y_t, \boldsymbol{\beta}_\omega) = \prod_{t=1}^{T} \prod_{j=1}^{J} \left[\beta_{\omega,j}\right]^{1\{y_t=j\}}$$

[2] **Conditional independence**

$$f_\omega(\mathbf{y}, \boldsymbol{\beta}_\omega) = \prod_{t=1}^{T} p_{\omega,t}\left(y_t, \boldsymbol{\beta}_{\omega,t}\right) = \prod_{t=1}^{T} \prod_{j=1}^{J} \left[\beta_{\omega,t,j}\right]^{1\{y_t=j\}}$$

IDENTIFICATION: Basics [4]

[3] **Conditional homogeneous Markov**

$$
\begin{aligned}
f_\omega\left(\mathbf{y}, \boldsymbol{\beta}_\omega\right) &= p_\omega(y_1) \prod_{t=2}^{T} p_\omega\left(y_t \mid y_{t-1}, \boldsymbol{\beta}_\omega\right) \\
&= p_\omega(y_1) \prod_{t=2}^{T} \prod_{j=1}^{J} \left[\beta_{\omega,j}(y_{t-1})\right]^{1\{y_t=j\}}
\end{aligned}
$$

[4] **Conditional non-homogeneous Markov**

$$
\begin{aligned}
f_\omega\left(\mathbf{y}, \boldsymbol{\beta}_\omega\right) &= p_{\omega,1}(y_1) \prod_{t=2}^{T} p_{\omega,t}\left(y_t \mid y_{t-1}, \boldsymbol{\beta}_{\omega,t}\right) \\
&= p_{\omega,1}(y_1) \prod_{t=2}^{T} \prod_{j=1}^{J} \left[\beta_{\omega,j,t}(y_{t-1})\right]^{1\{y_t=j\}}
\end{aligned}
$$

IDENTIFICATION: Basics [5]

• The previous discussion implicitly assumes that the researcher knows the true number of mixtures L. This is quite uncommon.

• We will study the identification L and present identification results and tests for a lower bound on L.

IDENTIFICATION: EM Algorithm when the model is not identified

• When a model is not identified, standard gradient search algorithms that maximize the likelihood function $\ell(\boldsymbol{\theta})$ (e.g., Newton methods, BHHH) do never converge and eventually reach points where a matrix is singular, e.g., the Hessian matrix or the matrix of the outer-product of the scores.

• **"Unfortunately"**, this is not the case when using the EM algorithm. The EM algorithm will converge to a point even if the model is not identified. In fact, it will converge very quickly.

• Of course, the convergence point depends on the initial value $\boldsymbol{\theta}^{(0)}$. Different initial values will return different convergence points for the EM algorithm.

• Therefore, one needs to be very careful when using the EM algorithm. The research needs to verify first that identification conditions hold.

EM Algorithm when the model is not identified   Example

• $\mathbf{Y} \in \{0, 1\}$ is a single Bernoulli random variable $(T = 1)$. There is only one free probability in the distribution of $\mathbf{Y}$, i.e., $P(y = 1)$. The sample is $\{y_n : n = 1, 2, ..., N\}$. Model:

$$P(y_n = 1) = \sum_{\omega=1}^{L} \pi_\omega \ \ [\beta_\omega]^{y_n} \ \ [1 - \beta_\omega]^{1-y_n}$$

The vector of model parameters is $\boldsymbol{\theta} = (\boldsymbol{\pi}, \boldsymbol{\beta}) = (\pi_\omega, \beta_\omega : \omega = 1, 2, ..., L)$.

• It is clear that the model is not identified for any $L \geq 2$, i.e., 1 restriction and $2L - 1$ parameters.

EM Algorithm when the model is not identified   Example

• However, given an arbitrary initial value $\boldsymbol{\theta}^0$, the EM algorithm always converges in one iteration to the following estimates of π_ω and β_ω: [Exercise: Prove this]

$$\widehat{\pi}_\omega = \frac{N_0}{N}\left[\frac{\pi^0_\omega(1-\beta^0_\omega)}{\sum_{\omega'=1}^{L}\pi^0_{\omega'}(1-\beta^0_{\omega'})}\right] + \frac{N_1}{N}\left[\frac{\pi^0_\omega\beta^0_\omega}{\sum_{\omega'=1}^{L}\pi^0_{\omega'}\beta^0_{\omega'}}\right]$$

$$\widehat{\beta}_\omega = \frac{\left[\frac{\pi^0_\omega\beta^0_\omega}{\sum_{\omega'=1}^{L}\pi^0_{\omega'}\beta^0_{\omega'}}\right]N_1}{\left[\frac{\pi^0_\omega(1-\beta^0_\omega)}{\sum_{\omega'=1}^{L}\pi^0_{\omega'}(1-\beta^0_{\omega'})}\right]N_0 + \left[\frac{\pi^0_\omega\beta^0_\omega}{\sum_{\omega'=1}^{L}\pi^0_{\omega'}\beta^0_{\omega'}}\right]N_1}$$

• Note that these estimates depend on the initial values. Note also that the posterior probabilities $\{\pi^{post}_{\omega,n}\}$ remain at their initial values.

4. IDENTIFICATION UNDER CONDITIONAL INDEPENDENCE

• We start with a model where the T variables $(Y_1, Y_2, ..., Y_T)$ are i.i.d. conditional on ω. Later we relax the assumption of identical distribution.

• We follow Bonhomme, Jochmans, and Robin (JRRS, 2016) but concentrate on a model with discrete variables Y_t. They present results for both discrete and continuous observable variables.

• Model:

$$P(y_1, y_2, ..., y_T) = \sum_{\omega=1}^{L} \pi_\omega \; f_\omega(y_1) \; f_\omega(y_2) \; ... \; f_\omega(y_T)$$

where $y_t \in \{1, 2, ..., J\}$. L is known [more on this below].

• We have a sample $\{y_{1n}, y_{2n}, ..., y_{Tn} : n = 1, 2, ..., N\}$ with $N \to \infty$, and we are interested in the estimation of $\{\pi_\omega\}$ and $f_\omega(y)$ for any ω and y.

IDENTIFICATION UNDER CONDITIONAL INDEPENDENCE [2]

• First, it is important to note that the joint distribution $P(Y_1, Y_2, ..., Y_T)$ is fully nonparametrically identified from the sample $\{y_{1n}, y_{2n}, ..., y_{Tn} : n = 1, 2, ..., N\}$, i.e., it can be consistently estimated without imposing any restriction. We treat $P(.)$ as known to the researcher.

• Define the $J \times L$ matrix.

$$\mathbf{F} \equiv [\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_L] = \begin{bmatrix} f_1(1) & f_2(1) & \cdots & f_L(1) \\ f_1(2) & f_2(2) & \cdots & f_L(2) \\ \vdots & \vdots & & \vdots \\ f_1(J) & f_2(J) & \cdots & f_L(J) \end{bmatrix}$$

ASSUMPTION 1: Matrix $\mathbf{F}$ is full column rank. [Note that this assumption implies that $L \leq J$).

- We show below that Assumption 1:

(1) is easily testable from the data;

(2) is a necessary and (with $T \geq 3$) sufficient condition for identification.

IDENTIFICATION UNDER CONDITIONAL INDEPENDENCE [2]

- Suppose that $T \geq 3$. Let $(t_1,\ t_2,\ t_3)$ be the indexes of three of the T variables (any 3 of the T variables). For arbitrary $y \in \{1, 2, ..., J\}$, define the $J \times J$ matrix:

$$\mathbf{A}(y) \equiv \left[a_{ij}(y)\right] = \left[\Pr\left(y_{t_1} = i\ ,\ y_{t_2} = j \mid y_{t_3} = y\right)\right]$$

- The model implies that (with $p(y) \equiv \Pr(y_t = 1)$):

$$\begin{aligned}
a_{ij}(y) &= \sum_{\omega=1}^{L} \Pr(\omega | y_{m_3} = y)\ \ \Pr(y_{m_1} = i\ ,\ y_{m_2} = j \mid \omega, y_{m_3} = y) \\
&= \sum_{\omega=1}^{L} \pi_\omega \frac{1}{p(y)}\ f_\omega(i)\ \ f_\omega(j)\ \ f_\omega(y) \\
&= \left[\begin{array}{ccc} f_1(i) & \cdots & f_L(i) \end{array}\right]\ diag\left[\pi_\omega\right]\ diag\left[\frac{f_\omega(y)}{p(y)}\right]\ \left[\begin{array}{c} f_1(j) \\ f_2(j) \\ \vdots \\ f_L(j) \end{array}\right]
\end{aligned}$$

IDENTIFICATION UNDER CONDITIONAL INDEPENDENCE [3]

• And in matrix form, we have that:

$$\underset{(J \times J)}{\mathbf{A}(y)} = \underset{(J \times L)}{\mathbf{F}} \; \underset{(L \times L)}{\mathbf{\Pi}^{1/2}} \; \underset{(L \times L)}{\mathbf{D}(y)} \; \underset{(L \times L)}{\mathbf{\Pi}^{1/2}} \; \underset{(L \times J)}{\mathbf{F}'}$$

where $\mathbf{\Pi} = diag\left[\pi_\omega\right]$, and $\mathbf{D}(y) = diag\left[\frac{f_\omega(y)}{p(y)}\right]$.

• The matrix in the LHS is identified. The matrices in the RHS depend of parameters π_ω and $f_\omega\left(y\right)$ that we want to identify.

• Define $J \times J$ matrix $\mathbf{A} \equiv \mathbb{E}\left[\mathbf{A}(y)\right] = \sum_{y=0}^{J} p(y)\ \mathbf{A}(y)$.

LEMMA: Matrix $\mathbf{F}$ has full column rank if and only if $rank(\mathbf{A}) = L$.

• We will see how this result provides a direct test of identification.

IDENTIFICATION UNDER CONDITIONAL INDEPENDENCE [4]

- **Proof of Lemma:**

- By definition, $\mathbf{A} = \mathbf{F}\ \mathbf{\Pi}^*\ \mathbf{F}'$, where $\mathbf{\Pi}^*$ is the diagonal matrix

$$\mathbf{\Pi}^* = \mathbf{\Pi}^{1/2} diag \left[\mathbb{E} \left(\frac{f_\omega(y)}{p(y)} \right) \right] \mathbf{\Pi}^{1/2}$$

- Since $\mathbf{\Pi}^*$ is a diagonal matrix with elements different than zero, and $\mathbf{A} = \mathbf{F}$ $\mathbf{\Pi}^*\ \mathbf{F}'$, we have that the $rank(\mathbf{A})$ is equal to the number of linearly independent columns of $\mathbf{F}$, such that $rank(\mathbf{A}) \leq L$. And in particular, $rank(\mathbf{A}) = L$ if and only if $rank(\mathbf{F}) = L$.

IDENTIFICATION UNDER CONDITIONAL INDEPENDENCE [5]

THEOREM: Under Assumption 1 (that implies $L \leq J$) and $T \geq 3$, all the parameters of the model $\{\pi_\omega\}$ and $\{f_\omega(y)\}$ are point identified.

• **Proof of Theorem:** The proof proceeds in three steps: (1) identification of diagonal matrix $\mathbf{D}(y)$; (2) identification of $f_\omega(y)$; and (3) identification of π_ω. The proof is constructive, and as we will see later it provides a simple sequential estimator.

• **[1] Identification of diagonal matrix** $\mathrm{D}(y)$.

- Since $\mathbf{A}$ is a square $(J \times J)$, symmetric, and real matrix, it admits an eigenvalue decomposition: $\mathbf{A} = \mathbf{V}\ \mathbf{\Lambda}\ \mathbf{V}'$.

[1] Identification of diagonal matrix $\mathrm{D}(y)$. [cont.]

• Since $rank(\mathbf{A}) = L \leq J$, only L of the eigenvalues in the diagonal matrix $\mathbf{\Lambda}$ are different to zero. Therefore, $\mathbf{A} = \mathbf{V}_L \ \mathbf{\Lambda}_L \ \mathbf{V}'_L$, where $\mathbf{\Lambda}_L$ is the $L \times L$ diagonal matrix with non-zero eigenvalues, and $\mathbf{V}_L$ is the $J \times L$ matrix of eigenvectors such that $\mathbf{V}'_L \mathbf{V}_L = \mathbf{I}_L$.

• Define the $L \times J$ matrix $\mathbf{W} = \mathbf{\Lambda}_L^{-1/2} \mathbf{V}'_L$. So far, all the matrix decompositions are based on matrix $\mathbf{A}$. So it is clear that matrix $\mathbf{W}$ is identified.

• Matrix $\mathbf{W}$ has a useful property. For any value of $y \in \{0, 1, ..., J\}$, we have that:

$$\begin{aligned}
\mathbf{W} \ \mathbf{A}(y) \ \mathbf{W}' &= \left[\mathbf{\Lambda}_L^{-1/2} \mathbf{V}'_L\right] \ \left[\mathbf{F} \ \mathbf{\Pi}^{1/2} \ \mathbf{D}(y) \ \mathbf{\Pi}^{1/2} \ \mathbf{F}'\right] \ \left[\mathbf{V}_L \ \mathbf{\Lambda}_L^{-1/2}\right] \\
&= \mathbf{U} \ \mathbf{D}(y) \ \mathbf{U}'
\end{aligned}$$

with $\mathbf{U} \equiv \mathbf{\Lambda}_L^{-1/2} \mathbf{V}'_L \ \mathbf{F} \ \mathbf{\Pi}^{1/2}$.

[1] Identification of diagonal matrix $\mathrm{D}(y)$. [cont.]

• It is straightforward to verify that matrix $\mathbf{U}$ is such that, $\mathbf{U}\mathbf{U}' = \mathbf{I}_L$. Therefore, the expression $\mathbf{W}\ \mathbf{A}(y)\ \mathbf{W}' = \mathbf{U}\ \mathbf{D}(y)\ \mathbf{U}'$ means that $\mathbf{U}\ \mathbf{D}(y)\ \mathbf{U}'$ is the eigenvalue-eigenvector decomposition of matrix $\mathbf{W}\ \mathbf{A}(y)\ \mathbf{W}'$.

- Since matrix $\mathbf{W}\ \mathbf{A}(y)\ \mathbf{W}'$ is identified, this implies that diagonal matrix is also identified.

- Note that the identification of the elements of $\mathbf{U}$ and $\mathbf{D}(y)$ is up-to-relabelling of the $\omega' s$ because any permutation of the columns of $\mathbf{U}$ and $\mathbf{D}(y)$ is a valid eigenvalue-eigenvector decomposition of matrix $\mathbf{W}\ \mathbf{A}(y)\ \mathbf{W}'$.

[2] Identification of $f_\omega(y)$.

• Remember that: $\mathbf{D}(y) = diag\left[\frac{f_\omega(y)}{p(y)}\right]$. Therefore, if $d_\omega(y)$ is the $\omega - th$ element in the main diagonal of matrix $\mathbf{D}(y)$, we have that:

$$f_\omega(y) = \mathbb{E}\left[d_\omega(y)\ 1\{y_t = y\}\right]$$

and $f_\omega(y)$ is identified. In other words, given $d_\omega(y)$ we can obtain a consistent estimator of $f_\omega(y)$ as:

$$\widehat{f}_\omega(y) = \frac{1}{NT}\sum_{n=1}^{N}\sum_{t=1}^{T} d_\omega(y_{nt})\ 1\{y_{nt} = y\}$$

• **[3] Identification of** π_ω.

• The model implies that,

$$p(y) = \sum_{\omega=1}^{L} \pi_\omega \; f_\omega \left(y \right)$$

• And in vector form:

$$\mathbf{p} = \mathbf{F} \; \boldsymbol{\pi}$$

where $\mathbf{p}$ is the $J \times 1$ vector of unconditional probabilities $(p(y) : y = 1, 2, ..., J)'$, and $\boldsymbol{\pi}$ is the $L \times 1$ vector of probability mixtures.

• Since $\mathbf{F}$ is full column rank, we have that $(\mathbf{F}'\mathbf{F})$ is non-singular and $\boldsymbol{\pi}$ can be uniquely identified as:

$$\boldsymbol{\pi} = \left(\mathbf{F}'\mathbf{F} \right)^{-1} \left(\mathbf{F}'\mathbf{p} \right)$$

5. ESTIMATION METHODS

• The previous proof of identification is constructive and it suggests the following sequential estimation procedure:

Step 1: Method of moments (frequency) estimation of the matrices $\mathbf{A}$ and $\mathbf{A}(y)$;

Step 2: Estimation (construction) of matrix $\mathbf{W}$ using an eigenvalue-eigenvector decomposition of matrix $\mathbf{A}$;

Step 3: Estimation (construction) of matrices $\mathbf{U}$ and $\mathbf{D}(y)$ using an eigenvalue-eigenvector decomposition of matrix $\mathbf{W}\ \mathbf{A}(y)\ \mathbf{W}'$;

Step 4: Method of moments estimation of $f_{\omega}(y)$ from the elements of diagonal matrix $\mathbf{D}(y)$;

Step 5: Least squares estimation of $\boldsymbol{\pi}$ as $(\mathbf{F}'\mathbf{F})^{-1}(\mathbf{F}'\mathbf{p})$.

ESTIMATION [2]

• This estimator is consistent and asymptotically normal (root-N when variables are discrete). It is also straightforward from a computational point of view (e.g., no problems of multiple local maxima or no convergence). But it is not asymptotically efficient. Also, the construction of valid asymptotic standard errors for this 5-step estimator using delta method is cumbersome. Bootstrap methods can be applied.

• Asymptotic efficiency can be achieved by applying 1-iteration of the BHHH method in maximization of the (nonparametric) likelihood function and using the consistent but inefficient estimator as the initial value. This one-step-efficient approach provides also correct asymptotic standard errors.

6. IDENTIFICATION AND TESTS OF THE NUMBER OF MIXTURES

- Kasahara and H., and K. Shimotsu (JRSS, 2014)
- Kasahara and Shimotsu (JASA, 2015)

7. IDENTIFICATION UNDER MARKOV STRUCTURE

- Kasahara and Shimotsu (ECMA, 2009)

8. IDENTIFICATION USING EXCLUSION RESTRICTIONS

• The previous identification results are based on the assumption of independence between the T variables $(Y_1, Y_2, ..., Y_T)$ once we condition on the unobserved type ω and possibly on observable exogenous variables $\mathbf{X}$.

• All the NP identification results using this conditional independence approach require $T \geq 3$, regardless the number of points in the support of Y_t.

• This is a very negative result because there are many interesting applications with $T = 2$ (two endogenous variables) where we can easily reject the null hypothesis of no unobserved heterogeneity, but we cannot identify a NPFM model using only the conditional independence assumption.

IDENTIFICATION USING EXCLUSION RESTRICTIONS [2]

• **Henry, Kitamura, and Salanie** (QE, 2014) propose an alternative approach to identify NPFM. Their approach is based on an exclusion restriction.

• Let Y be a scalar endogenous variable ($T = 1$) and let X and Z be observable exogenous variables. Consider the NPFM model:

$$\begin{aligned} P(Y \mid X, Z) &= \sum_{\omega=1}^{L} \Pr(\omega \mid X, Z) \;\; \Pr(Y \mid \omega, X, Z) \\ &= \sum_{\omega=1}^{L} \pi_{\omega}(X, Z) \;\; f_{\omega}(Y \mid X, Z) \end{aligned}$$

For notational simplicity, I will omit variable X (it does not play an important role) such that all the results can be interpreted as conditional on a particular value of X (i.e., X is discrete).

IDENTIFICATION USING EXCLUSION RESTRICTIONS [3]

• Model: $$P(Y \mid Z) = \sum_{\omega=1}^{L} \pi_\omega(Z)\ f_\omega(Y \mid Z)$$

ASSUMPTION [Exclusion Restriction]: $$f_\omega(Y \mid Z) = f_\omega(Y)$$

ASSUMPTION [Relevance]: There are values z_0 and z_1 in the support of Z such that $\pi_\omega(z_1) \neq \pi_\omega(z_0)$

• Variable Z enters in the mixing distribution π_ω but not in the component distributions f_ω. Similarly as with IV models, the identification strength of these assumptions depends on the strength of the dependence of $\pi_\omega(Z)$ on Z.

EXCLUSION RESTRICTION. Example 1. Misclassification Model

• The researcher is interested in the relationship between variables Y and ω where $\omega \in \{1, 2, ..., L\}$ is a categorical variable: $Pr(Y|\omega)$.

• However, ω is not observable, or is observable with error. The researcher observes the categorical variable $Z \in \{1, 2, ..., |\mathcal{Z}|\}$ that is a noisy measure of ω, i.e., there are misclasifications when using Z instead of ω.

• In this model, $\Pr(Y \mid \omega, Z) = \Pr(Y \mid \omega)$, i.e., given the correct category ω, the noisy category Z becomes redundant. [Exclusion Restriction].

• $\Pr(\omega \mid Z)$ depends on Z, i.e., Z is not complete noise and it contains some information about ω. [Relevance].

EXCLUSION RESTRICTION. Example 2. Demand Model

• Consider the following demand model using individual level data in **a single market**:

$$Y = d\left(X, \omega, \varepsilon\right)$$

Y = Quantity purchased of the product by a consumer;

X = Vector of exogenous consumer characteristics affecting demand: e.g., income, wealth, education, age, gender, etc.

ω = Unobserved consumer characteristics that can be correlated with X (endogenous unobservable)

ε = Unobserved consumer characteristics independent of (X, ω)

• The researcher is interested in the estimation of $\Pr\left(Y|X, \omega\right)$.

EXCLUSION RESTRICTION. Example 2. Demand Model

• Suppose that the researcher can classify consumers in **different groups, e.g., according to their geographic location / region**. Let Z be the observable variable that represents the geographic location of the consumer.

• [Exclusion Restriction]. $\Pr(Y \mid X, Z, \omega) = \Pr(Y \mid X, \omega)$, i.e., given (X, ω) a consumer's location is redundant to explain her demand. A single common market without transportation costs.

• [Relevance]. $\Pr(\omega \mid X, Z)$ depends on Z. After controlling for X, the unobservable ω a different probability distribution across locations.

EXCLUSION RESTRICTION. Example 3. Local Market Competition

• Game of oligopoly competition in a local market, e.g., game of market entry. Sample of M local markets. Model:

$$Y = g\left(X, \omega, \varepsilon\right)$$

$Y =$ Number of active firms in the local market;

$X =$ Vector of exogenous market characteristics: e.g., population, income, input prices, etc.

$\omega =$ Unobserved market characteristics that can be correlated with X (endogenous unobservable)

$\varepsilon =$ Unobserved consumer characteristics independent of (X, ω)

• The researcher is interested in the estimation of $\Pr\left(Y|X, \omega\right)$.

EXCLUSION RESTRICTION. Example 3. Local Market Competition

- Let Z_m be the average value of X in local markets nearby market m.

- [Exclusion Restriction]. $\Pr(Y \mid X, Z, \omega) = \Pr(Y \mid X, \omega)$, i.e., competition is independent across markets; given market characteristics (X, ω) the characteristics of other nearby markets Z are irrelevant.

- [Relevance]. $\Pr(\omega \mid X, Z)$ depends on Z. If ω is spatially correlated $(cov(\omega_m, \omega_{m'}) \neq 0)$ and and ω is correlated with X $(cov(\omega_{m'}, Z_{m'}) \neq 0)$, then $Z = X_{m'}$ may contain information about ω_m $(cov(\omega_m, X_{m'}) \neq 0)$.

Henry, Kitamura, and Salanie (HKS)

- Consider the model: $$P(Y \mid Z) = \sum_{\omega=1}^{L} \pi_{\omega}(Z)\, f_{\omega}(Y)$$

- They show that the parameters of the model, $\{\pi_{\omega}(Z),\ f_{\omega}(Y \mid Z)\}$ are identified up to $L(L-1)$ constants. These unknown constants belong to a compact space, and this implies that $\{\pi_{\omega}(Z),\ f_{\omega}(Y \mid Z)\}$ are partially identified. HKS derive the sharp bounds of the identified set.

- Under some additional conditions, the model can be point-identified.

- Here I illustrate these results for the case with $L = 2$ types or components.

Henry, Kitamura, and Salanie (HKS) [2]

• Consider the NPFM model with $L = 2$:

$$P(Y \mid Z) = [1 - \pi(Z)]\ f_0(Y) + \pi(Z)\ \ f_1(Y)$$

where Y and Z are scalar variables, and for simplicity suppose that they have discrete support.

• The model parameters are $\{\pi(z) : z \in \mathcal{Z}\}$ and $\{f_0(y),\ f_1(y) : y \in \mathcal{Y}\}$. $\#$ Parameters $= |\mathcal{Z}| + 2(|\mathcal{Y}| - 1)$.

• Restrictions: $\#$ free probs in $P(Y \mid Z) = (|\mathcal{Y}| - 1)\ |\mathcal{Z}|$.

• Order condition for point identification: $|\mathcal{Y}| \geq 3$ and $|\mathcal{Z}| \geq 2(|\mathcal{Y}| - 1)/(|\mathcal{Y}| - 2)$.

Henry, Kitamura, and Salanie (HKS) [3]

• Consider $y \in \mathcal{Y}$ (we show identification pointwise in y). Let z_0, $z_1 \in \mathcal{Z}$ be such that $\pi(z_0) \neq \pi(z_1)$. For convenience, let z_0 and z_1 be $z_0 = \arg\min_{z \in \mathcal{Z}} P(y \mid z)$ and $z_1 = \arg\max_{z \in \mathcal{Z}} P(y \mid z)$, such that $P(y \mid z_1) - P(y \mid z_0) > 0$ and it takes its maximum value.

• The model (and exclusion restriction) implies that:

$$P(y \mid z_1) - P(y \mid z_0) = [\pi(z_1) - \pi(z_0)]\,[f_1(y) - f_0(y)]$$

• And for any $z \in \mathcal{Z}$,

$$r(z) \equiv \frac{P(y \mid z) - P(y \mid z_0)}{P(y \mid z_1) - P(y \mid z_0)} = \frac{\pi(z) - \pi(z_0)}{\pi(z_1) - \pi(z_0)}$$

Note that for any $z \in \mathcal{Z}$, $r(z) \in [0, 1]$ with $r(z_0) = 0$ and $r(z_1) = 1$.

Henry, Kitamura, and Salanie (HKS) [4]

• **Test of Exclusion Restriction + # Components (L) assumptions.**

• Suppose that $|\mathcal{Y}| \geq 3$ such that there are two values $y, y' \in \mathcal{Y}$. Let $r(y,z)$ and $r(y',z)$ be the probability ratios associated with y and y', respectively.

• The model implies that:

$$r(y,z) - r\left(y',z\right) \equiv \frac{P(y \mid z) - P(y \mid z_0)}{P(y \mid z_1) - P(y \mid z_0)} - \frac{P(y' \mid z) - P(y' \mid z_0)}{P(y' \mid z_1) - P(y' \mid z_0)} = 0$$

Since is NP identified, we can construct a [Chi-square] test of this restriction.

Henry, Kitamura, and Salanie (HKS) [5]

- Define the unknown constants: $\alpha \equiv \pi(z_0)$ and $\beta \equiv \pi(z_1) - \pi(z_0)$. Since $r(z) = [\pi(z) - \pi(z_0)] / \pi(z_1) - \pi(z_0)$, we have that:

$$\pi(z) = \alpha + \beta\, r(z)$$

- And it is straightforward to show that:

$$\begin{aligned} f_0(y) &= P(y \mid z_0) - \frac{\alpha}{\beta}[P(y \mid z) - P(y \mid z_0)] \\ f_1(y) &= P(y \mid z_0) + \frac{1-\alpha}{\beta}[P(y \mid z) - P(y \mid z_0)] \end{aligned}$$

So all the model parameters, $\{\pi(z) : z \in \mathcal{Z}\}$ and $\{f_0(y), f_1(y) : y \in \mathcal{Y}\}$, are identified from the data up to two constants, α and β.

Henry, Kitamura, and Salanie (HKS) [6]

• To obtain sharp bounds on the model parameters, we need to take into account that the model imposes also restrictions on the parameters α and β.

• Without loss of generality, we can make $\beta > 0$ (choosing the sign of β is like labelling the unobserved types; i.e., $\omega = 1$ is the type with a probability that increases when z goes from z_0 to z_1).

• HKS show that the model implies the following sharp bounds on (α,β):

$$\frac{1}{1-\delta_{\text{sup}}} \leq \frac{-\alpha}{\beta} \leq r_{\text{inf}}$$

$$r_{\text{sup}} \leq \frac{1-\alpha}{\beta} \leq \frac{1}{1-\delta_{\text{inf}}}$$

where

$$r_{\textsf{inf}} \equiv \textsf{inf}_{z \in \mathcal{Z} - \{z_0, z_1\}} \, r(z)$$

$$r_{\textsf{sup}} \equiv \textsf{sup}_{z \in \mathcal{Z} - \{z_0, z_1\}} \, r(z)$$

$$\delta_{\textsf{inf}} \equiv \textsf{inf}_{y \in \mathcal{Y}} \frac{P(y|z_1)}{P(y|z_0)}$$

$$\delta_{\textsf{sup}} \equiv \textsf{sup}_{y \in \mathcal{Y}} \frac{P(y|z_1)}{P(y|z_0)}.$$

• Using these sharp bounds on (α,β) and the expression that relate the model paramaters with the data and (α,β), we can obtain sharp bounds on the model parameters, $\{\pi(z) : z \in \mathcal{Z}\}$ and $\{f_0(y), \; f_1(y) : y \in \mathcal{Y}\}$.

Point Identification: Example. "Identification to infinity"

• Since $\pi(z) = \alpha + \beta\ r(z)$, we can test the monotonicity of function $\pi(z)$ by testing the monotonicity of the identified function $r(z)$.

• Suppose that $\pi(z)$ is a monotonic function.

ASSUMPTION: There are values z_L^* and z_H^* in $\mathcal{Z}$ such that $\pi(z) = 0$ for any $z \leq z_L^*$ and $\pi(z) = 1$ for any $z \geq z_H^*$. [For instance, $z_L^* = z_0$ and $z_H^* = z_1$].

Under this assumption, all the parameters of the model are point identified.

9. APPLICATION TO GAMES

• **Aguirregabiria and Mira (2015):** "Identification of Games of Incomplete Information with Multiple Equilibria and Unobserved Heterogeneity".

• This paper deals with the identification, estimation and counterfactuals in empirical games of incomplete/asymmetric information when there are **three sources of unobservables for the researcher**:

1. Payoff-Relevant variables, common knowledge to players **(PR)**;
2. Payoff-Relevant variables, players' private information **(PI)**;
3. Non-Payoff-Relevant or "Sunspot" variables, common knowledge to players **(SS)**;

• Previous studies have considered: only [**PI**]; or [**PI**] and [**PR**]; or [**PI**] and [**SS**]; but not the three together.

EXAMPLE (Based on Todd & Wolpin's "Estimating a Coordination Game within the Classroom")

• In a class, students and teacher choose their respective levels of effort. Each student has preferences on her own end-of-the-year knowledge. The teacher cares about the aggregate end-of-the-year knowledge of all the students.

• A production function determines end-of-the-year knowledge of a student: it depends on student's own effort, effort of her peers, teacher's effort, and exogenous characteristics.

• **PR unobs:** Class, school, teacher, and student characteristics that are known by the players but not to the researcher.

• **PI unobs:** Some student's and teacher's skills may be private info.

• **SS unobs:** Coordination game with multiple equilibria. Classes with the same PR (human capital) characteristics may select different equilibria.

WHY IS IT IMPORTANT TO ALLOW FOR PR and SS UNOBS. ?

[1] Ignoring one type of heterogeneity typically implies that we over-estimate the contribution of the other.

• **Example**: In Todd and Wolpin, similar schools (in terms of observable inputs) have different outcomes mainly because they have different PR unobservables (e.g., cost of effort); or mainly because they have selected a different equilibrium.

[2] Counterfactuals: The two types of unobservables (PR and SS) enter differently in the model. They can generate very counterfactual policy experiments.

CONTRIBUTIONS OF THE PAPER

- We study identification when the **three sources of unobservables may be present** and in a **fully nonparametric** model for payoffs, equilibrium selection mechanism, and distribution of PR and SS unobservables.

- Specific contributions. **IDENTIFICATION:**

    1. Under standard exclusion conditions for the estimation of games, we show that the payoff function, and the distributions of PR and SS unobserved heterogeneity are NP identified.

    2. Test of the hypothesis of "No PR unobservables" (it does not require "all" the exclusion restrictions);

DISCRETE GAMES OF INCOMPLETE INFORMATION

• N players indexed by i. Each player has to choose an action, a_i, from a discrete set $\mathcal{A} = \{0, 1, ..., J\}$. to maximize his expected payoff.

• The payoff function of player i is:

$$\Pi_i = \pi_i(a_i, \boldsymbol{a}_{-i}, \mathbf{x}, \omega) + \varepsilon_i(a_i)$$

- $\boldsymbol{a}_{-i} \in \mathcal{A}^{N-1}$ is a vector with choices of players other than i;

- $\mathbf{x} \in \mathcal{X}$ and $\omega \in \Omega$ are exogenous characteristics, common knowledge for all players. $\mathbf{x}$ is observable to the researcher, and ω is the **Payoff-Relevant (PR) unobservable**;

- $\boldsymbol{\varepsilon}_i = \{\varepsilon_i(a_i) : a_i \in \mathcal{A}\}$ are private information variables for player i, and are unobservable to the researcher.

BAYESIAN NASH EQUILIBRIUM

• A **Bayesian Nash equilibrium (BNE)** is a set of strategy functions $\{\sigma_i(\mathbf{x}, \omega, \varepsilon_i) : i = 1, 2, ..., N\}$ such that any player maximizes his expected payoff given the strategies of the others:

$$\sigma_i(\mathbf{x}, \omega, \varepsilon_i) = \arg\max_{a_i \in \mathcal{A}} \; \mathbb{E}_{\boldsymbol{\varepsilon}_{-i}} \left(\; \pi_i(a_i, \; \sigma_{-i}(\mathbf{x}, \omega, \boldsymbol{\varepsilon}_{-i}), \; \mathbf{x}, \; \omega) \; \right) + \varepsilon_i(a_i)$$

• It will be convenient to represent players' strategies and BNE using **Conditional Choice Probability** (CCPs) **functions:**

$$P_i\left(a_i \mid \mathbf{x}, \omega\right) \equiv \int 1\left\{\sigma_i(\mathbf{x}, \omega, \boldsymbol{\varepsilon}_i) = a_i\right\} \; dG_i(\boldsymbol{\varepsilon}_i)$$

• In this class of models, existence of at least a BNE is guaranteed. There may be **multiple equilibria**.

MULTIPLE EQUILIBRIA

• For some values of $(\mathbf{x}, \omega)$ the model has multiple equilibria. Let $\Gamma(\mathbf{x}, \omega)$ be the set of equilibria associated with $(\mathbf{x}, \omega)$.

• We assume that $\Gamma(\mathbf{x}, \omega)$ is a discrete and finite set (see Doraszelski and Escobar, 2010) for regularity conditions that imply this property.

• Each equilibria belongs to a particular "type" such that a marginal perturbation in the payoff function implies also a small variation in the equilibrium probabilities within the same type.

• We index equilibrium types by $\tau \in \{1, 2, ...\}$.

DATA, DGP, AND IDENTIFICATION

- The researcher observes T realizations of the game; e.g., T markets.

$$Data = \{ \ a_1, \ a_{2t}, ..., a_{Nt}, \ \mathbf{x}_t : t = 1, 2, ..., T \ \}$$

- DGP.

(A) $(\mathbf{x}_t, \omega_t) \sim i.i.d.$ draws from CDF $F_{x,\omega}$. Support of ω_t is discrete (finite mixture);

(B) The equilibrium type selected in observation t, τ_t, is a random draw from a probability distribution $\lambda(\tau|\mathbf{x}_t, \omega_t)$;

(C) $\boldsymbol{a}_t \equiv (a_1, a_{2t}, ..., a_{Nt})$ is a random draw from a multinomial distribution such that:

$$\Pr(\boldsymbol{a}_t \mid \mathbf{x}_t, \omega_t, \tau_t) = \prod_{i=1}^{N} P_i(a_{it} \mid \mathbf{x}_t, \omega_t, \tau_t)$$

IDENTIFICATION PROBLEM

• Let $Q(\boldsymbol{a}|\mathbf{x})$ be the probability distribution of observed players' actions conditional on observed exogenous variables: $Q(\boldsymbol{a}|\mathbf{x}) \equiv \Pr(\boldsymbol{a}_t = \boldsymbol{a} \mid \mathbf{x}_t = \mathbf{x})$.

• Under mild regularity conditions, $Q(.|.)$ is identified from our data.

• According to the model and DGP:

$$Q(\boldsymbol{a}|\mathbf{x}) = \sum_{\omega \in \Omega} \sum_{\tau \in \Upsilon(\mathbf{x},\omega)} F_{\omega}(\omega|\mathbf{x}) \; \lambda(\tau|\mathbf{x},\omega) \; \left[\prod_{i=1}^{N} P_i(a_{it} \mid \mathbf{x}_t, \omega_t, \tau_t; \boldsymbol{\pi}) \right] \tag{1}$$

• *The model is (point) identified if given* Q *there is a unique value* $\{\boldsymbol{\pi}$, F_{ω}, $\lambda\}$ *that solves the system of equations (1).*

IDENTIFICATION QUESTIONS

• We focus on three main identification questions:

1. Sufficient conditions for point identification of $\{\boldsymbol{\pi}, F_{\omega}, \lambda\}$;
2. Test of the null hypothesis of No PR unobservables;
3. Test of the null hypothesis of No SS unobservables;

• With a nonparametric specification of the model, is it possible to reject the hypothesis of "No SS unobservables" and conclude that we need "multiple equilibria" to explain the data?

THREE-STEPS IDENTIFICATION APPROACH

- Most of our identification results are based on a three-step approach.

- Let $\xi \equiv g(\omega, \tau)$ be a scalar discrete random variable that represents all the unobserved heterogeneity, both PR and SS. ξ does not distinguish the source of this heterogeneity.

- Let $H(\xi|\mathbf{x})$ be the PDF of ξ, i.e., $H(\xi|\mathbf{x}) = F_\omega(\omega|\mathbf{x})\ \lambda(\tau|\mathbf{x}, \omega)$

STEP 1. NP identification of $H(\xi|\mathbf{x})$ and CCPs $P_i(a_i|\mathbf{x},\xi)$ that satisfy restrictions:

$$Q(a_1,a_2,...,a_N \mid \mathbf{x}) = \sum_{\xi} H(\xi|\mathbf{x}) \left[\prod_{i=1}^{N} P_i(a_i \mid \mathbf{x}, \xi)\right]$$

• We use results from the literature of **identification of NPFM based on conditional independence restrictions**.

STEP 2. Given the CCPs $\{P_i(a_i|\mathbf{x}, \xi)\}$ and the distribution of ε_i, it is possible to obtain the *differential-expected-payoff function* $\widetilde{\pi}_i^P(a_i, \mathbf{x}, \xi)$.

• $\widetilde{\pi}_i^P(a_i, \mathbf{x}, \xi)$ is the expected value for player i of choosing alternative a_i minus the expected value of choosing alternative 0. By definition:

$$\widetilde{\pi}_i^P(a_i, \mathbf{x}, \xi) \equiv \sum_{a_{-i}} \left(\prod_{j \neq i} P_j(a_j|\mathbf{x}, \xi) \right) \; [\pi_i(a_i, \boldsymbol{a}_{-i}, \mathbf{x}, \omega) - \pi_i(0, \boldsymbol{a}_{-i}, \mathbf{x}, \omega)]$$

• Given this equation and the identified $\widetilde{\pi}_i^P$ and $\{P_j\}$, we study the identification of the payoff π_i.

• We use exclusion restrictions that are standard for the identification of games.

STEP 3. Given the identified payoffs π_i and the distribution $H(\xi|\mathbf{x})$, we study the identification of the distributions $F_\omega(\omega|\mathbf{x})$ and $\lambda(\tau|\mathbf{x},\omega)$.

• Testing the null hypothesis of "No PR heterogeneity" does not require steps 2 and 3, but only step 1.

• This three-step approach does not come without loss of generality. Sufficient conditions of identification in step 1 can be 'too demanding'. We have examples of NP identified models that do not satisfy identification in step 1.

IDENTIFICATION IN STEP 1

• Point-wise identification (for every value $\mathbf{x}$) of the NP finite mixture model:

$$Q(a_1,a_2,...,a_N \mid \mathbf{x}) = \sum_{\xi} H(\xi|\mathbf{x}) \ \left[\prod_{i=1}^{N} P_i(a_i \mid \mathbf{x}, \xi)\right]$$

• Identification is based on the independence between players' actions once we condition on $(\mathbf{x}, \xi)$.

• We exploit results by Hall and Zhou (2003), Hall, Neeman, Pakyari, and Elmore (2005), and Kasahara and Shimotsu (2010).

IDENTIFICATION IN STEP 1 (II)

• Let L_ξ is the number of "branches" that we can identify in this NP finite mixture.

PROPOSITION 1. Suppose that: (a) $N \geqslant 3$*; (b)* $L_\xi \leqslant (J+1)^{int[(N-1)/2]}$*; (c)* $\mathbf{P}_{Y_j}(\xi=1)$, $\mathbf{P}_{Y_j}(\xi=2)$, ..., $\mathbf{P}_{Y_j}(\xi=L_\xi)$ *are linearly independent. Then, the distribution* H *and players' CCPs* P_i*'s are uniquely identified, up to label swapping.* ■

• We cannot identify games with two players.

• With $N \geqslant 3$ we can identify up to $(J+1)^{int[(N-1)/2]}$ market types.

IDENTIFICATION IN STEP 2 (two players)

• In a binary choice game with two players, i and j, the equation in the second step is:

$$\widetilde{\pi}_i^P(\mathbf{x},\xi) \equiv \alpha_i(\mathbf{x},\omega) + \beta_i(\mathbf{x},\omega)\ P_j(\mathbf{x},\xi)$$

where:

$$\begin{aligned} \alpha_i(\mathbf{x},\omega) &\equiv \pi_i(1,0,\mathbf{x},\omega) \\ \beta_i(\mathbf{x},\omega) &\equiv \pi_i(1,1,\mathbf{x},\omega) - \pi_i(1,0,\mathbf{x},\omega) \end{aligned}$$

• We know $\widetilde{\pi}_i^P(\mathbf{x},\xi)$ and $P_j(\mathbf{x},\xi)$ for every $(\mathbf{x},\xi)$, and we want to identify $\alpha_i(.,.)$ and $\beta_i(.,.)$. This is "as if" we were regressing $\widetilde{\pi}_i^P(\mathbf{x},\xi)$ on $P_j(\mathbf{x},\xi)$.

IDENTIFICATION IN STEP 2 [2]

• From the first step, we do not know if ξ is PR or SS unobserved heterogeneity. The worst case scenario for identification in the second step is that all the unobservables are PR:

$$\widetilde{\pi}_i^P(\mathbf{x}, \xi) \equiv \alpha_i(\mathbf{x}, \xi) + \beta_i(\mathbf{x}, \xi) \; P_j(\mathbf{x}, \xi)$$

• Then, the "parameters" $\alpha_i(\mathbf{x}, \xi)$ and $\beta_i(\mathbf{x}, \xi)$ have the same dimension (sources of variation) as the known function $\widetilde{\pi}_i^P(\mathbf{x}, \xi)$ and $P_j(\mathbf{x}, \xi)$ and identification is not possible without additional restriction.

• This identification problem appears even without unobserved heterogeneity:

$$\widetilde{\pi}_i^P(\mathbf{x}) \equiv \alpha_i(\mathbf{x}) + \beta_i(\mathbf{x}) \; P_j(\mathbf{x})$$

IDENTIFICATION IN STEP 2 [3]

ASSUMPTION [Exclusion Restriction]. $\mathbf{x} = \{\mathbf{x}^c, z_i, z_j\}$ *where* $z_i, z_j \in \mathcal{Z}$ *and the set* $\mathcal{Z}$ *is discrete with at least* $J+1$ *points, and*

$$\pi_i(a_i, \boldsymbol{a}_{-i}, \mathbf{x}, \omega) = \pi_i(a_i, \boldsymbol{a}_{-i}, \mathbf{x}^c, z_i, \omega)$$

[Relevance] *And there are* $z_i^0 \neq z_i^1$ such that $P_j(\mathbf{x}^c, z_j, z_i^0, \xi) \neq P_j(\mathbf{x}^c, z_j, z_i^1, \xi)$.

PROPOSITION 3. Under the Exclusion Restriction + Relevance assumptions, the payoff functions π_i *are identified.* ■

IDENTIFICATION IN STEP 3

• Let $\mathbf{\Pi}_i(\mathbf{x})$ be the matrix with dimension $J(J+1)^{N-1} \times L_\xi$ that contains all the payoffs $\{\pi_i(a_i, \boldsymbol{a}_{-i}, \mathbf{x}, \xi)\}$ for a given value of $\mathbf{x}$. Each column corresponds to a value of ξ and it contains the payoffs $\pi_i(a_i, \boldsymbol{a}_{-i}, \mathbf{x}, \xi)$ for every value of $(a_i, \boldsymbol{a}_{-i})$ with $a_i > 0$.

• If two values of ξ represent the same value of ω, then the corresponding columns in the matrix $\mathbf{\Pi}_i(\mathbf{x})$ should be equal.

• Therefore, the number of distinct columns in the payoff matrix $\mathbf{\Pi}_i(\mathbf{x})$ should be equal to L_ω. That is, we can identify the number of mixtures L_ω as:

$$L_\omega(\mathbf{x}) = \text{Number of distinct columns in } \mathbf{\Pi}_i(\mathbf{x})$$

PROPOSITION 5. Under the conditions of Propositions 1 and 3, the one-to-one mapping $\xi = g(\omega, \tau)$ and the probability distributions of the unobservables, $F_\omega(\omega|\mathbf{x})$ and $\lambda(\tau|\mathbf{x}, \omega)$, are nonparametrically identified. ■

TEST OF HYPOTHESIS "NO PR UNOBSERVABLES"

•

TEST OF HYPOTHESIS "NO SS UNOBSERVABLES"